

Multi-Word Verb Usage by Turkish Learners of English

A Corpus-Based Study

Seval Özen

**Multi-Word Verb Usage by Turkish
Learners of English: A Corpus-Based
Study**

Inaugural-Dissertation
zur
Erlangung der Doktorwürde
der Philologischen Fakultät
der Albert-Ludwigs-Universität
Freiburg i. Br.

vorgelegt von

Seval Özen, geb. als Kömürcü
aus Bergama, İzmir

WS 2020/2021

Erstgutachterin: Prof. Dr. Brigitte Halford

Zweitgutachter: Prof. Dr. Dr. h.c. Christian Mair

Vorsitzender des Promotionsausschusses der Gemeinsamen Kommission
der Philologischen und
Philosophischen Fakultät: Prof. Dr. Dietmar Neutatz

Datum der Disputation: 21.07.2021

Table of Contents

Table of Contents	iii
List of Tables	ix
Abbreviations	xii
Acknowledgements.....	xiv
Chapter 1: Introduction.....	1
1.1 Background	1
1.2 Aims and scope of this study.....	11
1.3 English learning and teaching in Turkey.....	17
1.4 The structure of thesis.....	26
Chapter 2: Lexicon.....	29
2.1 The role of the lexicon in linguistics and language teaching.....	29
2.2 The lexicon and the foreign language learner	44
2.3 Phraseology and multi-word units in native and non-native speaker language and their role in language learning and use	68
2.3.1 Phraseology	68
2.3.2 The notion of multi-word units – their definition, characteristics and categorization	72
2.3.3 Multi-word units in native language and their role in L1 acquisition and use.....	82
2.3.4 Multi-word units and the learner	112
Chapter 3: Investigating multi-word verbs in a learner corpus	193

3.1 The notion of ‘multi-word verbs’ – definition and categories	193
3.2 Multi-word verb categories investigated in this study	201
3.2.1 Phrasal verbs	201
3.2.2 Prepositional verbs	219
3.2.3 Phrasal-prepositional verbs	222
3.2.4 Verb-noun collocations	227
3.3 Summary	236
Chapter 4: Data and Methodology.....	237
4.1 Preliminaries	237
4.2 Corpora.....	240
4.3 Tools.....	256
4.3.1 Software for Retrieval and Analysis.....	256
4.3.2 Reference Tools	258
4.4 ‘Acceptability’ or ‘Appropriateness’ and the Role of the Native Speaker	260
4.5 The Extraction of Multi-Word Verbs	262
4.5.1 The extraction of phrasal and phrasal-prepositional verbs	266
4.5.2 The extraction of verb-noun collocations ..	270
4.6 Some words on terminology	272
4.7 Summary	279
Chapter 5: Multi-word verbs in learner writing – a quantitative approach	281
5.1 Preliminaries	281

5.2 Phrasal and phrasal-prepositional verbs in TICLE	285
5.2.1 Overall results regarding phrasal and phrasal- prepositional verb usage	286
5.2.2 Frequency of individual phrasal verbs	290
5.2.3 Frequency of individual phrasal-prepositional verbs	313
5.2.4 Productivity of verbs and particles in forming phrasal and phrasal-prepositional verbs	319
5.2.5 Extralinguistic factors correlating with the quantitative use of phrasal and phrasal- prepositional verbs	331
5.2.6 Synopsis: The quantitative use of phrasal and phrasal-prepositional verbs	352
5.3 Prepositional verbs in TICLE	354
5.3.1 Overall results regarding prepositional verb usage	355
5.3.2 Frequency of individual prepositional verbs	357
5.3.3 Productivity of verbs and prepositions in forming prepositional verbs	363
5.3.4 Synopsis: The quantitative use of prepositional verbs	371
5.4 Verb-noun collocations in TICLE	372
5.4.1 Overall results regarding the use of verb-noun collocations	374
5.4.2 Frequency of individual verb-noun collocations	375

5.4.3 Productivity of verbs in forming verb-noun collocations.....	383
5.4.4 Synopsis: The quantitative use of verb-noun collocations.....	404
Chapter 6: Multi-word verbs in learner writing – a qualitative approach.....	407
6.1 Preliminaries.....	407
6.2 Erroneous and unnatural usages in the context of phrasal, prepositional and phrasal-prepositional verbs.....	411
6.3 Erroneous and unidiomatic usages in the context of verb-noun collocations.....	456
6.4 Synopsis: The qualitative use of multi-word verbs.....	494
Chapter 7: Summary and conclusions	499
7.1 Introductory remarks.....	499
7.2 Summary of the main findings	500
7.3 Pedagogical implications	509
7.4 Limitations of the study and ways forward	524
Bibliography.....	532
Appendices	595
Appendix 1. Frequency list of phrasal verb types per million words (raw).....	595
Appendix 2. Frequency list of phrasal-prepositional verb types per million words (raw)	606

Appendix 3. Productive verbs that combine with particles to form phrasal and phrasal-prepositional verbs.....	607
Appendix 4. Particles	609
Appendix 4a. Particle productivity (combination with X verb types).....	609
Appendix 4b. Frequency of particles per million words	610
Appendix 5. Frequency list of prepositional verb types in TICLE (raw)	611
Appendix 6. Productive verbs that combine with prepositions to form prepositional verbs	618
Appendix 7. Productivity (combination with X verb types) and frequency (raw) of prepositions.....	619
Appendix 8. Frequency lists of verb-noun collocations in TICLE	620
Appendix 8a. Group I verb-noun collocations (verb + (determiner) + noun) occurring in the essays of 3 or more learners.....	620
Appendix 8b. Group II verb-noun collocations (verb + noun + preposition)	629
Appendix 8c. Group III verb-noun collocations (verb + prepositional phrase)	630
Appendix 9. Frequent noun collocates used in delexical verb-noun collocations in TICLE.....	632
Zusammenfassung in der deutscher Sprache.....	637

List of Figures

Figure 1: Concordance lines retrieved for <i>go back (to)</i> in TICLE	269
Figure 2. Total number of verb tokens compared to the total number of phrasal and phrasal- prepositional verb tokens in TICLE and LOCNESS	290
Figure 3: Number of different particle combinations with ten productive verbs in TICLE, along with the corresponding figures from LOCNESS	322
Figure 4. Number of verb types combined with the productive particles in TICLE, along with the corresponding figures from LOCNESS	324
Figure 5. Text length and frequency of phrasal and phrasal-prepositional use in TICLE.....	343
Figure 6. Correlation between text length and frequency of phrasal and phrasal-prepositional verb use in TICLE	344
Figure 7. Correlation between topic and frequency of phrasal and phrasal-prepositional verb use in TICLE	349

List of Tables

Table 1. Diagnostic frames for phrasal and prepositional verbs.	210
Table 2. Types of phrasal verbs investigated in the present study.....	211
Table 3. Distribution of essay topics in TICLE	246
Table 4. Corpus information	255
Table 5. The 25 most frequent phrasal verbs in TICLE compared with LOCNESS.....	293
Table 6. <i>continue</i> versus phrasal verbs <i>go on</i> , <i>carry on</i> and <i>keep on</i>	294
Table 7. <i>bring about</i> versus <i>cause</i>	296
Table 8. <i>sum up</i> , <i>summarize</i> , <i>in sum/summary</i> and some alternatives with <i>conclude</i>	299
Table 9. Frequencies of <i>carry out</i> and <i>point out</i>	304
Table 10. The 25 most frequent phrasal verbs in LOCNESS along with their frequencies in TICLE	310
Table 11. The most frequent phrasal-prepositional verbs in TICLE along with their frequencies in LOCNESS	314
Table 12. The productive verbs in TICLE and the particles they combine with.....	322
Table 13. The frequencies of the six most frequent particles pmw (raw)	330

Table 14. The impact of the independent variable text length on the use of phrasal and phrasal-prepositional verb use.....	339
Table 15. Results of linear regression analysis (stepwise) for the overall phrasal and phrasal-prepositional verb frequency for each learner.	341
Table 16. Results of linear regression analysis (enter method) for the relative phrasal and phrasal-prepositional verb frequency for each learner.	342
Table 17. Descriptive results of the one-way ANOVA test (dependent variable relative frequency – topic (n>9)).....	348
Table 18. The 25 most frequent prepositional verbs in TICLE	358
Table 19. The productive verbs in TICLE and the prepositions they combine with.....	366
Table 20. Number of prepositional verb types formed with the productive prepositions in TICLE, along with raw frequency of each preposition....	368
Table 21. The most frequent verb-noun collocations in TICLE	376
Table 22. Most productive verb types used in the formation of verb-noun collocations, along with absolute frequencies (percentage they represent)	384
Table 23. Types of error and unidiomatic usage attested in the production of phrasal-, prepositional- and phrasal-prepositional verbs in TICLE.....	415

Table 24. Types of error and unidiomatic usage attested in the production of verb-noun collocations in TICLE	459
--	-----

Abbreviations

Corpora and Dictionaries

BNC	<i>British National Corpus</i>
COCA	<i>Corpus of Contemporary American English</i>
ChiCLE	Chinese component of the <i>International Corpus of Learner English</i>
G-ICLE	German component of the <i>International Corpus of Learner English</i>
GloWbE	<i>Corpus of Web-Based Global English</i>
ICLE	<i>International Corpus of Learner English</i>
I-ICLE	Italian component of the <i>International Corpus of Learner English</i>
LGSWE	<i>Longman Grammar of Spoken and Written English</i>
LOCNESS	<i>Louvain Corpus of Native English Essays</i>
SWeCLE	Swedish component of the <i>International Corpus of Learner English</i>
TICLE	Turkish component of <i>International Corpus of Learner English</i>
OALD	<i>Oxford Advanced Learner's Dictionary</i>
OED	<i>Oxford English Dictionary</i>

Other abbreviations:

CIA	Contrastive Interlanguage Analysis
EFL	English as a foreign language
ELT	English language teaching
ESL	English as a second language
IL	Interlanguage
L1	First language

L2	Second language
L3	Third language
LCR	Learner Corpus Research
NS	Native speaker
NNS	Nonnative speaker
SLA	Second language acquisition
TL	Target language
MWV	Multi-word verb

Acknowledgements

Writing a thesis can be dull, gloomy and stressful at times. Fortunately, I was lucky to have people around me whose support and love reduced the number and length of those moments appreciably. I owe an enormous debt of gratitude to them.

First and foremost, I wish to express gratitude to my supervisor Prof. Dr. Brigitte Halford. I am indebted to her for showing me respect and support as both her student and assistant over the past decade, which flew away far too quickly. She has been a constant source of advice and encouragement in all matters, academic and beyond. Additionally, I warmly thank my second supervisor Prof. Dr. Christian Mair for sparking my interest in corpus linguistics and for offering valuable comments on portions of the early drafts of this thesis.

I am deeply grateful to all my friends for being there whenever I needed. I am profoundly indebted to Christine Kollmar and Jada Schumacher for their meticulous readings of early versions of my manuscript. For the past three years they devoted an inordinate amount of time and energy to help me mold this thesis into its present shape. Despite the distance, they have always been close to me; their endless help, support and love have been invaluable. I am also thankful to them for enduring all the grammatical challenges I put them through. I can truthfully say that without their help, I would have never completed this thesis.

Warm and heartfelt thanks to Prof. Dr. Göz Kaufmann who saw me through this work with a unique combination of intellectual and emotional support. His detailed and insightful corrections helped me avoid many errors. Along with Dr. Zeynep Köylü, the two of them were extremely helpful to me in navigating the statistics.

I also owe thanks to my friends and earlier students Patrice Noé Ross, David L. Norcross, Christian Blake Dauphinais and Sonya Stinsolotti for their readiness to share their thoughts on the problematic constructions in my data. I also would like to thank Alison Nagel and Christopher Wilson who kindly responded to my questions about some multi-word units and Holly Bayer who proofread this very section. As true friends do, Selma Kahraman, Anne Eichner and Kübra Arslan contributed indirectly to this thesis by continuously encouraging me during the past years.

Sincere thanks to Prof. Dr. Marcus Callies for inviting me as a guest lecturer to give a talk on my PhD topic. The follow-up discussion with him and his research group was a great source of inspiration.

My deepest appreciation goes to Christel and Josef Gaida – my daughter’s ‘Oma’ and ‘Opa’ – for their tremendous support during the entire process of writing this thesis. It was deeply reassuring to know that whenever needed, they were there for me.

I would also like to thank my parents Sevim and Adil Kömürçü as well as my grandmother Meliha Kömürçü for their endless love, support and prayers. I am indebted to my mother Semra Özen for her endless

strength and for always trying to make our lives easier – even from such a long distance.

I owe special thanks to my daughter Umay Feraye for the joy and happiness she brings into my life. Her mere existence was a great motivation to work harder. Finally, I am eternally grateful to my husband Ali for encouraging and comforting me, for always believing in me, and not least, for reminding me as well as for enabling me to take a break occasionally. My PhD was self-sponsored. Full-time parenting, with all the joys and struggle it includes continued on as I worked. My husband and I patiently postponed lots of things for the last three years in order for me to finish this thesis. I hope to make up for the times I could not be with him and our daughter while I focused my attention on this project.

Chapter 1: Introduction

*Learn a new language, but keep the old;
one is silver and the other gold.*
(Patsy M. Lightbown)

1.1 Background

After the first attention-shift from grammar to lexis seen in the 1960s, the focus of the last 15 years or so has turned towards lexical groups, i.e. to multi-word units such as collocations, phrasal verbs, chunks, etc. (this shift will be discussed at more length in Chapter 2). From the 1980s onwards, the increasing availability of large electronic corpora has made it possible to explore vast amounts of authentic language data both qualitatively and quantitatively without difficulty and uncover the highly patterned nature of language use. Whereas formerly multi-word units were treated as peripheral exceptions and linguistic oddities, a substantial body of research now exists to show that natural language use, both written and spoken, largely consists of recurrent word sequences. Therefore, these recurrent word sequences have increasingly been regarded as important building blocks for language acquisition and use. In the context of foreign or second language learning¹, multi-

¹ In Second Language Acquisition (SLA) research, there is a traditional distinction made between ‘second language acquisition’ and ‘foreign language learning’ according to the environment the ‘target language’ (TL) is learned: ‘Second language acquisition’ is generally used to refer to learning that takes place in an environment where the TL is

word units constitute a particularly interesting phenomenon since they are known to cause problems for language learners.

The present work, which is situated in Learner Corpus Research (LCR), focuses on one variety of learner English, namely Turkish learner English, and measures the learners' use of multi-word verbs – a particular category of multi-word units – since their acquisition and active usage cause great difficulty even for advanced learners of English.

Ellis and Barkhuizen (2005: x) suggest that “[o]ne way of learning about how learners acquire an L2 is by studying the language they produce.” Following their suggestion, written productions of Turkish learners at a level ranging from (higher-)intermediate to advanced will be examined. The corpus-based investigation of learner language offers the opportunity to gain important insights into the difficulties learners

a common part of daily communication and learning happens mostly without guidance of teaching or books; ‘foreign language learning’, on the other hand, is used to refer to language learning via formal instruction, i.e. in a classroom situation. In later work on applied linguistics ‘foreign language learning’ is used as a “blanket term” to cover both natural second language acquisition, i.e. through exposure, and second language learning in classroom environments (Ringbom 1987: 27) (for the effect of contextual differences these two situations have on learners, see Ringbom 1987: 26-31 and for the details of the differences of these two learning environments see e.g. Wilkins 1972: 150-156). However, the situation in reality is more complex, as shown in recent years (e.g. Gilquin & Granger 2011; Edwards & Laporte 2015). In the present study the terms ‘(language) learner’ and ‘L2 learner’ are used interchangeably to refer to those who have been learning another language after having acquired their mother tongue. Accordingly, ‘language learning’ is used for the learning of another language after acquisition of one’s mother tongue is complete.

encounter at different stages of the learning process. Learner corpus research can therefore contribute not only to a better understanding of learner language, but also to the improvement of teaching and learning materials.

Multi-word verbs are complex both in terms of their grammatical form and their lexical meaning, and they are required to be learnt, stored and retrieved as a whole (see Wray and Perkins 2000). Siyanova and Schmitt (2007: 119) state that “[u]nless a learner knows that a string of words is a multi-word verb, they are likely to try to decode the meanings of the individual words”, which may work in some cases but not all, since the “meanings of multi-word verbs vary on a cline of transparency.” The multi-word verbs that carry non-literal, idiomatic meaning present the greatest problems for learners due to the “mismatch” of the idiomatic meaning and the meanings of the individual constituents of the verb (Siyanova & Schmitt 2007: 120). This “mismatch” can be considered ‘illogical’ by the learner at times – an example provided by Cornell (1985: 274) well illustrates this point: “why should one be *laid up* with an illness when one is *lying down*?” Moreover, as noted by Cornell (1985: 270) for phrasal verbs, “[t]he reduction in the learning problem represented by non-idiomatic phrasal verbs is to some extent cancelled out by the phenomenon of polysemy.” That is, the polysemy of some multi-word verbs add to their difficulty, e.g. *to turn the volume down* meaning ‘to lower the intensity/lessen’ versus *turn someone down* (idiomatic phrasal verb) meaning ‘reject or refuse someone’. The problem caused by the polysemy factor is exacerbated by the fact that there are

some multi-word verbs that have more than one idiomatic meaning in addition to non-idiomatic one(s), as can be seen in the example sentences of *break up* (1-7) – the first six meanings of *break up* and the sentences illustrating the meanings are taken from Cowie and Mackin (1975) and the very last one is from Bywater (1969: 99):

(1) *The ship was caught in a south-easterly gale. She went aground and started to **break up**.* ('disintegrate, shatter')

(2) *I know what he has to endure, but there is no sign of him **breaking up** under the strain yet.* ('crumble, disintegrate – mentally or physically')

(3) *The meeting was threatening to get out of hand and the police were forced to **break it up**.* ('disperse or scatter (often by force)')

(4) *When do you **break up** for the Easter holidays?* ('disperse (at the end of a term)')

(5) *Their marriage has virtually **broken up** – they are hardly ever in each other's company.* ('dissolve, come to an end')

(6) *The ships were **broken up** and sold as scrap to dealers.* ('divide in pieces by cutting, smashing')

(7) *The weather is **breaking up**.* ('The good weather is coming to an end').

Gardner and Davies (2007) have found that the most frequent 100 phrasal verbs in the 100-million-word BNC, which make up nearly half of all phrasal verbs in the corpus, have in total 559 potential meanings

– that is, 5.6 meanings per phrasal verb on average (p. 353). The researchers also noted that *break up* has the highest number of meanings with 19 (2007: 352).

Restricted co-occurrence of multi-word verbs, in the sense that they only occur with specific context words or collocates, is another learning burden for learners: *make a mistake* but not **do a mistake*. As collocational restrictions vary from one language to another, incongruent collocations between their L1 and L2 are a notable source of problems for learners. For example, the learners face the tasks to differentiate and avoid non-target-like word sequences in their L2 (see Section 2.3.3).

Another difficulty of multi-word verbs for learners may be related to their grammatical constraints, or, as referred to by Cornell (1985: 274), to their grammatical “peculiarities”: *come by*, for example, cannot normally be used in the passive unlike its equivalents *acquire* and *obtain* or unlike *start* and *begin*, which can be followed by either the gerund or infinitive form, or *set about* is followed only by the gerund (Cornell 1985: 275). What is more, there are some word sequences that are actually two different verb structures resulting in different meanings and uses. For instance, the multi-word verbs in ‘They *ran over* the bridge’ and ‘They *ran over* the cat’, despite their formal similarity, are different multi-word verb types: the former being a prepositional verb, and the latter a phrasal verb. Whereas phrasal verbs allow the particle movement, unless the noun phrase following the verb is a personal pronoun, prepositional verbs do not allow this movement in any case: **They ran*

the bridge over versus *They ran the cat over* (see Section 3.2). Confusion, occasionally even at advanced levels of proficiency, also occurs due to prepositions and particles, which tend to be difficult to acquire as a result of various reasons proposed in the literature such as their multifunctionality, semantic complexity, polysemy, arbitrariness, etc.: *look after* vs. *look for* or *make out* vs. *make up* (Cornell 1985: 273f). All of these features of multi-word verbs can be seen as factors that increase the learning burden for learners.

As a result of the difficulty of multi-word verbs, some learners of English have been found to avoid using them for fear of making mistakes and have a tendency to use, if available, their one-word counterparts instead, regardless of their L1 background (e.g., Dagut & Laufer 1985; Hulstijn & Marchena 1989; Laufer & Eliasson 1993; Liao & Fukuya 2004; Schmitt & Redwood 2011; Riguel 2014). However, although there are usually one-word counterparts, as in the case of phrasal verbs and verb-noun combinations, “there is usually no total congruence, as the whole concept of synonymy is indeed always a question of degree” (Waibel 2007: 37). It has generally been agreed in the literature that synonyms in an absolute sense, i.e. where the two words can be used interchangeably in all their possible contexts of use, are very rare (e.g., Carter & McCarthy 1988: 28f; cf. Jackson & Amvela 2007: 108ff). For instance, as for phrasal verbs and their one-word equivalents, Sinclair (1989: iv) state that they cannot be substituted appropriately for one another because they often have a different range of use, meaning or

collocation. As noted by Cornell (1985: 274f), phrasal verbs are usually more specific in meaning compared to their one-word equivalents and carry undertones which their users must be cognizant of, e.g. *put up with* cannot be used in a positive manner unlike its often provided one-word verb equivalent *tolerate*; *to put up with other people's opinions* is not the same as *to tolerate other people's opinions* (see Cornell 1985: 274f. for a host of further examples). Furthermore, stylistic aspects of language are another relevant issue in the choice of 'equivalents'. Multi-word verbs show a tendency to be more "colloquial in tone" (Biber et al. 1999: 409) compared to their Latinate one-word alternatives. Therefore, one-word verbs are usually more appropriate in a formal register. In Sinclair et al.'s words, one-word verbs are often "out of place" and learners opting for them "run the risk of sounding pompous or just unnatural" (1989: iv). Moreover, some multi-word verbs do not have one-word equivalents; equivalents are either multi-word verbs themselves (e.g. *to make up for* = *to compensate for*, *make an effort* = *put effort into something*) or paraphrasing is necessary to explain their precise meaning (e.g. *to get (a)round to* often suggests procrastination or, as Darwin and Gray (1999: 66) explain, "*to show off* is to show off, not *to impress another with one's prowess by performing difficult yet completely unnecessary feats*").

In terms of idiomaticity and fluency, multi-word verbs are of paramount importance for attaining a native-like command of English. As noted

by Bywater (1969: 97) for phrasal and phrasal-prepositional verbs, multi-word verbs are an integral part of English. Avoidance or underuse of them by learners will result in unidiomatic language and stylistic deficits. Guo (2006: 2) states that “[t]he features which are used by the native speakers but not by the learners would be necessary for the learners if they wish to achieve the naturalness and the nativeness of the native speaker English.” In a similar vein, Wray and Perkins (2000: 2), endorsing Pawley and Syder (1983), argue that it is often the failure to use native-like multi-word units that give away advanced L2 learners as non-native. The following quotation nicely summarizes the matter:

The plain fact is that what distinguishes the writing and, above all, the speech of a good foreign student from those of an Englishman is that what an Englishman writes or says is full of these expressions [compound verbs], whereas most foreigners are frightened of them, carefully avoid them, and sound stilted in consequence (Bywater 1969: 97).

Some types of multi-word verbs, i.e. verb-particle combinations, are a feature of the Germanic language family to which English belongs (Darwin & Gray 1999; Dagut & Laufer 1985; cf. Celce-Murcia & Larsen-Freeman 1999: 425). Therefore, learners of English who are native speakers of non-Germanic languages may be particularly unfamiliar with them, and this may result in their lack of ability to deal appropriately with them (Siyanova & Schmitt 2007: 120). Turkish, belonging to the Turkic language family, is an agglutinative language. It is a head

final, left-branching, ‘free’ word order language²(Özsoy 2019: 1). As a result of belonging to different linguistic families, Turkish and English share few structural and lexical commonalities. With regard to lexicon, the only common ground between these two languages is a group of borrowings from French, e.g. *kalite* (=quality), *enflasyon* (=inflation), some of which are false friends, e.g. *sempatik* (=nice) (Thompson 2009: 224). Some English abstract words do not have Turkish equivalents, causing in turn confusion between – for instance – *mind*, *idea*, *opinion* and *thought* (Thompson 2009: 225). Groups of “not-quite-synonymous” English words (e.g. *win/earn/pass (an exam)* or *shorten/abbreviate/abridge*) often have only one Turkish counterpart (ibid.), i.e. Turkish ‘equivalents’ of English words cover a larger semantic range. Although some aspects of the English verb system have their counterparts in Turkish, there are some crucial differences. As an agglutinative language, Turkish predominantly makes use of suffixation. While English has prepositions, Turkish has postpositions. The functions of some prepositions are thus fulfilled in Turkish by case suffixes. Being a postpositional and head-final language, Turkish, therefore, has a different structure of multi-verb construction compared to English – especially concerning the lack of phrasal, prepositional, and phrasal-prepositional verbs. Multi-word verbs are quite common in Turkish (Thim 2012: 39) and compounding is a highly productive process in their formation. Turkish multi-word verbs are mainly divided

² See the studies in Özsoy (2019) for different aspects of the word order phenomena in Turkish.

into three groups: multi-word verbs composed of a noun and an auxiliary verb, multi-word verbs composed of a verb and an auxiliary verb³ and semantically contracted multi-word verbs (Savran 2008). Simple verbs are conjoined with nouns, adjectives, adverbs, verbal components, and pronominal components which have nominal, dative, ablative, or genitive suffixes. Such structural and the earlier mentioned lexical differences between English and Turkish (i.e. incongruencies in lexical “gridding” (Dagut 1977)) are believed to pose problems for Turkish-speaking learners of English in the acquisition and use of multi-word verbs in their TL. The impact of typological distance regarding multi-word verbs between L1 and L2 in the learning success of learners has already been demonstrated in previous research (e.g., Dagut & Laufer 1985; Sjöholm 1995; Wang 2016). Besides the tendency towards preferring one-word equivalence of multi-word verbs, some expected deviations in the use of multi-word verbs are omitting particles or prepositions (the units that lack in Turkish), choosing wrong particles and/or prepositions, using the wrong word order – for instance, when the separation of the verb and particle is necessary if the direct object is a pronoun as in “**Can you drop off me at the station?*” – and, lastly, unnatural, albeit creative, combinations, mostly due to direct translation from the mother tongue.

³ There are three types of auxiliary verbs in Turkish: (i) bound auxiliaries: suffixes – (y)Abil, -(y)Iver, -(y)Ayaz, -(y)Adur, -(y)Akal, (ii) free auxiliaries: verbs such as *ol*- (which is also one of the forms of the copula), *et*-, *gel*-, *dur*-, *kal*-, *düş*-, *bulun*-, *eyle*- and *buyur*-, (iii) *i*-/(*y*)-: forms of the copula (Göksel & Kerslake 2005: 157).

Some of the issues mentioned in this section will be taken up again and elaborated at various points of this study. The remainder of this chapter is organized as follows: Section 1.2 provides the aims and scope of this study, including the research questions at the core of the present study. Section 1.3 addresses some important issues regarding English learning and teaching situation in Turkey. Finally, a sketch of what will be undertaken in subsequent chapters is provided in Section 1.4.

1.2 Aims and scope of this study

Given the structural, semantic, and contrastive difficulties of multi-word verbs (see Chapter 3), they have proven to pose problems to learners of English both in their acquisition and active usage. Taking its cue from corpus-based studies on the use of English multi-word verbs by foreign learners of various L1 backgrounds (e.g., Nesselhauf 2005; Waibel 2007; Mazaherylaghab 2013; Chen 2013a, 2017; Wang 2016), the present study focuses on Turkish learner English and attempts a detailed, descriptive investigation of English multi-word verb usage in Turkish learners' written productions. In other words, the study aims to identify the difficulties faced by Turkish learners of English with regard to the use of multi-word verbs. Although there have been a few small-scale studies investigating the phraseological knowledge of Turkish-speaking learners of English, this study is the first to delve into a large-scale, exhaustive investigation of multi-word verbs in the interlanguage of this learner group.

The Turkish sub-corpus (TICLE) of the *International Corpus of Learner English* (ICLE) is the basis for this study, and the *Louvain Corpus of Native English Essays* (LOCNESS) functions as the native speaker control corpus. The learner data will be explored both in quantitative and qualitative terms – in terms of frequency of occurrence and also with respect to semantic and stylistic considerations. In order to achieve this, all multi-word verbs – including both transparent and idiomatic ones – will be extracted from the learner corpus in order to guarantee a detailed investigation. No pre-defined list will be used since basing the analysis on an existing frequency list is seen as a limit to the analysis from the start. For two groups of multi-word verbs, namely phrasal and phrasal-prepositional verbs, which have been reported to be avoided or underused by some learner groups due to L1 influence, i.e. L1-L2 dissimilarity (e.g., Dagut & Laufer 1985; Hulstijn & Marchena 1989; Schmitt & Redwood 2011), the control corpus LOCNESS will be investigated in order to identify the points of similarity and difference with respect to the use of these two multi-word verb categories between the Turkish-speaking learners of English and their native English speaker counterparts. Thus, the methodology employed in a part of this study is what Granger (1996b) termed ‘Contrastive Interlanguage Analysis’ (CIA). In CIA, there are two types of comparison employed: a comparison of native and non-native varieties of one and the same language, i.e. native speaker data (L1) vs. learner data (L2), and a compar-

ison of several non-native varieties, i.e. L2 vs. L2. In this study, investigations in the case of phrasal and phrasal-prepositional verbs are based on only the former type of comparison, namely L1 vs. L2.

This study goes beyond the previous research on the use of multi-word verbs in learner language. Instead of restricting the research to only one group of multi-word verbs, four different categories of multi-word verbs will be investigated in the present study: namely, phrasal, prepositional, and phrasal-prepositional verbs, as well as verb-noun collocations. The goal is to arrive at a more complete picture of the learning process of multi-word verbs, and to thoroughly capture difficulties learners experience in the use of varied types of multi-word verbs. This larger group of verbs will allow for comparisons between different categories of multi-word verbs and enable, in turn, more general statements about the acquisition of multi-word units by foreign language learners. Moreover, the fact that all multi-word verbs (rather than a limited number of pre-determined multi-word verbs) will be extracted allows for investigation of not only common multi-word verb types but also of less frequent ones.

Restricting the study to one L1 group was deemed necessary mainly for practical reasons: here, a large group of verbs are being investigated. Moreover, as a number of studies have already shown, the native language of a learner influences lexical learning in an L2 (e.g., Zughoul 1991; Ard and Homburg 1992; Harley & King 1989) and accordingly L1 plays a role in multi-word verb use (e.g., Biskup 1992; Wang 2016; Nesselhauf 2005; Wolter & Gyllstad 2011; Yamashita & Jiang 2010).

Moreover, as recently demonstrated by Wang (2016) in the context of delexical verb-noun collocations, learner groups with different L1 backgrounds may encounter different problems at various stages of the learning process with which they need special help.

Before moving on, it is important to note one thing at the outset: Granger (1998b: 16) has pointed out that the most advantageous access to learner language is to investigate the data with an open mind, rather than starting out from a fixed hypothesis. Therefore, although the subsequent data analysis will center around some central research questions, it will not be restricted to these questions as the study is intended to be exploratory and informative in nature.

An important aim of this study is to determine whether, and if so, to what extent, the learner's L1 influences the use of multi-word verbs in English. The typological distance resulting in the structural differences in the verb system of Turkish and English is believed to prompt the Turkish learners to use some multi-word verb categories to a minor extent. Therefore, the first research question addresses whether Turkish learners of English avoid, underuse or unexpectedly overuse the multi-word verb categories, which lack a structurally equivalent form in Turkish, in their writing compared to native speakers. Are there any multi-word verb categories that appear to be comparatively easy for the learners? If so, what are these categories? What role does the learners' L1 play in this regard?

The supposed underuse of certain multi-word verb categories raises the question of how Turkish learners deal with their lack of knowledge or compensate for it. Do they try to avoid certain multi-word verbs in English, or do they try to be creative to produce new ones, maybe by translating Turkish expressions word for word into English? Do they give new meanings to existing multi-word verbs, or do they resort to other linguistic means instead (for instance, in the case of phrasal and phrasal-prepositional verbs)? Since teachers and/or course books usually provide one-word Latinate equivalents of phrasal verbs (Side 1990), e.g. *pick up* = *receive*, investigation of the learner language to determine whether the Latinate words are preferred instead of phrasal and phrasal-prepositional verbs will be a worthwhile endeavor.⁴ So, too, the study aims to determine the possible sources of the errors. To this purpose, all uses of multi-word verbs by learners, focusing on the deviations, will be analyzed. As for the deviated uses, they will be examined to determine whether the verb or its company is erroneously selected or whether a given multi-word verb is used in the wrong context. In essence, what types of errors do the learners frequently make, and what are the possible explanations for the common error types? What might be the factors that contribute to the difficulty of (certain) multi-word verbs?

⁴ A note of caution should be added here to say that the data is based on free essay writing and does not come from elicitation tasks where learners have the option to choose between a phrasal verb and the Latinate one-word equivalent.

Since the effect of a learner's L1 on SLA is limited (i.e. selective and varies across individual learners), the present study also aims to analyze possible effects of other factors in the use of multi-word verbs (or lack thereof). The fact that quite a few learner-related variables are recorded on the ICLE-CD, such as age, years of classroom exposure to English, exposure to the target language in its natural environment, or command of other known languages as well as external variables such as use of reference tools (see Chapter 4), enables the researcher to examine the data to evaluate the possible impact of these variables on the learner's production of multi-word verbs. Some of these variables are believed to be more relevant than others as regards multi-word verb use of learners and will, thus, be investigated in the context of two categories of multi-word verbs, namely, phrasal and phrasal-prepositional verbs – the two multi-word verb categories reported to be avoided and/or underused by some learner groups.

Lastly, what does the learner behaviour tell us about the acquisition and storage of multi-word verbs and lexical chunks in general?

In sum, the ultimate aim in this work is to provide an exhaustive account of the way Turkish learners of English at intermediate to advanced level of proficiency use multi-word verbs with the hope of throwing some light on how multi-word verbs are acquired and, in turn, of advancing the existing knowledge of the nature of vocabulary learning in a foreign language. The answers provided in this study will, in any case, not answer all of the questions raised here. However, it is hoped that the study

will lay a foundation from which future answers may be discovered. In addition, the results of this study can provide valuable input for those working in the field of English Language Teaching (ELT), materials development, and testing and can provide important information for future empirical studies involving language learners. Knowing why and how a specific learner group makes use of multi-word verbs and what kinds of problems they have in their acquisition may be useful in helping teachers to choose appropriate materials and tasks that would allow learners to address potential difficulties.

The data extraction yielded 6,129 tokens in the learner corpus. Due to this high number, it is impossible to discuss and present all instances of multi-word verb usage found in the data. To avoid strain on a reader of the study while still illustrating the main points, only representative examples from each group of multi-word verbs will be analyzed.

1.3 English learning and teaching in Turkey

It seems necessary to comment on ELT in Turkey since the learning environment has a significant impact on language development and, in turn, on the proficiency level attained by the learner. The following exploration is intended to give a brief account of potentially influential external and environmental factors for both success and failure in foreign language learning, including the status of English, the foreign language teaching traditions, and the amount and nature of English input in Turkey.

Notably, English is largely learned through formal schooling as a foreign language and used as such in Turkey. That is, regarding the use and status of English, Turkey belongs to the “expanding circle countries” (Kachru 1992). Having no official status, English in the expanding circle countries, as pointed out by Bamgbose (2003: 421), owes its enormous prestige mainly to its instrumental value. The instrumental function of English in Turkey predominates in three domains: trade, tourism, and education (Inal & Özdemir 2015: 135; see Arik (2020) for an overview of functions English serves in Turkey). The prominent role English plays in the first two domains combined with its symbolism of modernization and eliticism have positioned it “as a *sine quo non* for both economic and interpersonal purposes”, resulting in an ever-growing demand for English language education (Inal & Özdemir 2015: 135f; italics in original). The Turkish national education system has thus reshaped its foreign language policy over the years according to these new demands.⁵

By way of the educational reform of 1997, which led to the extension of the obligatory primary education in Turkey from five to eight years, English started to be taught from an earlier age in state (i.e. public) schools as a compulsory subject. English was offered for two hours a week during fourth and fifth grades, followed by four hours a week

⁵ See, for instance, Kırkgöz (2009, 2007) for a fuller account of the policy changes in ELT in the Turkish education system; see Inal and Özdemir (2015) for the resultant restructuring in teacher language education programs.

from grades six through eight (Bayyurt 2012: 305). In 2012, the obligatory education requirement was raised to 12 years (4+4+4 model, i.e. four years of primary, secondary and high school), and English has since been offered from second grade on at state schools (two hours a week for second through fourth grades). In secondary school, English is offered three hours a week during fifth and sixth grades and four hours a week during seventh and eighth grades. In high school (grades 9 through 12), the number of the English hours offered ranges between two to five hours a week depending on the type of school. Some high schools offer one year of intensive English (20 hours per week) as a preparatory program to their students in their first years. In Turkey, there are private schools for all levels of education, and they offer English earlier and more intensively than state schools. The number of private English-medium schools with English preparatory programs has been increasing since the 1980s (see Inal and Özdemir 2015: 138). Owing to the differences in the amount of time dedicated to English learning in the curriculum, the quality and range of materials provided, the methods applied, and the qualification of teachers, the role of English differs to a great extent across types of schools (Doğançay-Aktuna 1998: 31). These factors, in turn, create great discrepancies in the quality and extent of instruction available to the learner in different schools. In highly competitive private secondary schools, classes are taught in a foreign language, which is in most cases English, and, accordingly, those who go to such private schools are more likely to have a higher proficiency in English. Contrarily, in the less competitive state-owned

schools, the language competence of students does not usually develop beyond the basics, mostly as a result of crowded classes and a lack of quality materials and qualified EFL teachers. Many graduates of public high schools can only be categorized as false beginners, even after English language instruction for three to six hours a week throughout high school (Doğançay-Aktuna 1998: 32). According to Education First English Proficiency Index (2020), for instance, Turkey has a low English proficiency (69th among 100 countries, 33th among 34 European countries; see also Arik (2020: 519)).

According to Işık (2008), there are two main factors behind the general low proficiency level attained in foreign languages in Turkey: the application of ever-existing traditional methods of foreign language teaching and learning and the defects in foreign language planning. Related to the first problem, Kilimci and Can (2009: 209) similarly state that the teaching of English in Turkey was traditionally based on “a teacher-centered transmission model” and “[t]he predominant method employed was grammar translation with a focus on grammar and vocabulary at the expense of communication.” Teachers may find teaching in a ‘translational’ way (that is, through the learners’ shared L1) as much easier and more efficient than teaching through a language which they have not yet mastered. In Richards and Rodger’s (2014: 7) view, Grammar-Translation Method’s persistence “may be due to (a) the limited command of spoken English of language teachers, (b) the fact that this was the method their teachers used, (c) it gives teachers a sense of

control and authority in the classroom, and (d) it works well in large classes.”

With regard to the defects in foreign language planning, Işık (2008) maintains that foreign language planning and policies must be based on empirical data and explains the deficit seen in the language planning process in Turkey as follows:

Although foreign language education is much emphasized and supported by funds, the decisions taken are not based on scientific data. The decisions mostly depend on the political powers’ opinions and bureaucrats’ personal views. The Board of Education and Discipline in charge of foreign language policy and planning is not in a condition to undertake this duty. The members of this council are recruited on the basis of assignment. The members are not selected from among academicians with academic reputation all over the country, with adequate knowledge and experience, and according to academic criteria. In this respect, whether the member has the necessary qualities to direct policies and planning cannot be guaranteed. In addition, since the duties are based on assignment and dismissal, it is difficult to maintain continuity in terms of foreign language policies. Due to such deficiencies, a foreign language education policy which depends on the country’s aims and realities has not been formed as of yet (Işık 2008: 21; translation from the original by Bayyurt 2012: 34)

In a similar vein, Rakıcıoğlu-Söylemez (2016: 124) points out that foreign language education systems in Turkey are not supervised efficiently so as to provide quality assurance because there is no supervisory organization at the national level to undertake this mission (see

Rakıcıoğlu-Söylemez 2016 for an overview of the problems encountered in foreign language education in Turkey). Zok (2010: 10), on the other hand, argues that one of the problems with ELT in Turkey is that the overall teaching quality has decreased over the years as a result of the implementation of alternative, shorter methods of teacher training, such as ‘distance education’, so as to meet the increasing demand for English teachers (see Part 1 in the edited volume by Bayyurt and Bektaş-Çetinkaya (2012) on the social and historical account of English language teacher education in Turkey). All these observations and statements are equally plausible and valuable as an explanation for the general low proficiency attained by Turkish EFL learners.

From the 1990s onwards, there have been major attempts to revise the ELT syllabi to encourage a greater focus on communication and SL methodology and to move instruction away from a focus on written language and traditional grammar-based FL methods (see Sariçoban and Sariçoban 2012: 33f. for a number of influences on Turkish education). It is, however, hard to judge whether SL methodologies have taken hold in the teaching of English in Turkey. Reporting the fact that the Council of Europe has been encouraging the implementation of Communicative Language Teaching (CLT) in Turkey, Sariçoban and Sariçoban (2012: 38) observe that the revised ELT curriculum, albeit more comprehensive and elaborate in many aspects, has not been prescriptive in the implementation of new decisions regarding the integration of CLT into the curriculum. The researchers further claim, “it would not be wrong to

state that, today, in most foreign language classrooms, the methods suggested by the imported foreign course books are made use of, combined with the intuitive decisions of teachers.” The findings of a case study conducted by Kırkgöz (2006), which aimed at exploring what had already been attained since the 1997 educational reform with regard to classroom practices and use of methodology of teachers in state-owned primary schools, reflect these observations and claims. The study has revealed that CLT did not seem to have the expected impact on the ELT teachers’ classroom practices since most classroom activities were based on traditional methods of teaching – the proposed communicative learning environment was not created by most teachers. Other problems at the instructional level identified by the study were: large class sizes, making it difficult to implement the syllabus effectively, insufficient teaching time allocated to cover the curriculum, inadequate resources in the classrooms, and textbooks that did not support the proposed communicative teaching methodology (Kırkgöz 2007: 223; 2009: 679). Thus, one could conclude that at the level of a whole teaching culture, a specific country, or academic field and in the practices of individual teachers, one or the other tradition can generally be recognized as dominant regardless of the recent attempts and advances.

From my personal experience, there is one other thing worth a brief mention here since it is related to the learners at hand – the imbalance between the four language skills attained in the Turkish EFL context. In Turkey, after the completion of high school, all graduates have to take a national written exam called *Üniversite Giriş Sınavı* [University

Entrance Exam] in order to be admitted to a university. Those who graduated from language departments of high schools and who would like to pursue their studies in a field related to foreign language education need to take a second written language test in addition to the University Entrance Exam– the so-called *Yabancı Dil Sınavı* (YDS), which consists of multiple-choice questions. The format of this test has become known over the years and has a considerable effect on the way English is taught, especially to those aiming to take this language test – the ones who have foreign languages as their main subject in high school. Getting a place at a university to study in English Language Teaching Programs is quite competitive, which results in an exam-orientated teaching at high schools in Turkey. In most schools, foreign languages are taught in such a way that the speaking, listening, and even writing skills are neglected since these skills are not tested in the YDS. Consequently, students are mainly trained to answer multiple-choice questions focusing largely on grammar knowledge and on translating sentences between two languages and less so on lexical knowledge. Given the lack of focus on the speaking skill, the fact that Turkish learners of English “are not reaching a level in which conversation in English with any measure of confidence is possible” (Zok 2010: 11) should not come as a surprise. The weakness in their writing skill can also be easily spotted (e.g., Kıray & Kömürcü 2014).

Other than the formal instructional input the learners receive, they rarely associate or communicate with native speakers of English outside of the classroom. They are, of course, exposed to the English language

through the Internet, English movies, television programmes, music, and similar other means, but the learners' exposure to authentic English on a daily basis is very limited compared to other countries where English is taught as a FL, e.g. Sweden or Norway. Although radio stations widely broadcast English-language popular music and there are a handful of specialty TV channels that broadcast movies and television series with Turkish subtitles (e.g. CNBC-e), broadcasting is almost exclusively in Turkish (Inal & Özdemir 2015: 136); foreign films, television series, and TV shows on popular TV channels are always dubbed into Turkish (see also Arik (2020: 520) for the presence of English in Turkish media). Accordingly, the main language of print journalism is Turkish with few exceptions – to my knowledge, there are only three daily newspapers (namely, *Daily Sabah*, *Hürriyet Daily News* and *Aydınlık*) printed in English. Moreover, unlike in most European countries where it is common for students to spend some time abroad in the TL environment either as part of their study program or during holidays for leisure purposes, it is very rare for the learners in Turkey to go abroad. This fact can indeed be verified by the number of students involved in this study who reported having experience of living in an English-speaking country: only 4 out of 278. Therefore, it would not be inaccurate to state that unless the learners themselves seek opportunities for further input and chances for interaction in the target language, for the majority of the learners in Turkey, English input is largely confined to what is provided in the instructional setting and in the teaching materials.

Given the inadequate natural exposure to the target language and the problems at the instructional level, it is no wonder that despite the prominent role English has gained in the Turkish education system and ongoing improvements in the infrastructure of ELT in Turkey, only 17 percent of the Turkish population have been reported to speak English well enough to be able to have a conversation (Eurobarometer Report of European Commission 2006: 13). As a conclusion, it can be stated that English learning in the expanding circle countries has different levels of prioritization and varied methodologies of teaching, resulting in diverse language levels attained by the learners – and there is apparently still much more to undertake from an instructional point of view at all levels of ELT education in Turkey to increase the lower general English language proficiency of learners.

1.4 The structure of thesis

This study is divided into seven chapters. Following this introductory chapter, the second chapter focuses first on lexicon and then on multi-word units in general terms and in relation to native speakers and language learners. Chapter two also reviews the role (and presumed difficulty) of multi-word units in language learning and use and discusses the importance of these units for foreign and second language learners. A brief literature review on learner-related research in this area is also provided. The contrastive problems English multi-word verbs present to foreign language learners in general—and to Turkish learners in particular—are addressed, in turn justifying the choice of multi-word verbs

as a subject of investigation. The third chapter is concerned with the notion of multi-word verbs and outlines the different multi-word verb categories investigated in the study. Chapter four is devoted to the methodological aspects of the present study. The computer tools relevant for the analysis are described; furthermore, the corpora investigated are presented in detail. The definition and classification in respect to multi-word verbs are clarified, and the methods of data extraction are specified. The chapter concludes with a clarification of the terminology used. Chapters five and six present results from the quantitative and qualitative analyses, respectively. The final chapter sums up the findings and presents their implications for lexical learning process in L2 in general terms and with respect to vocabulary instruction. Acknowledging the limitations of the present study, the chapter concludes with proposals for further research.

Chapter 2: Lexicon

Words are the leaves of the tree of language, of which, if some fall away, a new succession takes their place.
(John French)

2.1 The role of the lexicon in linguistics and language teaching

The lexicon – though central to language acquisition and use (e.g., Laufer 2002; Lewis 2001; Gass 1988) as it is the essential basis to communication – has been in and out of fashion in the long and diversified history of language teaching. Regardless of the changes in theoretical insights as well as in applied approaches within linguistics and correspondingly second language acquisition (SLA) research, one fact remained consistent for a long time in the history of language teaching: the study of language learning was limited to the analysis of grammar learning (Bogaards 1996: 367) since lexicon did not play a major role in the earlier foreign or second language teaching methods or approaches. Although the importance of the lexicon in the process of learning a foreign or second language is now widely acknowledged, this has not always been the case. Despite its obvious importance in learning

foreign language, the lexicon remained, in Lewis's (2002) words, "something of a Cinderella".⁶

The two aspects of language – namely, lexicon and grammar – have been kept apart traditionally: it used to be thought that these two aspects could be studied separately "as if they belong to two different worlds" (Bogaards 1996: 358). Lexicon was neglected for a long time in SLA research, largely due to structuralism dominating the study of languages with its emphasis on phonology and syntactic patterning. With the shift to transformational linguistics in the late 1950s and 1960s and Chomsky's influence in theoretical linguistics, syntax increasingly included the regular aspects of language (Atkins et al. 1994: 18) and the view of lexicon being "somewhat peripheral, an irritating irregularity in an otherwise ordered grammar" (Carter & McCarthy 1988: 41) was strengthened. In turn, the common view that became prevalent in linguistics led to language being considered as a structured system and the preoccupation of linguists was exclusively with the aspects of language that could be studied reliably, i.e. scientifically – namely, phonology and grammar (Wilkins 1972: 109), what Meara (1984: 230), referred to as "the more manageable parts of a language". Grammar as a 'closed' system was considered "systematic and regular and therefore analysable as a set of generalisations and rules" (Waibel 2007: 1). Lexicon, on the other hand, as an infinite, 'open' system, where new items can be added quite freely

⁶ "Lexis is the core or heart of language but in language teaching has always been the Cinderella" (Lewis 2002: 89).

at any time so as to adjust itself to the changing conditions and to meet the evolving needs of human communication, was believed to be “random, chaotic and not organisable in terms of a rule-governed system analogous to syntax” (Waibel 2007: 1). As Bloomfield (1933: 274) expressed it, lexicon was just “an appendix of the grammar, a list of basic irregularities”. As a result, lexicon came to be regarded by linguists as “a boundless chaos” (Carter & McCarthy 1988: 38), “an inherently messy part of our linguistic competence” (Meara 1984: 230).

As the common concern was grammar, mastery of a foreign language was equated with learning the structures of the language. As observed by Lewis (2002: 11), “[t]he tacit assumption was that once you had mastered the system, you could generate any sentence you wished”. The importance of the lexicon was thus relegated to a secondary level in the process of teaching foreign languages.⁷ While learners were still acquiring grammar, the range of vocabulary taught was deliberately kept restricted to the level essential to exemplify target structural patterns so that “the learner’s power of acquisition” could be focused on grammar – what was considered to be the most important part of learning a language at the time. Attempts at learning new words before having mastered the basic grammar structures used to be regarded as a hindrance for learners since it would distract them from observing and using the syntactic patterns of the target language (Judd 1978: 72). Fries (1945), whose ideas had a great influence in the field of L2 teaching for a long

⁷ See Meara (1984: 229f.) for the reasoning behind the ‘unjustified neglect’ of lexicon.

time – until the mid 1980s – asserted that the main problem of learning a new language was not learning its vocabulary but mastering its sound system and its grammatical structure, i.e. the matters that native speakers acquire as unconscious habits as children.⁸ In a rather firm way, he expressed his belief by stating that “[s]uch students, with fluency in vocabulary but with no basic control of either the sound system or the structure, are almost without exception hopeless so far as ever achieving a satisfactory control of English is concerned” (Fries 1945: 3). Acknowledging the limitation of vocabulary expansion during the very early stages of foreign language learning, Twaddell (1973) expressed a similar view to Fries (1945). He suggested that the initial focus should be on habit formation of the fundamentals of a foreign language – its pronunciation and basic grammatical patterns – and only after these habits have developed vocabulary should be expanded. Thus, it was then believed that what all learners needed at first was “sufficient vocabulary” to practice and master the sound system and the syntactic

⁸ Fries (1945: 2) stated that “[m]astery” of a language must mean something other than knowing “all the words” of the language” as there are always areas of expertise whose language include special terms which are not known by all native speakers. He went on to remark that “[i]t is true, however, that whenever we think of language and language learning we usually think of mastering the vocabulary – of learning the “words” (ibid.). As an explanation for this common view the researcher put forward our experience with our mother tongue; we master the sound system and the structural devices of our mother tongue so early in our life that they become unconscious habits in our early childhood; as a result, we cannot remember the learning process (ibid.). However, learning the lexicon of our mother tongue is “a constantly developing mastery” and we are often very conscious about it (Fries 1945: 2).

structure of the target language (Fries 1945: 3).⁹ This belief has been evidenced by Sinclair and Renouf's (1988) observations of the traditional syllabi: in most of them a small number of common words were used, without an independent content, to indicate the structural frameworks. This observation has led Sinclair and Renouf (1988: 141) to conclude that "[p]ride of place is given to the grammar, and the vocabulary is clearly secondary."

Lexicography, however, started to draw attention as early as the 1960s, especially after the tendency to look at lexis in a syntagmatic way as opposed to a paradigmatic way (e.g., Householder & Saporta 1962; Sinclair 1966; Halliday 1966). J. R. Firth, who brought the term *collocation* into prominence in lexical studies, is regarded as the pioneer of this tendency. According to Firth (1957), one central part of the meaning of a given word is the words with which it typically occurs, i.e. its collocations. Firth (1957: 196) illustrated this claim by stating that "[o]ne of the meanings of *night* is its collocability with *dark*, and of *dark*, of course, collocation with *night*".¹⁰ Pointing out the restrictive nature of collocability (1957: 195), Firth suggested not to treat words as if they

⁹ As for the selection of these early lexical items, Twaddell (1973: 63) states the following comments: "the chief virtue of a vocabulary item is its versatility and its usefulness in a variety of grammatical constructions, and as an additional advantage its usefulness in practice of the characteristic phonetic features of the FL. In the beginning approaches to FL learning, one vocabulary item which performs a number of such pedagogical purposes is better than several which together would perform the same purposes."

¹⁰ See Firth (1968: 20ff.) for over some thirty formal types of collocation of the verb *get*, which he uses as an example to make his point.

had “isolate” meaning and occurred and could be employed in free distribution (1968: 18). Halliday (1966) and Sinclair (1966) took Firth’s theory of meaning one step further and highlighted the importance of lexical collocations in an integrated lexical theory, e.g. adjective-noun and verb-noun collocations. In his thought-provoking paper, Halliday (1966) defended the setting-up of lexis as an independent part of language forms: he suggested devising methods appropriate to the description of lexical patterns in a language in view of a lexical theory that was “complementary to, but not part of, grammatical theory” (Halliday 1966: 148).¹¹ Sinclair (1966), in his paper entitled “Beginning the Study of Lexis”, asserted the necessity of looking at the internal patterns of language from two “interpenetrating” aspects – lexis and grammar – so as to describe language more accurately since lexis, describing the tendencies of items to collocate with one another, provides facts about languages that cannot be obtained by grammatical analysis, in which the patterns of language are analyzed by a number of different available choices (Sinclair 1966: 411). Sinclair explained this as follows: “One lexical item is not chosen rather than another, lexical items do not contrast with each other in the same sense as grammatical classes contrast. There are virtually no impossible collocations, but some are much more likely than others” (Sinclair 1966: 411).

¹¹ See also Halliday (1961) where he removed the distinction between grammar and lexis seen in traditional approaches to language and incorporated them into one unified resource for the expression of meaning – he considered lexis as the “most delicate grammar” (p. 267).

With the publication of Chomsky's 'Remarks on Nominalization' in 1970, in which nominalizations were captured in both a syntactic and a lexical way, recognizing the existence of regular processes in the lexicon, the lexicon became prominent in generative linguistics (cf. Bogaards 1996: 362f). Along with the aforementioned advent of looking at the lexicon in a syntagmatic way in 1960s and 1970s, which instigated a renewed interest in the study of the lexicon, dissatisfaction with the secondary status of vocabulary in foreign language teaching in favor of syntax started to be voiced by some. Wilkins (1972), for instance, who was one of the first to propose an increased role for vocabulary, pointed out the equivalent significance of grammar and vocabulary: "The obvious fact is that to communicate at all seriously and adequately through the language a command of both grammar and vocabulary is necessary" (Wilkins 1972: 110). Learning how to construct grammatical sentences but not having the vocabulary to convey what one wishes to would yield nothing that is of very practical use to the learner; one would be literally "at a loss for words", but learning only words and no structure would be equally useless (Wilkins 1972: 110). "The fact is that while without grammar very little can be conveyed, without vocabulary *nothing* can be conveyed" (Wilkins 1972: 111) because "[w]hat we normally think of as 'vocabulary items' – nouns, verbs and adjectives – do indeed contain more information than is carried by grammatical elements" (Wilkins 1972: 111). Besides, it is not possible to teach these two aspects of language separately, as Sinclair and Renouf (1988: 143) expressed in their discussion of mixed syllabi:

it is almost impossible to teach grammar without in passing teaching some vocabulary. Vocabulary fleshes out the structures, introduces variety and promotes practice of the structure in question. [...] In the same way, a class devoted to expanding its vocabulary will not be able to avoid syntax, assuming that it is not exclusively committing lists to memory.

As syntactic errors were found to be a natural development phenomenon in second language learning (Selinker 1972; Dulay & Burt 1973, 1974; Corder 1967), some criticism against the traditional applications of vocabulary teaching started to be voiced towards the end of 1970s. Judd (1978: 75), for instance, who was in favor of massive vocabulary expansion at very early stages, criticized that vocabulary use was not seen and taught as a vital skill in its own right, but rather as a “part of wider areas of language learning such as reading and listening comprehension” (ibid.: 72) and suggested starting to work on ESL learners’ vocabulary enrichment instead of devoting excessive time to improve the syntactic performance of the learners. He justified his suggestion by stating: “No one can deny the assertion that one major aspect of fluency in a language is control over the lexicon” (Judd 1978: 71).

By the end of 1970s, the importance of vocabulary in language teaching had been affirmed, following the changes in the field: there was a shift of emphasis to communicative competence and accomplishment of real-world tasks, a focus on functional competence and use of the target language rather than formal correctness. Advocacy of communication practice was to be seen even at the earliest stages of learning since communication started to be regarded as the primary goal of language. Dik

(1978: 5), for instance, regarded language as an instrument of social interaction in his Functional Grammar and argued that the main function of language was communication. Along with the increased interest in communicative language teaching (e.g., Savignon 1972, 1983; Widdowson 1978; Munby 1978; Brumfit & Johnson 1979; Candlin 1981; Canale & Swain 1980; Canale 1983), the learner was placed on the center stage and the lexicon started to be considered as a resource for the needs of the learners in communication (see e.g., Blum and Levenston 1978 and 1980 for lexical strategies).¹² Eventually, even the importance of vocabulary over grammar started to be expressed by some. Bogaards (1996), for instance, viewing the sentences of language as buildings, which necessitate elements used in the construction (words) and the rules to determine how this happens (grammar rules), stresses the importance of vocabulary over grammar as follows:

Rules and elements are interdependent. This does not mean, however, that they are equally important. Whereas one may play with marbles without having any rules at all, it is totally impossible to do anything with only rules and no marbles. Or, to go back to the context of building, it may be possible to protect oneself from wind and rain just by piling up building material without having a clear plan as to how to do this, but if one only has the structure of a house, or the knowledge of building principles, one will get wet (Bogaards 1996: 358).

¹² For a review of the developments in vocabulary teaching between 1930s and 1980s, see Carter and McCarthy (1988: 39-57).

According to Bogaards (1996), language consists of lexical elements that call up grammatical structures, not, as commonly viewed, of grammatical constructions into which lexical elements can be inserted. In his words: “syntactic structures are not autonomous constructions with their own status, independent of the lexicon” (ibid.: 359). He then goes on to remark that rules and structures are indeed “extremely sensitive to any individual lexical meaning of the words used” (Bogaards 1996: 361).

It should, however, be noted that although the lexicon could not be considered as the “neglected component” of second language acquisition towards the 1980s anymore, Gass (1988) observed that the majority of the lexical studies dealt only with descriptive aspects of the lexicon rather than concentrating on the establishment of a theory of the lexicon and that the lexicon was still only discussed “as a side issue rather than as *the* major area of research” (Gass 1988: 93). Observing the lack of lexical theory regarding the learning of a second or foreign language nearly a decade later, Appel (1996) explains the absence of the lexicon as an issue in the textbooks written on second language acquisition with the lack of information on the lexicon in L2 acquisition and impossibility to understand how words and their meanings are acquired in L2.¹³ These observations suggest that despite the rapid change in the status

¹³ See Bogaards (1996: 369ff) for his suggestions on the types of elements that could be contained in a lexical language acquisition theory.

of the lexicon, lexical items were still considered more or less individually in their capacity as building blocks in communication with the capacity to trigger certain grammatical constructions.

The interest in truly syntagmatic aspects of the lexicon took a flying leap during the 1980s, especially after Bolinger's remarks on the formulaic nature of language. Bolinger has drawn attention to the fact that much of actual language use is repetitive and much more based on lexical units of varying sizes than had been generally thought. In his own words: "the amount of language that comes ready made is vastly greater than supposed" (Bolinger 1971: xiv). Similarly, in a later piece of work he has stated that "our language does not expect us to build everything starting with lumber, nails and blueprint. Instead it provides us with an incredibly large number of prefabs, which have the magical property of persisting even when we knock some of them apart and put them together in unpredictable ways" (Bolinger 1979: 96 [1976: 1]) (see also Saussure 1916/1966: 177; Palmer 1938/1961: v; Hymes 1962/1968: 126) and has also pointed out that most of these prefabs are not completely but only partially fixed. By referring to the neurophysiological and psychological evidence in Ladefoged (1972), indicating that storing multi-word sequences are more economical compared to computing them during language use, Bolinger has further claimed that the "human mind is less remarkable for its creativity than for the fact that it remembers everything" (1979: 97 [1976: 2]). Fillmore (1979: 100) has also commented insightfully on the formulaic nature of language: "a large part of our ability to get along well in a language is our facility with

formulaic expressions.” Since then, many similar claims have been made – see most notably Pawley and Syder (1983) whose research findings accord closely with the statements of Bolinger and Fillmore. Pawley and Syder (1983: 215) report that “by far the largest part of the English speaker’s lexicon consists of complex lexical items” and that most of these complex lexical items are semi-productive (*ibid.*: 216f). In a similar vein, Nattinger (1988: 76) has characterized language use as “basically a ‘compositional’ process, one of ‘stitching together’ preassembled phrases into discourse”. Acknowledging this characterization of language, Sinclair (1991: 109-115) has asserted that the production of texts, i.e. language production, involves a constant alternation between two processes which he calls the ‘open choice principle’ and the ‘idiom principle’ – the former refers to the process in which each word is chosen from a large range of options to fill a certain slot in a sentence by satisfying ‘local restraints’ (i.e., grammaticality) whereas the latter to the use of preconstructed multi-word combinations, which is, according to Sinclair, the dominant mode of processing (these two modes of processing are explored further in Section 2.3.1). Not very long after, with the advances in computer technology along with the advent of electronic corpora, which has further motivated the study of lexicon and the search for lexical patterns, empirical support for the mainly intui-

tion-based claims about the formulaic nature of language was provided.¹⁴ The availability of computerized corpora has made it possible to explore vast amounts of authentic language data both qualitatively and quantitatively without difficulty and uncover the highly patterned nature of language use. One main finding of computerized corpus-based analysis of language has been the fuzzy boundary found between lexicon and grammar, which in turn resulted in the questioning of the traditional lexicon and grammar dichotomy. Words, as soon realized, are not combined freely with one another; they have their own characteristic combinations. Moreover, their meaning mostly depends on their company, as Firth had maintained. That is, words are not “self-contained units of meaning” which fill the grammatical slots in a sentence, as formerly viewed (Krishnamurthy 2003: 289).¹⁵ These facts about words have instigated a renewed interest in lexical studies, especially in collocations and phraseological items, earlier neglect of which may

¹⁴ The investigation of word combinations in natural language use dates back to the earliest studies of collocations by J. R. Firth (1957), who brought the notion of ‘collocation’ and ‘collocability’ into prominence to express the habitual co-occurrence of words. However, large-scale investigations have only become feasible with the arrival of the computer in the late 20th century.

¹⁵ The following comments from Wray (2002: 13) neatly summarize the matter: “Whereas it was previously possible to imagine that words combined fairly freely, their restrictions attributable to context and pragmatics, and to easily definable social signaling, it is now clear that, once you actually map out the patterns of distribution for words, no such piecemeal and superimposed explanation is possible. Words belong with other words not as an afterthought but at the most fundamental level.”

be attributable to the fact that they rest on the borderline between grammar and lexicon, which made their detection and elicitation difficult before the advent of the computer as a powerful research tool.

The availability of large corpora has had a shaping influence on our understanding of how language is used. Numerous studies based on large collections of authentic text from various sources have thus far shown us that natural language use, both written and spoken, largely consists of recurrent word sequences (e.g., Kjellmer 1991; Altenberg 1991, 1998; Altenberg & Eeg-Olofsson 1990; Renouf & Sinclair 1991; Sinclair 1991, 2004; Biber et al. 1999; Moon 1998; Erman & Warren 2000; Stubbs 2001; Cortes 2002; O’Keeffe et al. 2007; Aas 2011). Moreover, at all levels of language production, from complete utterances to smaller units such as words and phrases, only very few of these ready-made units are found to be semantically and grammatically completely ‘frozen’: Instead, a great number of them, as Altenberg (1998: 120f) put it, “occupy a position along the cline between fully lexicalized units and free constructions” which indicates the difficulty or impossibility of drawing a sharp line between lexicon and grammar (see also Peters 1983 and Pawley and Syder 1983). Thanks to the insights gained from corpus-based studies focusing on aspects of lexis-grammar co-selection, it is now known that the dichotomy between grammar and lexicon does not hold true for language use and it has been widely accepted that words have their own grammar. Thus, the two areas which have traditionally been kept separate, in linguistic analysis and theory as well

as in language teaching, are in fact inseparable and inextricably intertwined, i.e. lexicon and grammar (vocabulary and syntax) are interdependent. Römer (2009), for instance, demonstrates the interdependence of lexicon and grammar by presenting a case study on the use of the introductory *it* pattern in apprentice and expert academic writing. The lexicon has thus become viewed as a dynamic rather than as a static entity and recent years have seen an increasing interest in the search for rules and regularities in words. Correspondingly, an increasing number of current language models have abandoned the traditional distinction between lexicon and grammar, e.g., connectionist/emergentist models of language (e.g., Rumelhart and McClelland 1986; Seidenberg 1994; McClelland 2010; McClelland et al., 2010) and exemplar-based models of language (e.g., Beekhuizen et al., 2013; Bod 1998, 2009; Goldinger 1996; Pierrehumbert 2001). The progressive integration of grammar and lexicon finds its final expression in the language approach known as Construction Grammar (e.g., Fillmore et al. 1988; Goldberg 1995, 2006) in which language knowledge is viewed as having knowledge of ‘constructions’ which are pairings of form and meaning or function that range from morphemes to words and abstract syntactic frames (for extensive coverage of the latest ideas in Construction Grammar see Hoffman and Trousdale 2013).

The foregoing account of the role of lexicon in linguistics can be recapitulated and concluded as follows: After the first attention-shift from grammar to lexicon seen in 1960s, the focus of the last 15 years or so

has turned towards lexical groups, i.e. to multi-word units such as collocations, prefabs, chunks, etc. Whereas formerly they were treated as peripheral exceptions and linguistic rarities, a substantial body of research now exists to show the central role they play in language learning and use. As a concomitant of the increasing acknowledgement of their importance they have become a recognized aspect of vocabulary and pedagogy (e.g., Nattinger & DeCarrico 1992; Lewis 2000; Altenberg & Granger 2001; Willis 1990, 2003; Wray 2002; McCarthy & O'Dell 2005; Meunier 2012; Siyanova-Chanturia & Pellicer-Sánchez 2019). In the context of foreign or second language learning, multi-word units constitute a particularly interesting phenomenon since they are known to cause problems for learners. It is now widely accepted that it is necessary for any language learner to be able to produce not only intelligible and acceptable multi-word units for successful communication but also units preferred by native speakers in the target language.

2.2 The lexicon and the foreign language learner

As the lexicon is “the basis of accurate and fluent communication” (Rudzka-Ostyn 2003: v) and “lexical competence is at the heart of communicative competence” (Meara 1996: 35), the study of the lexicon remains to be of paramount importance in relation to foreign language acquisition. This importance has been repeatedly expressed by many scholars in the literature. Levelt, for instance, who asserts that the lexicon is the driving force in sentence production, points out that “grammatical and phonological encoding are mediated by lexical entries”

(1989: 181). This assumption, though in relation to speech production in L1, applies equally well to foreign languages. Accordingly, Saville-Troike (1984: 199) contends that “vocabulary knowledge is the most important area of second language L2 competence.” The following quotation from Hatch (1983) provides a good explanatory statement for the importance of the lexicon:

When our first goal is communication, when we have little of the new language at our command, it is the lexicon that is crucial. If we can but find the words, we know we can take care of our immediate needs. The words [...] will make basic communication possible. (Hatch 1983: 74).

Given the fact that there are tens of thousands of words in a language, the L2 learner is faced with a challenging and formidable task. Besides, only very few words have one single meaning. Fries (1945: 40) has asserted that “the actual number of clearly separable senses covered by most of the commonly used words in English is enormous” and that words usually have from fifteen to twenty meanings. There is empirical evidence revealing the difficulty learners have with polysemy. Regarding the meaning differentiation, in a study of lexical guessing Bensoussan and Laufer (1984), for instance, found that polysemes were the most frequent cause of incorrect guessing and, in turn, of incomprehension of words: Being familiar with one of the meanings of a polyseme, learners did not seek another meaning even though the meaning they knew did not make sense in the given context, e.g. *course* in the sense of *duration* was often interpreted as *dish*; *abstract* as opposed to *concrete* was interpreted as *summary*, etc. With regard to production, there is also

empirical evidence for the avoidance of what Levenston (1979) calls “unreasonable polysemy”. Levenston (1979: 152) refers to Kantor’s (1978) study in which English speaking learners of Hebrew were found to have acquired one meaning of a polyseme and could not bring themselves to use it in its other meanings; the Hebrew verb *lidchot*, for instance, which means both *postpone* and *reject*, were not used by the learners in the sense of *reject* since it just did not seem logical that a word could have two such “incompatible” meanings and, moreover, “even lead – with objects like ‘the proposed meeting’ – to most unfortunate ambiguities” (Levenston 1979: 152).¹⁶ As explained by Lennon (1996: 35), the learners “over-rely on their ideas of core meaning of polysemous verbs and may be derailed by treacherous translation equivalents in L1.”

In addition to polysemy, the fact that one meaning can be represented by different words is another source of difficulty for language learners. Accordingly, vocabulary learning, as pointed out by Laufer-Dvorkin (1991: 1), is not only an absolute necessity but it is also “never-ending”: “Long after the acquisition of phonology, morphology and syntax has been completed, the learner will still be encountering new words and expanding his lexicon” (ibid.). New words constantly emerge, meanings of words keep changing; they are widened or narrowed, become

¹⁶ As for the question of which meaning is preferred and which is rejected, Levenston states that the sequence and intensity of acquisition are decisive factors and adds that “[i]n controlled, formal language situations the first meaning encountered is also likely to be the most frequent” (Levenston 1979: 153).

restricted to a specific context, and new combinations arise. Correspondingly, lexicon is often considered the greatest source of problems for foreign language learners; choice of the appropriate word or expression in a given context is a real challenge, since “[t]he vocabulary of a language is sensitive to a wide range of co-textual and contextual considerations” (Pietilä et al. 2015: 2). The following quotation from Pietilä et al. (2015: 2) furnishes an excellent summary statement of the matter:

[...] words enter into a myriad of relationships: for instance, they can combine appropriately only with particular items in collocations or bundles, they enter into numerous cohesive chains, have particular currencies, may become fashionable or obsolete. Moreover, words may differ almost imperceptibly in shades of meaning, bear various connotations, and invoke distinct cultures. They may be charged with evaluative potential and radically change the tone or formality of a passage.

That is, knowing a word well involves more than being familiar with its spoken and written form and its most common (referential) meaning: word knowledge also involves knowing a word’s affective and pragmatic meaning, its structure and syntactic behavior, its synonyms, its relations with other words, its common collocations, and constraints on its use.¹⁷ Normally only after having learned some kind of basic or “ker-

¹⁷ These constraints include geographical restrictions, currency, age, gender restrictions and frequency (Nation 2006: 449).

nel” meaning of a word, the learner acquires other, secondary meanings,¹⁸ its associations, collocates and specific meaning(s) in different collocations and use of the word in fixed expressions (see Hemchua and Schmitt 2006). That means, as Appel (1996: 387) has rightly pointed out, in addition to being a quantitative process, lexical development also has qualitative aspects. Therefore, as suggested by Faerch, Haastrup and Phillipson (1984: 100), “we should think of vocabulary knowledge as a continuum between ability to make sense of a word and ability to activate the word automatically for productive purposes”, rather than making the overly simplified distinction between ‘active’ and ‘passive’ vocabulary (see also Melka-Teichroew 1982). Having a productive knowledge of a word then means being able to use it appropriately in a given context and register in right combinations (for a summary of the main characteristics of knowing a word in a second or foreign language, see, for instance, Carter 2012: 226f.; Ringbom 1987: 35-38; Laufer-Dvorkin 1991: 5-7).

Word learning in another language is in most cases not an easy task. Some difficulties result from the words themselves due to the intrinsic difficulty related to their form, meaning and usage; pronounceability, inflectional and derivational complexity, specificity, idiomaticity, register restrictions and polysemy are all contributing factors in the ease

¹⁸ Gass (1988: 101), for instance, claims that the learner first learns the prototypical or the central, core meaning of words and only after that more peripheral meanings (see Takahashi et al.’s (1988) study cited in Gass 1988: 102).

and difficulty of learning a word (Laufer 1990a; 1997). It is not uncommon that learners make unsuccessful lexical choices or that they confuse phonologically similar words in their target language (see e.g., Meara 1984) or homonyms. As opposed to native speakers, who have sorted out a complex system of similarities and distinctions among words, learners' word knowledge may not be complete, i.e., some features of a word may have been mastered by the learner but not all. As asserted by Laufer (1990a: 295), "the plurality of features to be learnt increases the probability of a word being problematic and therefore only partially learnt, since problems can arise from one or more of the areas." Whereas some features of words present a heavier 'learning burden' to the learner,¹⁹ partial word knowledge can mostly be attributed to an inadequate exposure to words, i.e. target language. During the acquisition of our first language, we have the luxury of encountering words numerous times in the course of our early development and always in a variety of meaningful contexts and these consecutive encounters strengthen our understanding of what makes a word special and help us to "narrow down its range of operations in the language" (Martin 1984: 130). As Martin explains further, "[w]e slowly develop a "feel" for what kinds of linguistic and sociolinguistic neighborhoods it [a word] tends to inhabit. Only gradually do we come to appreciate the subtle distinctions between a given word and its nearest cousins in the lexicon" (ibid.). This luxury we have during our first language acquisition is not enjoyed

¹⁹ See Lado (1972: 281ff.) for the seven patterns of difficulty in vocabulary.

by foreign language learners; the input is rather limited, words are sometimes presented out of context and sometimes isolated. In addition, incongruency between words in L1 and the target language adds to the difficulty of the word learning process.²⁰ As pointed out by Wilkins (1972: 130), “[v]ocabulary learning is learning to discriminate progressively the meanings of words in the target language from the meanings of their nearest ‘equivalents’ in the mother-tongue” and achieving this discrimination is not possible without massive exposure to the target language (Wilkins 1972: 132; see also the Continuous Lexical Disambiguation Model proposed by Sonaiya 1991). Empirical research has already shown that the mother tongue of a learner influences word learning in the process of L2 acquisition (e.g., Zughoul 1991; Ard and Homburg 1992; Harley & King 1989).²¹

Learners often consider word learning as their main task in trying to master the target language (Appel 1996: 382) since they often experience difficulty in expressing themselves precisely. As shown by Cumming (1990: 491ff), learners, while writing, often put a considerable amount of mental effort into searching for the ‘right’ word or phrase to express an idea, generating and assessing possible alternative linguistic choices. The hurdle of word search faced by learners often results in a

²⁰ See Laufer-Dvorkin (1991: 7-26) for a summary of intralexical and interlexical factors that affect learning of words.

²¹ cf. Hemchua and Schmitt (2006). In their study analyzing Thai university students’ English compositions for lexical errors, the identified sources of errors were mainly from L2 intrinsic difficulty rather than L1 transfer.

continuous desire for vocabulary expansion and in turn in an excessive amount of time spent on the elaboration of L2 vocabulary (see Manchón et al. 2007). As Ellis (1985: 104) rightly points out “[v]ocabulary is the area that learners seem most conscious of”. For instance, a respondent in a study conducted by Leki and Carson (1994), who investigated ESL students’ perceptions of the writing instruction they received, expressed the frustration with the imprecision with which he was to express himself as follows:

I wish I had learned more words and had increased my vocabulary significantly. Sometimes I simply run out of words necessary to express what I am actually thinking. I usually find a substitute word, but often times that word would only be semi-fitting with my thoughts (Leki & Carson 1994: 91).

In the same study, the importance of vocabulary from a learner’s point of view was mentioned as a writing need more frequently compared to other language skills – 38% (Leki & Carson 1994: 90). Similarly, questioned about the most important aspect of EFL writing, most of the participants in Porte’s (1997) study ranked the “range of vocabulary” to be the most significant to them. Furthermore, the fact that learners experience greater difficulties with vocabulary rather than the grammar of the target language has already been confirmed by several studies (e.g., Grauberg 1971; Schlue 1977; Dechert 1984; Lennon 1991a; Erkaya 2012; Darus & Subramaniam 2009; Porte 1997; Götz 2015; see also

Meara 1984).²² In past research on communication strategies, a great proportion of the problems learners encountered during language production has been found to be lexis-related; the main obstacle to the encoding process is insufficient L2 lexical knowledge (e.g., Kellerman 1991; Poulisse 1993; Yule 1997; Hendrickson 1976; cf. also Dörnyei & Kormos 1998).²³ Besides, as pointed out by Meara (1980: 221), learners themselves openly acknowledge that they are faced with difficulties when it comes to vocabulary and as soon as they have got over the initial phase of acquiring the target language, most learners consider vocabulary learning “their greatest single source of problems.” If pronunciation is excluded, what differentiates an advanced foreign speaker from a native speaker is mostly the choice of words as well as variety of words in a given context. As stated by Hasselgren (1994: 250), “[t]he differences between a learner’s lexis and a native speaker’s will often give him away as a ‘foreigner’ even if his language is ‘impeccable’.” In most cases, choice of an inappropriate word or of an expression results in the learner’s language sounding unnatural, nonnative-like. There is, unfortunately, no linear relationship between proficiency and the number of lexical errors produced. Martin (1984: 130) claims that “[a]s the

²² Meara (1984: 229) refers to a study by Blaas (1982) in which lexical errors were found to outnumber grammatical errors by three or four to one.

²³ Cf. Levelt’s (1989, 1993, 1995) model of speech production, which was originally developed for monolingual communication but has been widely used as a theoretical framework in L2 production research; lemma retrieval is attributed a central role in the speech process. Poulisse (1993), for instance, based her model of L2 verbal communication on Levelt’s (1989) model.

fluency of advanced language students increases, so too does the number of vocabulary errors they generate, both in speaking and writing” (see also Arabski 1979) since with the increasing proficiency language learners attempt to express more sophisticated ideas, which in turn requires not only the knowledge of more advanced grammar but also the use of more sophisticated vocabulary. Martin’s claim tallies well with Lennon’s (1991a) results; having found nearly half of the errors of advanced learners to be lexical or to have a lexical element to them, the researcher has concluded that the advanced learner’s main problems appear to be with lexis and preposition choice (Lennon 1991a: 43). Arabski (1979), who similarly has found the lexical errors to be the most frequent types of errors in the writings of beginner and intermediate level Polish learners, in advanced writings, the 2nd most, states that lexical errors cover different types of errors on each level: On the advanced level, “37% of lexical errors consist in a lack of semantic precision (synonymy, paraphrase, etc.) and as many as 49% of errors are those involving the application of primary counterpart” (Arabski 1979: 99f) – ‘a primary counterpart’ is defined by Arabski (1979: 48) as “an item which in IL represents the whole group (list) of L2 translation equivalents” and according to him, it causes underdifferentiation (ibid.). Regarding the lexical errors of advanced learners, Martin (1984: 130), on the other hand, states that they display one or more types of “dissonance” between a given word and its context; namely, on the stylistic, syntactic, collocational, and semantic levels.

The importance of lexical knowledge is also emphasized by the fact that grammatical as well as phonological errors could be more easily ignored since it is still possible to interpret the intended message by means of other contextual information (Gass 1988: 92). As grammar is the “essentially predictable” component of language (Rifkin & Roberts 1995: 530), grammar errors have been found not to impede communication to a significant degree (Guntermann 1978). Lexical errors, on the other hand, can result in misunderstandings of the intended message or at least in an increase in the burden of interpretation of the message (Hemchua & Schmitt 2006: 3; Gass 1988: 92) since the semantic and communicative load of an utterance is determined by the lexicon (Waibel 2007: 2; Wilkins 1972: 111; see Johansson 1978: 26-28). Accordingly, lexical errors are judged more negatively. In a study investigating the reactions of both native and non-native professors to written productions of one Chinese and one Korean student in terms of content and language, Santos (1988) found content to receive lower ratings than language but lexical errors to be rated as the most serious errors. Santos (1988: 84) explained his findings by stating that “[i]t is precisely with this type of error [lexical error] that language impinges directly on content; when the wrong word is used, the meaning is very likely to be obscured.” Accordingly, concerned with the communicative effect of errors, Johansson (1978) found lexical errors to cause more problems of interpretation and to induce a higher degree of irritation in native speakers compared to grammatical errors (pp. 65-72). Johansson (1978: 71) explains this finding, which he expected, as follows:

By definition, grammar deals with the rule-governed, predictable aspects of language, the lexicon with the idiosyncratic, unpredictable aspects. A grammatical choice is a part of a closed system, a lexical choice involves a selection from among members of an open class. In interpreting sentences with lexical errors the receiver is therefore faced with a more difficult task, since the possibilities of correction are much more varied and his knowledge of the language offers less help in predicting the correct form.

Khalil's (1985) findings correspond to Johansson's (1978) results: semantically deviant utterances of Arab EFL learners were considered less intelligible and were interpreted with less accuracy by native speakers compared to utterances with grammatical deviances. Politzer's (1978) study, results of which point in the same direction, deserves special attention as vocabulary errors are compared with errors recurring in the realm of phonology, confusion of case endings, verb morphology, gender confusion, and word order: Native speakers of German graded vocabulary errors of English speakers of German as the most serious (Poltzer 1978: 257; see also Olsson 1972, 1973 and Guntermann 1978).²⁴ On the basis of these findings, one can then conclude that, as has already been pointed out by Page (1990: 170), grammatical correctness is not always needed for accurate communication since ungram-

²⁴ Cf. Albrechtsen, Henriksen and Færch (1980). In their study on native speakers' evaluations of spoken interlanguage of Danish learners of English in terms of lexical, syntactical and morphological errors, they have shown that lexical errors do not necessarily prevent successful (efficacious) communication if semantically related words are used or if there is enough "contextual support" to help for the correct interpretation of the erroneously used words.

matical utterances do not always relay messages less precisely (see Bogaards 1996: 358f.). As suggested by Chafe (1980: 170), though in the context of hesitational phenomena in native speech, since the main aim of the speaker is to adequately verbalize his thoughts, it is quite unlikely for him to be interested in grammaticality, of course unless there is some special reason to pay attention to it, but there is usually not. In a similar vein, Di Pietro (1980), who is for considering the “tool aspect” of language, calls attention to the fact that humans are more than “grammar-machines”; “[w]hile guided by the conventions of their grammars, humans use language to assert participation in a community of fellow beings and to communicate meanings and intentions” (Di Pietro 1980: 314). For an accurate communication to take place or for verbalizing our thoughts with precision, it is difficult to over-emphasize the prominence of lexicon, which is nicely illustrated by a quotation taken from Bogaards (1996: 358):

As a rule, vocabulary is more important than grammar. Indeed, with only a few loose elements and no syntactic connection, something happens anyway. Take, for example, the following sequence: *Storm fall tree dead*. These words call up a rather clear image, even though it is not at all certain which grammatical categories they belong to and regardless of the fact that they can mean many different things. [...] Furthermore, the foreign language learners’ need for lexical elements, in early as well as in later stages of the learning process, is often far greater than the need for grammatical rules. Travellers take dictionaries along when going abroad, not grammar books (Hatch 1978). Moreover, it is worth noting that in situations where getting a

message across quickly and precisely is of the utmost importance (in telegrams, panic situations or times when emotions are high), people limit themselves to what is essential, namely, the lexicon, and reduce the syntax to the very minimum.

In addition to the irritation and difficulty in comprehensibility they cause, lexical errors play a decisive role in the quality assessment of learners' essays. For instance, in an analytical assessment of students' writings done by Astika (1993), vocabulary has been found to be the best predictor of a learner's writing proficiency score, compared to other four composition components investigated – namely, content, organization, language use and mechanics (see also Grobe 1981). Likewise, investigating the role of the lexical component in the phase of assigning a quality score to compositions written by (high-)intermediate ESL learners through four lexical richness measures (lexical variation, error-free variation, percentage of lexical error, and lexical density), Engber (1995) has found that the diversity of lexical choice and the correctness of lexical form have a significant effect on reader judgments of the quality of an essay. The researcher concludes by supporting the basic premise that “the lexicon is an integral component in both the construction and interpretation of meaningful written text” (Engber 1995: 151).

Despite the importance of lexical proficiency to communication and (academic) success in foreign language education, it is very rare that lessons or courses are organized to focus specifically on vocabulary (Gass 1988: 95f.). Learners, just as native speakers, have an internalized

L2 lexicon but, as demonstrated by Meara (1984: 231ff.) through word association tasks, there are major differences between the lexicon of these two groups, i.e. in the way they store and handle words. Given the lack of concentration on vocabulary combined with the relatively confined exposure to the target language, especially in instructional settings, and the reliance on L1 or on other languages known to the learner during the L2 learning process, it is not a surprise that there are differences in the lexicon of native speakers and that of foreign language learners. The learner's lexicon, compared to that of the native speaker, is more unstable and more loosely organized, and the associations of words are quite often not semantically motivated but instead depend on some kind of phonological link with the stimulus word (Meara 1984: 232-234). These differences result from the fact that learning vocabulary in L2 does not happen independently of the already existing semantic system of the learner's L1, which only partially overlaps with the semantic system of the target language. The learner's reliance on L1 in the process of L2 acquisition was expressed long ago by Lado (1957: 2) in the following sentence:

[...] individuals tend to transfer the forms and meanings, and the distribution of forms and meanings of their native language and culture to the foreign language and culture – both productively when attempting to speak the language and to act in the culture, and receptively when attempting to grasp and understand the language and the culture as practiced by natives.

Blum and Levenston (1978: 409) assert that all L2 learners have the initial assumption that for every word in their mother tongue there is a

single translation equivalent in the L2 as a working hypothesis and, indeed, “thinking in the mother-tongue” is the only way learners can ever begin communicating in their L2. The researchers further state that “positive transfer” is the primary way of increasing one’s control of the L2 vocabulary (ibid.). There is some evidence from research on vocabulary development indicating that transfer may have a facilitative effect on learning – also beyond lexical items that show overt similarity (see Ard and Homburg 1992 [1983]). Despite the occasional facilitative effect of L1 transfer in vocabulary learning, because of the “natural” assumption of the learner that there is a one-to-one lexical relationship between his/her mother tongue and the target language, the incongruencies in lexical “gridding” between L1 and L2 are a potential source for the learner’s difficulties and errors – what Dagut (1977: 227) refers to as “divergences” from the lexical norms. As for the potential difficulty caused by L1 for the learner, Dagut (1977: 224) states the following:

[...] it is only to be expected that the language-learner will intuitively transfer to the foreign language the lexical segmentation of experience that his own tongue has accustomed him to finding normal and natural [...] and that he will have psychological difficulty in adopting the different “world-view” embodied in the lexical segmentation of the foreign language.

According to Dagut (1977: 227), although “one-to-many semantic relationship” is seen in both directions between L1 and L2, the interference is always from “the “coarser” to the “finer” grid, i.e. in the direction of making *too few* semantic distinctions” (italics in the original). That is,

an L1 item that has more than one L2 equivalents will be troublesome for learners since they often tend to choose one of these equivalents and “consistently let it do the job of both or all of them, spreading its area of meaning to cover the semantic space of the L1 source item” (Hasselgren 1994: 251), resulting in what Arabski (1979: 32ff) refers to as “underdifferentiation errors”. The generally “finer gridding” of English lexicon compared to, for example, Norwegian or Hebrew, has already been found to be a major source of lexical misselections by Hasselgren (1994) and Dagut (1977), respectively. Moreover, words limited to one specific register are harder to learn for the foreign language learner: to be able to choose the appropriate lexical item for each register, the learner needs to familiarize himself with phenomena that go beyond mere linguistics, such as “the socially-defined relationships between individuals in the language community” (Laufer 1990a: 302).

Being aware of the lexical difficulties, the learner applies certain strategies in lexical usage. Levenston and Blum (1977) name the overuse of “general” items as one of the major characteristics of learner language – a “general” term is defined as “usually a common word with a wide range of meanings, learnt in the early stages” (Levenston & Blum 1977: 55).²⁵ The researchers explain the use of this strategy by stating that a general term is reorganized as a hyperonym in the lexical system of the target language by the learner, partly due to L1 transfer, and it is then

²⁵ Carter (1987, 2012) refers to these items as “core” words and defines them as words which are ‘central’ in their meaning, neutral in discourse, i.e. in style or connotations, and collocate widely, such as *big* or *get*.

used in contexts which normally require a more specific term; during this overgeneralization, all register restrictions and collocational restraints are ignored (Levenston 1979: 152).²⁶ The dependence on and overuse of general words have already been confirmed, for example, for advanced Norwegian learners of English (Hasselgren 1994). Similarly, young L2 learners in a French immersion program have been found to show a higher tendency to use “utility” (i.e. higher frequency) verbs compared to native French speakers (Harley & King 1989).²⁷ Hasselgren (1994: 237) accounts for the learner’s clinging to the familiar words as follows:

Finding the right word in a second language is not easy. The choice may seem vast and time is rarely on our side. And the pressure not to produce a ‘bloomer’ intensifies as we advance. Little wonder then that, stripped of the confidence and ease we take for granted in our first language flow, we regularly clutch for the words we feel safe with: our ‘lexical teddy bears’.

This reliance on the so-called “lexical teddy bears” consecutively results in a lack of lexical variation in learner language (see e.g., Linnarud

²⁶ Conversely, the opposite direction of reorganizing can be observed in learner language, though not as commonly: “A specific term with clear semantic (and sometimes also morphological) restraints is raised in status and used as a hyperonym in all contexts” by the learner (Levenston & Blum 1977: 55).

²⁷ Cf. Sinclair (1991: 79) who has claimed that most learners avoid using common verbs as much as possible, especially in the cases where they are part of idiomatic sequences, and they rather “rely on larger, rarer, and clumsier words which make their language sound stilted and awkward.” As will be noted later in this section, the results obtained by Altenberg and Granger (2001) lend support to Sinclair’s claim.

1986; De Haan 1997; Harley and King 1989; see also Gilquin and Granger 2011).

It should, however, be noted that even though the learner prefers words that can be generalized for use in a high number of contexts rather than words with specific meanings, s/he has great difficulty with their use. The difficulty the learner has with these general words has been repeatedly expressed and illustrated in the literature; for example, the difficulty of high-frequency verbs (e.g., Sinclair 1991; Altenberg & Granger 2001; Lindner 1994; Lennon 1996; Howarth 1996; Zhou 2016; Nesselhauf 2005; Wang 2016). These verbs, which are highly polysemous²⁸, have many language-specific idiosyncrasies, making them “treacherous for foreign language learners”, although they display major semantic similarities across languages (Altenberg & Granger 2001: 174). Focusing on the verbs *put*, *go*, *recognise* and *take* in the speech productions of four advanced learners of English, Lennon (1996) has shown that even they produce high frequency verbs erroneously to a considerable extent and has stated that in most cases the posited verb was a high-frequency, ‘easy’, verb. That is why Lennon maintains that the availability versus non-availability of lexical items to the learner is not an adequate explanation for the learner’s lexical problems in production

²⁸ According to Altenberg and Granger (2001: 174), the polysemous nature of high-frequency verbs results from two types of meaning extension: (a) one universal tendency creating more general, abstract, delexicalized or grammaticalized uses, and (b) various language-specific tendencies resulting in specialized meanings, collocations, and idiomatic uses.

and states that his findings reject the reasoning that “learners are ‘over-extending’ verbs available in their lexicon to compensate for lack of rarer, more ‘difficult’ or specialized verbs” (1996: 23). On the basis of his findings, Lennon argues that even advanced learners have not figured out the meaning and usage boundaries prevailing among some very common verbs, lacking information with regard to their collocational possibilities/restrictions (1996: 28 and *passim*). Focusing on the verb *make*, Altenberg and Granger (2001) have attained similar findings for advanced learners of English with French and Swedish L1 backgrounds; these learners have been found to have great difficulty with high frequency verbs, and whereas some of the problems these verbs cause are shared by the two groups of learners, some are L1-related (see also Gilquin 2007 and Hugon 2008). The problems of these two learner groups with the high-frequency verbs have been found to be especially with their delexical usage²⁹; both learner groups not only underused the

²⁹ High-frequency verbs are often used as delexical verbs, which are also known as ‘light verbs’ (e.g., *do, get, give, have, make, put* and *take*). Although these verbs are not purely grammatical in the same way as auxiliary verbs (lexically ‘empty’), they are not lexically ‘full’ either (Allerton 1984: 32). As for the function of this special class of verbs, Allerton states that they “collocate with a non-verb word (usually a noun) in a construction to form an expanded form of the verb” (*ibid.*). The semantic content of the combination is provided not by verbs, but by the accompanying nouns, e.g., *take a bath, make a decision*, etc. (see also Quirk et al. 1985: 750-752). Allerton further adds that “[t]he most common such construction is when a transitive ‘general verb’ combines with a deverbal noun (phrase), but there are others, such as the combination of intransitive ‘general verb’ with preposition plus deverbal noun (phrase), e.g., *go for a walk, be on the move*” (Allerton 1984: 32f.). According to Allerton, ‘general verbs’ constitute a subset the lexical verbs which participate in such constructions, and as opposed to the

delexical structures of *make* – lending support to Sinclair’s claim (see footnote 27) – but also misused them (Altenberg & Granger 2001: 178). The researchers have noted that the delexical use of the verb *make* accounts for the majority of learner errors with this high-frequency verb in the investigated corpus (Altenberg & Granger 2001: 178f.), e.g. **make a balance/poll/a step*, etc. What makes the delexical use of high-frequency verbs difficult is, as pointed out by Allerton (1984: 33), mostly arbitrary choice of these verbs: “the choice of *take* in *take a step* (rather than *make*, for instance) is semantically unmotivated and therefore likely to be largely language-specific (compare English *take/*make a step* and French **prendre/faire un pas*)” (Altenberg & Granger 2001: 174f.). Despite the low number of delexical verbs in English, they are extremely frequent in language use (Sinclair 1990) so that mastery of them cannot be avoided by the learner.

While trying to get the meaning across in the target language, in addition to the common use of words of general meaning, the learner applies other strategies in order to cope with her/his inadequate command of the appropriate vocabulary, especially under the pressure of on-line production³⁰: avoidance, which goes hand in hand with the use of target

lexical verbs, ‘general verbs’ form a closed subset, and are therefore comparable to prepositions (Allerton 1984: 33.)

³⁰ According to Lennon (1996: 23f.), under the pressure of spontaneous language production, the learners may face at least the following problems: (a) gaps in lexicon; (b) incorrect information about lexical items; (c) retrieval (accessing) problems of lexical items; (d) imperfectly established lexical items, i.e. ‘incomplete lexical entries’; (e) over-activation of items (two or more items may compete for a single ‘slot’); (f) retrieval of the ‘wrong’ item (possibly due to phonological or L1 contamination), and (g)

word's (near-)synonyms, co-hyponyms and superordinates, word-coinage, shift to L1 (Poulisse & Bongaerts 1994), use of cognates, and literal translation from L1 to L2 – the so-called ‘transliteration’³¹ – to name but a few. Some other strategies found by Hasselgren (1994) applied by the Norwegian learners of English are the use of perceived equivalents and associations, i.e. use of a word which either is phonologically similar or has a similar meaning to the target word.³² Having message adjustment in the form of generalizations, approximation and word coinage as major communication strategies used in his data, Hendrickson (1976: 8) mentions also the occasional omission of a lexical item altogether, either intentional or accidental. The use of all these communicative strategies, i.e. compensatory strategies, in most cases results in either non-existent items – in Hasselgren's term ‘invalid’ items – or a lexical deviance of some kind from the native speaker norm. Although the reliance on compensatory strategies changes depending on the proficiency level of the learner (e.g., Poulisse & Schils 1989; Arabski 1979: 38f), all foreign language learners employ one or

learner uncertainty and inability to identify the nature of the problem being experienced (e.g. whether the perceived gap represents a gap in the language, in store, or is an accessing problem).

³¹ As for the strategy of ‘transliteration’, Hasselgren (1994: 241) states that it reflects “the process of breaking down an L1 item and reassembling it with L2 parts, often resulting in a non-existent word or phrase”. The researcher notes that L1 compound words are especially prone to transliteration (ibid.).

³² Hasselgren (1994: 239) thinks that the word ‘strategy’ is not the appropriate one to be used in this context as it has connotations of deliberateness and she therefore opts for the term ‘route’ instead.

more strategies to overcome the communication problems they are confronted with at times.

There is one last, very important, point regarding lexicon that needs to be labored here briefly, even at the expense of overstating the obvious by now: Vocabulary learning also involves the learning of multi-word units. In all languages there are items which co-occur, some with higher frequency than others, some others which co-occur only when the need arises, and still others the co-occurrence of which seems not possible or acceptable, even though they are semantically compatible (Wilkins 1972: 126f). As opposed to the outdated assumption that words are chosen individually from the mental lexicon during language use, it is now known that many words are “either themselves patterns or form part of patterns which are quite fixed and stable and which are used routinely in relatively predictable situations” (Carter 2012: 84). As most of these co-occurrence patterns are not semantically motivated and differ from one language to another, as the aforementioned example *take a step*, learning, especially actively producing, multi-word sequences preferred by native speakers is anything but an easy task for learners. Even advanced learners show a lack of feeling for native speaker patterning (Hasselgren 1994: 256). Surprisingly, as mentioned earlier, this also extends to collocational knowledge of common verbs (Lennon 1996: 28). Investigating learner language, Hasselgren (1994: 253) states that “[t]he analysis of lexical dissonance revealed that wrong words are often wrong because they do not collocate with their neighbours.” The variety and extent of idiomatic multi-word sequences in any language is ample.

In order to avoid sequences in a target language that are deviant or idiosyncratic in some way from typical native speaker production, collocational information is an essential part of learning vocabulary. As Martin (1984: 133) states, “one does not really “know” a word until one knows its collocational profile”. This is now a common view widely held.

The general importance of the lexicon and the problems it causes to learners has gained a more central stage in SLA research, especially with the linguistic analyses claiming a more extended role for lexis in language. Along with the changes in the view of language knowledge, recent years have seen a growing interest in the role of multi-word units in both first and foreign/second language acquisition and a substantial body of research now exists to show their potential contribution to language learning and teaching.

2.3 Phraseology and multi-word units in native and non-native speaker language and their role in language learning and use

2.3.1 Phraseology

The combinatorial creativity has been considered as the fundamental feature of human language in most linguistic models (Van Lancker 2001). Although the combinatorial potential of language is an undeniable fact, much of language use has been shown to consist of memorized, prepatterned conventional expressions. Prepatterning in language has been observed and repeatedly expressed by several scholars, especially by Bolinger (1961, 1976). He has argued that language use is much more based on lexical units of varying sizes and repetitive, not particularly creative and language itself is much more memory-based than acknowledged. That is, according to Bolinger, language use in reality showcases an interplay between more or less remembered and newly created speech. More recently, John McHardy Sinclair, who has emphasized consistently the non-random nature of language, the need to analyze long texts and search for observable textual evidence of meaning, has taken Bolinger's ideas one step further. While not denying the fact that grammar enables endless combinatorial possibilities to its language users, Sinclair (1991) has argued that in practice such possibilities are ignored and some lexical combinations are used repetitively (see Bolinger 1979 [1976]). Moreover, he has eloquently argued against the assumption that the word is the primary unit of meaning and put

forward the hypothesis that units of meaning are ‘largely phrasal’ (Sinclair 1996: 82). This claim is based on his two central concepts that explain the way in which the meaning of a language text is interpreted: the ‘open choice principle’ and the ‘idiom principle’. The ‘idiom principle’ refers to the phraseological tendency of language, i.e., words do not occur at random and in isolation, they rather combine with each other to make meaning (Sinclair 1987, 1991, 2004: 29). That is, this principle enables a speaker to choose from an array of prefabricated multi-word units. As Sinclair (1987: 320; 1991: 110) himself puts it, “a language user has available to him or her a large number of semi-pre-constructed phrases that constitute single choices, even though they might appear to be analysable into segments.” One of the examples Sinclair provides to illustrate this co-operation on meaning making is the phrase *of course* (1991: 110f.): Even though it consists of two words, it functions as one word just like other one-word adverbials such as *sure*, *perhaps* or *maybe*. Sinclair (1991: 111) argues that the first component of the phrase is not the preposition given in grammar books and the second component is not the countable noun found in the dictionaries but these two components together take on meaning in the phrase. This is in contrast to the ‘open choice principle’ (also referred to as the ‘slot-and-filler’ model) which is, according to Sinclair (1987: 319; 1991: 109), “a way of seeing language text as the result of a very large number of complex choices”, i.e. the assumption is that each word is chosen individually to fill a certain slot in a sentence, the only constraint being grammaticalness. There are, however, restrictions on word choices that

go beyond those of grammar and register and the open choice principle does not provide for substantial enough restraints on consecutive word choices (Sinclair 1991: 110). That is why Sinclair has expressed the need for the idiom principle, which is, according to him, at least as essential as grammar in the explanation of how meaning arises in text (1991: 112). Based on his observations of authentic data, Sinclair has proposed that text production happens as a result of continuous alternation between these two modes of processing and native speakers apply the idiom principle before the open choice principle when they process language, i.e., the idiom principle being the dominating principle of text interpretation. In his own words:

For normal texts we can put forward the proposal that the first mode [of text interpretation] to be applied is the idiom principle since most of the text will be interpretable by this principle. Whenever there is a good reason, the interpretative process switches to the open choice principle, and quickly back again (Sinclair 1987: 324).

With these two principles, and his work on collocations, Sinclair has paved the way for a conception of language handled in two different ways: stored as sequences and as individual words simultaneously – dual-processing model of language, which was previously been proposed by several other scholars (e.g., Lounsbury 1963; Hopper 1988). Accumulated empirical evidence showing the prevalence of seemingly prefabricated language has then given support to dual processing theory. Support for this theory comes from a variety of sources: from an analysis of natural speech corpora, psycholinguistic studies on adults

2.3 Phraseology and multi-word units in native and non-native speaker language and their role in language learning and use

and children, and neurolinguistic observations (see, for instance, Van Lancker-Sidtis and Rallon 2004). Thus, there is now a strong argument for a model of language that alternates between “fixity and novelty” for the creation of meaning (Tannen ²2007: 48ff. [1989]). All these developments in the field have given rise to the interest in phraseology.

Phraseology is defined as “the study of the structure, meaning, and use of word-combinations” (Cowie 1994: 3168) and according to Altenberg (1998: 101) it is “a fuzzy part of language”. Until quite recently, it has been rather a neglected area of language description. Since it does not make a sharp distinction between grammar and lexicon and takes a holistic view, there was no place for it in the traditional “apparatus” of language analysis (Sinclair 2008: xv). Interest in phraseology has grown considerably over the last three decades. The main reason of this increasing interest in phraseology is the availability of computerized corpora. From the 1980s onwards, the increased availability of large electronic corpora has made it easier to study the patterning in language. With the awareness of the abundance and the importance of ready-made multi-word combinations in language use, the growing interest in the collocational aspects of the lexicon has resulted in the ‘rise’ of phraseology as a sub-discipline of lexicology. A growing number of scholars in corpus linguistics have been focusing on multi-word units, lexical patterns, constructions both in native and non-native language, just to name a few: Millar (2011), Schmitt (2004), the contributions in Granger and Meunier (2008), Meunier and Granger (2008), Römer and Schulze (2010), Römer and Schulze (2009), Hunston and Francis (2000).

This section is intended to provide an introduction to phraseology – more precisely to the use of multi-word units in native and non-native speaker language as far as is relevant to the present study without claiming to be full-scale. As a starting point, we will need to examine the definition and nature of multi-word units in general. Then we will move to their functions in native and non-native language.

2.3.2 The notion of multi-word units – their definition, characteristics and categorization

A multi-word unit is defined by Wray³³ (1999: 214) as follows:

a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar.

This definition encompasses various types of word sequences in different sizes; from word sequences that are idiomatic, immutable and syntactically irregular³⁴ (e.g., *by and large*) to word sequences that are transparent and contain slots for open class items (e.g., *NP be-TENSE sorry to keep-TENSE you waiting* – the so-called ‘lexicalized sentence stems’ in the terminology used by Pawley and Syder (1983)) (Wray and

³³ Wray (1999) adopts the term ‘formulaic sequence’.

³⁴ The multi-word sequences with a structure that cannot be produced or predicted with the general knowledge of the familiar rules of grammar are called by Fillmore et al. (1988: 505) “extragrammatical idioms”.

2.3 Phraseology and multi-word units in native and non-native speaker language and their role in language learning and use

Perkins 2000: 1). Thus, a multi-word unit is interpreted as a whole as if it were a single word in its own right and believed to be stored holistically.

There is a considerable terminological profusion concerning frequently recurrent sequences of words; they are variously called in the literature, among others, *prefabricated units*, *prefabs*, *phraseological units*, (*lexical*) *chunks*, *multi-word units/expressions*, *formulaic sequences/expressions*, *routine formulae*. Wray and Perkins (2000: 3), for instance, have identified over forty terms that have been used to refer to one or more type or subtype of formulaic language (see Wray 2002: 9 for a comprehensive overview of terms). Although some of the terms used to refer to sub-categories of multi-word units have acquired a rather fixed meaning (e.g., idioms are customarily defined as being non-compositional), some others are quite confusing (e.g., the term ‘collocation’ is used for different meanings) (Granger & Paquot 2008: 35). Using different terms to refer to the same units or the same terms to refer to vastly different units in the field is quite problematic. Cowie (1998a: 210) condemns the abundance of different terms in the field by referring to phraseology as “a field bedevilled by the proliferation of terms and by conflicting uses of the same term.”³⁵ In the present context, the terms ‘multi-word units’, ‘prefabricated units’ and ‘formulaic sequences’ are

³⁵ As a possible reason for this variation in terminology, Wray and Perkins (2000) rightly propose the multi-faceted nature of formulaic language: it has been distinguished according to its form, function, semantic, syntactic and lexical features, and lastly its relationship with novel language (p. 3).

used interchangeably since they are fairly neutral terms, i.e. they do not refer to any particular theoretical framework, which makes them appropriate for the present purpose.

Despite the diversified terminology, there are a few criteria differentiating multi-word units from free combinations – some of which have already been touched on. Their multi-word character is their first obvious characteristic: As mentioned earlier, multi-word units consist of at least two lexemes (e.g., collocations), but can also expand to whole sentences (e.g., proverbs). However, regardless of their multi-word character, multi-word units are lexicalized, that is, they behave like a single word and are stored as whole units in the mental lexicon. Being stored as whole units, they are also produced as units. That is, they are reusable; they do not need to be produced from scratch each time they are being used. Indeed, the reuse of existing combinations by members of a given speech community makes them habitual, and, in turn, the preferred ways of expression in that community, i.e. conventional and institutionalised (see Langacker 1987: 62). This tendency is clearly noticed when the ‘collocabilities’ of synonymous words are compared: e.g. whereas *perform*, *carry out* and *conduct* all collocate with *test* and *experiment*, only the first two collocate with *task* and only the last two with *experiment*, for no obvious reason: **conduct a task*, **perform a survey* (Cowie 1994: 3169). In most prepatterned phrases, numerous other elements are possible members both in theory and in practice but many sequences are so well-established as groups that they enjoy lexemic status, “simply by virtue of being more common” (Kjellmer 1991:

114). In some cases phraseological tendencies result from real-world circumstances – for instance, in English-speaking countries water is normally served in glasses rather than in mugs, cups or cans, so *glass of water* can be expected to be better established in English rather than *mug of water* (Kjellmer: 1991: 114). Similarly, *room* being a significant collocates of both *door* and *window* strongly supports the tendencies in the world. As Sinclair (2004: 29) states, “here language does little more than correlate with the world”. Many of the phraseological tendencies, however, “seem almost purely linguistic” (ibid.). That is, the lexical choices in the combinations are in most cases arbitrary: As Kjellmer (1991: 114) explains, “there is a tendency, independent of circumstances in the real world, for speakers of a language to arrange words in groups and to use the same arrangement whenever the need arises rather than to arrange the words or to choose alternative ways of expression” (the so-called ‘nativelike selection’ which I will expand on in Section 2.3.3). Halliday’s (1966) famous *strong tea* and *powerful engine* illustrate this point very well: Although these two adjectives should be interchangeable on semantic grounds, they are not on collocational grounds (Sinclair 2004: 29).

One further characteristic of multi-word units, which is closely related to the characteristic just mentioned in the preceding paragraph, is their fixedness. All multi-word units are fixed to at least some degree, either syntactically or lexically, and cannot be changed randomly in most cases. Some of them are completely fixed and cannot be modified in any way (e.g., *loud and clear* in which the order of the lexical items or

the lexical items themselves cannot be changed as in **clear and loud* or **loud and manifest* for no real-world reason (Kjellmer 1991: 114)), whereas some allow little variation in the choice of the lexical items of which they are made up (e.g., *rare as hen's teeth* can become *scarce as hen's teeth* or *like hen's teeth* but not **rare as chicken's teeth*).

One other criterion of multi-word units is their non-compositionality or idiomaticity. When a sequence of words is semantically non-compositional, its meaning cannot be deduced from the individual meanings of its constituent words. However, as pointed out by Wray (1999: 215), although semantic non-compositionality is a typical feature of multi-word units, it is not a defining one because if it were, many transparent formulaic word sequences would have to be disregarded; for instance, collocations, semantically relatively transparent phrases such as *make good time* or *kick the habit*, Pawley and Syder's (1983) 'lexicalized stems' (e.g., NP *tell-TENSE the truth*) or Nattinger and DeCarrico's (1992) 'lexical phrases' (e.g., *a while/month/year ago*), etc. Furthermore, non-compositionality can be seen as a continuum, with entirely transparent units at one end and entirely opaque ones at the other.³⁶ Between these two end points there is a whole range of gradations; some

³⁶ By referring to Warren's (2001) claim – the meanings of many transparent expressions are not entirely derivable from the meanings of the constituent words – Wiktorsson (2003: 17f) states that dividing expressions into compositional and non-compositional on the basis of whether they are opaque or not may be a too simplistic approach and demonstrates this convincingly through the use of the examples of *bed and breakfast* and *from coast to coast*.

elements of the units are transparent and some are more metaphorical than the rest of the phrase. For instance, in the phrase *get down to brass tacks*, the first three units are more transparent than the last two.

Regardless of their length and varied internal syntactic structures, one other common feature of multi-word units is the way they are pronounced: rather fast, with a single, continuous intonation contour and without internal interruption, i.e., dysfluencies such as hesitations and pauses (Wood 2006; Lin 2010; Wray 2004).

As briefly touched on earlier, there are a great variety of multi-word units, in the structural, functional and stylistic sense and they are pervasive in various registers. Some are fully fixed in form, i.e. unchangeable, such as proverbs (e.g., *Let's make hay while the sun shines*), anomalous phrases like *by and large* in which words are semantically opaque but syntactically irregular, frames such as *If I were you*, *Perhaps we could* and collocations such as *slightly different*, which is more natural than *mildly different* (Wray 2002: 7). Kjellmer (1991: 112) refers to fixed sequences as 'fossilized phrases' – they are considered fossilized in that the occurrence of one word suggests the other with great precision (left- or right-predictive phrases). The only kind of variability that normally occurs in the predicted word ('collocate'), as Kjellmer (ibid.) notes, is that represented by inflectional endings: *Cocker Spaniel(s)*. Some other phrases are fossilized but still allow limited variation – what Kjellmer (1991: 113) refers to as 'semi-fossilized phrases': In these phrases one word predicts very limited number of words, e.g., *moot*

point/question, nodding acquaintance/terms, inferiority/Oedipus/persecution complex or some idioms such as *have a weak/soft spot for* (ibid.). Similarly, the so-called ‘semi-preconstructed phrases’ (Wray 2002: 7), as their name indicates, are also flexible to some extent; they can be used in different tenses for different persons and accept closed class variation (e.g., pronouns). *NP have + TENSE really done it this time* would be an example for this group of phrases, which can be used as *You’ve really done it this time; I’d really done it this time; John’s really done it this time*, etc. (Wray 1999: 215). There are still some other sequences, referred to as ‘lexicalized sentence stems’ by Pawley and Syder (1983), which are much more flexible since they have one or more slots for open class items, e.g., *NP be-TENSE sorry to keep-TENSE you waiting* (Pawley & Syder 1983: 210). According to Moon (1998), this latter type is more common in language use; the researcher found that 48 percent of the examples obtained from the data she investigated were of this kind (p. 98). In a similar vein, Kjellmer (1991: 113) has claimed that our mental lexicon contains a lot more phrases than those belonging to the types of fossilized and semi-fossilized phrases; he refers to this third type as ‘variable phrases’ and asserts that their number is much greater than that of the other two types mentioned above. Variable phrases comprise sequences of words that co-occur a lot more frequently than their individual frequencies would suggest and one of the words in such a sequence predicts the other word(s) but in this type of phrases, as Kjellmer (ibid.) notes, the ‘prediction’ is to be

2.3 Phraseology and multi-word units in native and non-native speaker language and their role in language learning and use

considered more loosely: One word in such a sequence tends to co-occur with one or a few out of a great number of words that can also co-occur with it. Thus, “[t]endency’ is here a key concept” (ibid.). There are two types of such variable phrases: (a) the sequences consisting of two or more lexical words, some of them also incorporating function words, such as *glass of water*, *classical music*, *in the years to come*, *go to college*, etc., and (b) established sequences of one lexical word and one or more function word(s), such as *a number of*, *for a change*, *it appears that*, *must admit that*, etc. Kjellmer (1991: 114) notes that as they aid in lending structure to the lexicon, the variable type phrases are a more crucial part of the lexicon compared to the fossilized and semi-fossilized phrases despite their ‘inconspicuousness’, which indeed allows them to be used in a variety of situations. He further adds that they are the phrases every learner of a language needs to master before dealing with the other two types. He goes on to remark that “[i]ndeed, it is not until one considers this last type that the ubiquity and indispensability of set expressions become fully apparent” (Kjellmer 1991: 115).

There have been different attempts to categorize multi-word units, which has resulted in various categorizations. The three-way classification of Kjellmer (1991) mentioned in the preceding paragraph is only one of them. The following few examples shall suffice to give a glimpse of the situation. Alexander (1984), for instance, distinguishes five broad categories of fixed expressions found in English: idioms (phrasal verbs, ‘tournures’ like *kick the bucket* or *put the cat among the pigeons*, and irreversible binominals like *cash and carry*, *uphill and*

down dale and *bag and baggage*); discourse structuring devices (greetings and formulae like *long time no see*, and connectives and ‘gambits’ like *for a kick off*); proverbs like *You scratch my back and I’ll scratch yours* and proverbial (metaphorical) idioms like *the proof of the pudding*; catchphrases which include clichés like *when all is said and done* and slogans like *Your country needs you*; and quotations and allusions like *You’ve never had it so good* (Alexander 1984: 128f). Cowie (1988, 1994), on the other hand, divides word combinations into two major groups which are distinguished from one another with respect to the kinds of meaning their constituents convey and to the structural level at which they function: composites and formulae. Cowie (1988: 134f) defines composites as “word combinations more or less invariable in form and more or less unitary in meaning, which function as constituents of sentences (as objects, complements, adjuncts, and so on) and contribute to their referential, or propositional meaning” and adds that composites are “lexical building-blocks comparable in their syntactic functions to nouns, adjectives, adverbs and verbs.” This group is further subdivided into collocations and idioms. Even though collocations are transparent and usually show some lexical variability, according to Cowie (1994: 3169), they are marked by an arbitrary restriction of choice at one or more points, e.g. *heavy rain*, *light exercise*, *perform a task* or *make a comment*. Idioms include both figurative and pure idioms; figurative idioms have a figurative meaning but preserve a literal interpretation (e.g., *do a U-turn*, *mark time*). Pure idioms, on the other hand, such as *blow the gaff* or *spill the beans* have undergone ‘petrification’ and are

semantically non-compositional, i.e. the meaning of the whole can no longer be deduced from the combination of the original meanings of its parts (Cowie 1988: 135). The category of formulae includes ‘sentence-like’ units which function on the pragmatic level as autonomous utterances and their meanings are mostly a reflection of their function in discourse, such as greetings, enquiries, invitations, etc. (Cowie 1988: 132). In his later work, Cowie (2001) further divides the category of formulae into two subcategories: routine formulae, like *good morning*, or *see you soon*, which perform speech-act functions, and speech formulae, which are expressions used to organize discourse and indicate a speaker’s attitude to other participants and their messages (*you know what I mean, I beg your pardon?*). A very similar typology to that of Cowie’s has been suggested by Mel’čuk (1998) although his terminology differs.

More recently Erman and Warren (2000) have categorized multi-word units, in their terms ‘prefabs’³⁷, into three broad categories: (a) lexical prefabs, which are semantic units such as *break a rule*; (b) function prefabs, which serve either grammatical functions such as *a great deal of*, *instead of*, or pragmatic functions such as *and then*, *good evening*; and (c) reducibles, which are types of abbreviations, such as *let’s*, *you’re* (see also Van Lancker-Sidtis and Rallon 2004).

³⁷ Erman and Warren (2000: 31) define a ‘prefab’ as “a combination of at least two words favored by native speakers in preference to an alternative combination which could have been equivalent had there been no conventionalization.”

These and similar differences in the categorizations of multi-word units,³⁸ according to Granger and Paquot (2008: 35), match the differences in the preference of characteristics adopted in the categorization of multi-word units and prioritization of the preferred characteristics to a great extent. The researchers state that most categorizations give priority to one or more of five characteristics of these units: “(1) internal structure (e.g. verb + noun or verb + preposition); (2) extent: phrase- or sentence-level; (3) degree of semantic (non-)compositionality; (4) degree of syntactic flexibility and collocability; (5) discourse function” (Granger & Paquot 2008: 35).

Regardless of the varied ways of categorization, all multi-word units serve the same functions in language learning and use. These functions are the focus of the next section.

2.3.3 Multi-word units in native language and their role in L1 acquisition and use

Although, as observed by Chomsky (1965), native speakers have the inherent capacity to generate an infinite number of utterances that conform to the syntactic rules of their mother tongue, they rarely do so and they rather have a tendency to use recurrent multi-word units available

³⁸ Van Lancker-Sidtis and Rallon (2004) believe that the categorization of formulaic expressions is troublesome partly because of the “uncontrolled and nonstandardized nomenclature” and assert the fact that there are almost no methods for verification of categories (p. 211). The researchers thus attempt to develop better methods for the classification of formulaic expressions by means of an analysis of a screenplay.

2.3 Phraseology and multi-word units in native and non-native speaker language and their role in language learning and use

to them that they easily retrieve during language use (e.g., Pawley & Syder 1983; Wilkins 1972; Coulmas 1981; Wray 2002; McCarthy 2010; Siyanova-Chanturia & Martinez 2015). As put by Altenberg (1998: 122), they have at their disposal “more or less conventionalized building blocks that are used as convenient routines in language production.”

The reliance on multi-word units by native speakers has been an idea with a long history. As early as 1976, Fillmore pointed out that “an enormously large amount of natural language is formulaic, automatic, and rehearsed, rather than propositional, creative, or freely generated” (p. 9). In a later piece of work on fluency, Fillmore posited their importance for native speaker’s competence by stating that “I believe that a large part of our ability to get along well in a language is our facility with formulaic expressions” (1979: 100). Bolinger is another proponent of the view that the role of formulaic sequences is far more extensive in language than previously acknowledged, as noted in Section 2.3.2. He asserted, “speakers do at least as much remembering as they do putting together” (Bolinger 1976: 2).

As a result of speakers making only limited use of the possibilities of their language, many well-formed utterances never occur because, as put by Coulmas (1981: 150), “they do not conform to the idiomatic preferences of the speech community” (see also Wray 2002: 5ff and Kjellmer 1991: 114). Wilkins (1972: 127f) expresses a very similar view when he stated that “joining together semantically compatible

parts does not always produce an acceptable co-occurrence” because there is an agreed form, better said a ‘preferred’ form, to express a concept between members of a given society. Indeed, if native speakers made full use of their capacity to generate an infinite number of novel utterances with the syntactical rules and words available to them, they would not only be faced with an immense processing task but also not be accepted as exhibiting ‘nativelike control’ of their language (Pawley & Syder 1983: 193). According to Pawley and Syder (1983), a fundamental aspect of nativelike linguistic knowledge entails the mastery of numerous multi-word units, in sentence-length as well as partly “lexicalized sentence stems”, which can be easily chosen and combined together, allowing fluent idiomatic³⁹ language use (see also Cowie 1998b: 1).

Using the bus-or-taxi analogy for the way people use language – traveling along regular routes like buses or moving more freely like taxis – Aitchison (1987) has demonstrated that people, who in general try to be buses, become taxis only if the bus-route is insufficient. That is, according to Aitchison (1987), since saying something entirely new demands

³⁹ Pawley and Syder (1983: 220f) note that the term ‘idiomatic’ has two meanings. The first meaning is the one mentioned earlier in this paper: semantic non-compositionality. The Oxford English Dictionary (OED) defines the second meaning, which is the one intended by Pawley and Syder, as follows: “Peculiar to or characteristic of a particular language; pertaining to or exhibiting the expressions, constructions, or phraseology approved by the peculiar usage of a language, esp. as differing from a strictly grammatical or logical use of words; vernacular; colloquial”.

considerable on-line computation, which is expensive in terms of processing time and energy (pp. 3f), people start by using memory and available routines and only if this proves inadequate, they rely on computation (p. 14). Having developed the bus-or-taxi analogy of Aitchison (1987), Kjellmer (1991) expressed a similar view when he suggested that an act of speaking or writing was like driving a car; abiding by the traffic rules and following certain stretches of road one arrives at the destination. Here is the direct quote that, I think, requires no gloss:

If we know the lay of the land reasonably well, we do not stop at every street corner to make fresh decisions, but we do have to make minor decisions at most crossroads and major decisions at main junctions. There is only a limited number of ways of reaching a goal, and few personal variations ordinarily occur (like driving across a lawn, or on the pavement). In a similar fashion we normally have a goal in speaking or writing [...] We have to obey the rules laid down by the grammar of our language and we normally follow certain 'lexical stretches', i.e. well-established sequences of words. If we speak the language reasonably well, we do not stop at every word or every few words to get our bearings; a lexical stretch will often link up or overlap with other lexical stretches that will take us further along our chosen path. [...] Again, few personal deviations from the established pattern occur, such as choosing unexpected words or ungrammatical forms. So, just as in driving, we use semi-automated routines in speaking and writing; both traffic rules/grammatical rules and a road network/a set of lexical stretches are essential to ensure adequate communication (Kjellmer 1991: 122f.).

As operating with prepatterned word sequences is the common way to move through the discourse, by not choosing the expected sequence native speakers sometimes intentionally depart from normal patterns of language use to create a ‘surprise’ effect on their listeners or readers (Kjellmer 1991: 123; O’Dell & McCarthy 2008: 6). This manipulation is especially used as a device for humorous or artistic reasons by journalists, advertisers and poets (see Leech and Short 1981: 28). The familiarity of native speakers is what makes it possible for this act to work, i.e., to have the intended effect on the listeners or readers (Kjellmer 1991: 123ff.; see Howarth 1998a: 29f). Thus, the use of manipulations in the familiar word sequences as a device indicates, as rightly pointed out by Kjellmer (1991: 124), not only the extent to which multi-word units are fixed and predictable but also the indisputable position of multi-word units as an essential part of a native speaker’s overall language proficiency. Modifications of predicted word sequences are, however, not always welcome; for instance, in the case of learner language. As mentioned earlier, a very common phenomenon known to most foreign language learners is the fact that although some sentences are perfectly grammatical, a native speaker would never use them (Allerton 1984: 39).⁴⁰ In other words, proficiency in a language means possessing not only a large set of recurrent phrases

⁴⁰ The difficulty posed by the randomness and illogicalness of idiomatic language to foreign language learners will be more elaborated on in the next section (2.3.4).

but also the ability to construct permissible variants of them “within the limits drawn up by the selectional rules” (Kjellmer 1991: 126).

Just like the abundance of various terms used to refer to them and the analysis and classification of multi-word units (see Section 2.3.2), the storage and processing models that have been proposed by researchers also show variety (see Wray 2002, for a clear and comprehensive survey). However, regardless of different treatments, today it is generally agreed that word sequences which are lexically and/or syntactically fixed to some extent behave more like individual words rather than separately constructed sequences – ‘computed’ forms – and that they play a major role in language processing and use. As stated by Peters as early as 1983, “normal adult speakers actually store and call into play entire phrases that may be many words long – phrases that are not constructed from their ultimate grammatical constituents each time they are used” (p. 2). Peters has also posited that in terms of usage and storage, there is no difference between whole utterances and short units, i.e. words; they are treated in the same way. In her words: “To the language learner, they are all units, and are stored in the lexicon and retrieved as such” (Peters 1983: 89). Similarly, Ellis (1996) has claimed that the role of working memory in learning word sequences such as frequent collocations, phrases, and idioms is the same as for words; thus, they can be viewed as “big words” (p. 111). More recently, Snider and Arnon (2012) have presented findings showing parallels in the processing of words and larger sequences (e.g., *don't have to worry*).

Although it is impossible to find a definite answer for the question of how many phraseological units are generally used in speech or writing, it is now known for sure that they are very commonly used in language production. A number of writers have commented on their abundance in language use. Pawley and Syder (1983: 210), for example, have asserted that “[t]he number of single morpheme lexical items known to the mature English speaker is relatively small; a few thousand” but this number goes up, at least, to several hundreds of thousands if we are talking about the number of sentence-length expressions familiar to the ordinary, mature English speaker (Pawley & Syder 1983: 213). Conklin and Schmitt (2008: 75) presume that this estimation of Pawley and Syder does not include collocations or phrase-length expressions, nor the number of larger than sentence-length expressions known by native speakers (such as poems, nursery rhymes, song lyrics, speeches from plays). According to them, “[t]he addition of these types of formulaic sequences would suggest that the phrasal component of the mental lexicon is likely to be truly vast” (Conklin & Schmitt 2008: 75).⁴¹ Altenberg (1998: 102) estimates that the language used by adult native speakers may consist of up to 80 percent of phraseological units.⁴² In a more

⁴¹ Conklin and Schmitt (2008: 75) note that “we know of no study that has directly attempted to quantify the size of a proficient user’s phrasal lexicon, and so we are currently left with the assumption that because the formulaic component of discourse is large, the phrasal lexicon must be too.”

⁴² Cf. Moon (1998: 57). Moon, in contrast, believes that the corpus she investigated (18 million-word Oxford Hector Pilot Corpus) may consist of only 4-5 percent multi-word units – what she calls ‘fixed expressions’ (FEIs). Such different estimates are a consequence of what is considered as a multi-word unit in studies. For instance, whereas

2.3 Phraseology and multi-word units in native and non-native speaker language and their role in language learning and use

recent research project conducted by Erman and Warren (2000), the pervasiveness of multi-word units in both spoken and written language has been verified one more time; the researchers have found that about 55 percent of a text consists of prefabricated language (see also Van Lancker-Sidtis and Rallon 2004 and Jackendoff 1995).

The variety and high number of multi-word units can be seen as evidence to their substantiality. Some, accordingly, consider them as “building blocks” for language use. Arnon and Christiansen (2014), for instance, argue that both adults and children rely on multi-word units in language production and comprehension and that these units constitute a substantial part of native knowledge of language.

The functions multi-word units serve can be given as a reason for their abundance in language.⁴³ Firstly, they reduce the processing time and effort of the speaker (e.g., Aitchison 1987; Pawley & Syder 1983; Wray 1999: 215f., 2002: 15-18; Conklin & Schmitt 2008; Skehan 1998) because, as thoroughly accepted now, it is a lot more efficient to retrieve

Altenberg (1998) applied a low threshold significance for recurring multi-word units and considered any continuous sequence of words occurring more than once in identical form as a multi-word unit in nearly half a million word London-Lund Corpus, without differentiating between units of meaning and random sequences of words, Moon set the significance threshold at five words (1998: 57). Besides, Moon counted any sequences of two or more words as FEIs and her analysis was based on a pre-established list of 6,776 multi-word units identified as FEIs (1998: 44f.). Moreover, Altenberg counted *tokens* in his study, whereas Moon counted the *types*.

⁴³ For an elaborate treatment of the different functions of multi-word units, see Wray (2002: 93-102).

a prefabricated sequence compared to going through the process of generating the sequence from scratch and create a novel one (e.g., Wray 2002: 18). That is because of the fact that frequent multi-word sequences in a language can be stored as whole units, i.e. ‘chunks’, in long-term memory.⁴⁴ Thus, multi-word sequences serve as “shortcutting devices” during the real-time performance, allowing maximal expressive effectiveness with minimal processing (Peters 1983: 3, *passim*). As Bybee and Scheibman (1999: 576) explain, “frequency conditions autonomy in storage and renders internal analysis unnecessary”. The human brain has vast storage capacity and can effectively perceive patterns and generalize. However, although the amount of information that can be processed simultaneously is still an unresolved issue (Ramsey et al. 2004), what is known for sure is that the human brain is not as well equipped for executing several mental acts simultaneously or in rapid succession (Pawley & Syder 1983: 217f.). As such, humans have highly restricted abilities to effectively encode novel speech in advance or during a conversation (Pawley & Syder 1983: 191). Therefore, instead of resorting to “the congested forum of on-line analysis” (Wray 1999: 215), where single lexemes have to be combined

⁴⁴ As long ago as 1956, in the field of psychology George Miller coined the term ‘chunking’ in his classical review of short term memory and put forth the strategy of ‘chunking’ for linguistic processing as one way to circumvent the limitations of short-term memory (“seven, plus or minus two”) on the amount of information humans are able to receive, process, and remember. He stated that by organizing the input sequence into larger and larger chunks humans could increase the total number of bits of information the memory span contains (cf. Cowan 2001). Reviews of the role of ‘chunking’ in language acquisition can be found, for instance, in Schmidt (1992) and Ellis (1996).

according to the speech situation and where text is seen “as a series of slots which have to be filled from a lexicon which satisfies local constraints” (Sinclair 1987: 320), prefabricated and habitual multi-word units can be used in language production. By this way, within the constraints of real time conversation the limited cognitive resources, i.e. working memory – a limited capacity system (Baddeley 2000), are freed up to process additional tasks. The speaker is thus able to focus on some other aspects of the discourse; for instance, as suggested by Pawley and Syder (1983: 208), on the planning of larger units of discourse rather than on producing individual sentences or on matching the timing, tone and rhythm of his utterance to his conversational purpose, on the social (as opposed to the linguistic) aspects of interaction (Peters 1983: 3), on evaluating the ideas in the conversation or on undertaking another, unrelated activity (Wray 2002). The fact that native speakers generally produce fluent multi-clause utterances that go beyond the limits of human information processing capacity made Pawley and Syder (1983: 191) claim that “fluent and idiomatic control of a language rests to a considerable extent on knowledge of a body of ‘sentence stems’ which are ‘institutionalized’ or ‘lexicalized’.” As these word sequences are pronounced quite fast, with no internal interruption and a continuous intonation contour, their use enhances speech fluency by extending the length of runs between pauses and making pauses less frequent (Wood 2006; Lin 2010; Wray 2004). Along the same lines, Bolinger (1979 [1976]) claimed that the presence of formulaic sequences in adult speech enables greater fluency than would be possible

if each phrase had to be newly created (see also Janssen and Barber 2012). A study conducted by Kuiper and Haggio (1984) lends support to these claims: The researchers found the livestock auctioneer speech to be almost entirely made up of chunks (in their term ‘*oral formulae*’) and they attributed this feature to the high processing demands faced by the auctioneers, i.e. the need to be fluent and to keep the attention of a mobile audience. That is, by relying on a small number of low-flexibility phrases, auctioneers are able to perform fluently, without pauses or hesitations for long periods of time (see also Kuiper 1992, 1996, 2004). Thus, their contribution to fluency, especially in spoken discourse that occurs under heavy time constraints, is second function they serve (see Pawley and Syder 1983, 2000: 164, 195f.; Hunston and Francis 2000: 271).⁴⁵

One thing worthy of note related to the issues of multi-word units and fluency is Wray’s (2002) one observation: The researcher has noticed that there is one type of multi-word sequence used commonly by speakers which have semantically equivalent single-word counterparts – termed ‘circumlocutory’ by Wray (2002: 74), e.g., *take/make a decision* (*decide*), *draw a conclusion* (*conclude*), *a sea-change* (*a change*), *at the end of the day* (*really*) and *up hill and down dale* (*everywhere*) (ibid.). Questioning the use of more words rather than the available one-word equivalent, Wray contends that there are other causes for processing

⁴⁵ The term ‘fluency’ is used here to mean the rate, in both productive and receptive skills. See Fillmore (1979: 93) for different kinds of fluency.

pressure such as the struggle to retain fluency and the sustaining of output while planning what to say next. She considers these multi-word units “time-buyers” and states that they help the speaker while planning what to say next and adjust the pace and rhythm of an utterance (Wray 1999: 216). In her words:

One possible explanation of the preference for a longer over a shorter way of expressing an idea is that it buys time for planning and/or that it contributes to an even rhythm [...] If so, then the saving of processing effort is not simply about taking short cuts. It is about regulating production so that it is manageable, and this may mean, at times, taking the long way round, adding padding and/or establishing and maintaining a particular preferred rhythm and flow (Wray 2002: 75).

The use of multi-word units not only contributes to fluency in speech and ease of production but also supports perception and enhances comprehension. As put by Skehan (1998: 38), the use of multi-word units is “double edged”. Being familiar to the hearer, who also has a processing burden in real time conversation, the use of multi-word units lessens the task of decoding on the hearer’s part, i.e. they are retrieved faster because the hearer does not need to analyze each word exhaustively or attend to every word in order to understand the meaning (see also Hunston and Francis 2000: 270), which in turn enables fluent processing (Stubbs 1997; Skehan 1998: 38).

There have been some studies in support of the assertion that formulaic language is associated with processing advantages over creatively generated language, for native as well as non-native speakers. For instance, in a measurement of reaction times to grammaticality judgments, Jiang and Nekrosova (2007) found that both native and non-native speakers responded to formulaic sequences significantly faster and with fewer errors than they did to non-formulaic sequences. Conklin and Schmitt's (2008) results also gave support to this assertion: The comparison of reading times for formulaic sequences and matched non-formulaic phrases (control phrases of similar length and structure) by both native speakers and non-native speakers showed that the formulaic sequences were read more quickly than the non-formulaic control phrases by both groups and, moreover, this processing advantage was seen regardless of whether the formulaic sequences were used idiomatically or literally. This led the researchers to conclude that a formulaic sequence has a processing advantage even when it is used literally (Conklin & Schmitt 2008: 83). One other study that reported online processing advantage for multi-word units is Tremblay et al.'s (2011). In a self-paced reading test, Tremblay et al. (2011) found that lexical bundles⁴⁶ were read faster than comparable groups of words that did not make up lexical bundles, e.g. *He sat in the middle of the bullet train* vs. *He sat in the front of the*

⁴⁶ Adopting the suggestion in Biber et al.'s (1999), Tremblay et al. (2011: 572) define lexical bundles as "sequences of two, three, or four words that recur as wholes at least 10 times per million as well as sequences of more than four words occurring at least 5 times per million."

bullet train. In follow-up two word and sentence recall experiments, the researchers also found that sentences containing lexical bundles were recalled correctly more often compared to sentences that did not contain lexical bundles. Their results suggest that regular multi-word units leave memory traces in the brain. Additional support for the processing advantage of formulaic language comes from Underwood et al. (2004). By means of eye-movement methodology, Underwood et al. explored the processing of formulaic sequences of both native and non-native speakers – namely, familiar idioms (e.g., *to kill two birds with one stone*) and other formulaic sequences (e.g., *on the other hand, as a matter of fact*) embedded in extended contexts. They measured how often and for how long the terminating words in formulaic sequences were fixated upon. The results obtained were then compared to measurements of the same words in non-formulaic sequences (e.g., *throwing a stone*). Their aim was to test whether the familiarity with the formulaic sequence would facilitate the processing of the final word; that is, whether recognizing the formulaic sequence from the first words would result in less attention to the final word. Their results showed that native speakers fixated less often and for a shorter duration on the terminal words when these words were in formulaic sequences than when not. Underwood et al. (2004) took this as evidence in support of the claims of holistic representation and processing of formulaic sequences. Furthermore, the processing advantage of formulaic sequences seen in native speakers is only partly shared by proficient non-native speakers; although they also needed fewer fixations on the terminal words in the

formulaic sequences, they still required an equally long gaze time for these words just as for the terminal words in non-formulaic contexts. In a more recent study undertaken to investigate by means of a phrasal-decision task whether the frequency of compositional four-word expressions (e.g., *don't have to worry*) affects processing, Arnon and Snider (2010) found that participants responded faster to high frequency phrases than to less frequent ones.

As justified by the aforementioned studies, multi-word units play an important role in the way language is processed and used. One may raise the question how learner deviations from the target language formulaicity affect online processing – whether it would create greater processing demands for the native speakers. To my knowledge, there has been only one study that investigated the processing of multi-word units produced by language learners, namely, the study conducted by Millar (2011). In a self-paced reading experiment, Millar investigated the processing of learner collocations by native speakers that deviate from target language norms and found that sentences containing non-nativelike collocations were read significantly more slowly compared to sentences containing nativelike word sequences (e.g., *best partner* vs. *ideal partner*). In other words, deviations from the target norms led to “an increased and sustained processing burden” (Millar 2011: 129). Millar’s findings add to the growing body of evidence showing the processing ease of formulaic language and provide empirical support for its importance for language learning and use.

Third function of multi-word units is a social one. As pointed out by Wray (2002: 72ff), some formulaic sequences have the function of signaling identity and group membership in a certain linguistic group (see also Hymes 1968: 126f and Terkourafi 2002). As individuals we use certain multi-word units to express our identity but we also have the desire to sound like others in our speech community so as to be a part of it. The use of common multi-word units play a major role in maintaining the identity of the community since, as explained by Wray (2002: 92), “[s]uch groupwide sequences are cyclically harvested from and resown into the shared linguistic inventory of the community, as individuals both imitate the preferred forms of others and also contribute to the pool of idiomatic material from which others draw.” Moreover, using multi-word units that are in certain use in a given community is an efficient tool for persuading others to behave in a manner agreeable and favorable to us (Wray 2012: 231). Besides, as multi-word units or ‘partly lexicalized frames’ that have their own semantic entry demand less decoding, their use minimizes the risk of misunderstandings in discourse (Wray 2012: 231f).

Lastly, maybe most importantly, multi-word units play an essential role in language acquisition; they are the basis for language development in early first language acquisition and children’s second language acquisition. In the field of first language acquisition, it has long been noted that children’s speech includes certain memorized units (e.g., Brown 1973; Bloom et al. 1975; Braine 1976) and their importance has long been recognized. Clark (1974) and Peters (1977, 1983), for instance,

postulated the central role of multi-word units in the process of some children's first language acquisition (see Bretherton et al. 1983 and Bates et al. 1988). It has also been noted that sequences of words can be extracted and treated as a single lexical unit by children (e.g., Bolinger 1975: 100; Crystal 1997: 244; Plunkett 1993: 44) and only with the increasing knowledge of the language and with growing language experience, i.e., what parts of a sequence are substitutable or modifiable, unanalysed wholes are eventually analysed into their constituent parts. Drawing on data from only one child, Peters (1977) has demonstrated that two different strategies are employed during the process of L1 acquisition – 'gestalt' (from whole to parts) and 'analytic' types of speech (from parts to whole) – and has proposed that the choice of strategy to produce an utterance is related to the personality and the function of the utterance in a given context.⁴⁷ More recently, Tomasello and Brooks (1999) claim that children come equipped to use both strategies and they probably use them both to some extent for the acquisition of different aspects of language.

⁴⁷ The child Peters (1977) reported on used mostly analytic speech in referential contexts such as naming pictures in a book or labeling a quality while he preferred gestalt speech in more conversational contexts such as while playing with his brother, requesting something rather than demanding or opening conversations (p. 566). See also Nelson (1973: 21f) who demonstrated that there were striking individual differences in the children's speech during their L1 acquisition: There were children, who she termed 'referential', using a largely object-oriented language and those, who she termed 'expressive', using a more self-oriented, a social interaction language.

2.3 Phraseology and multi-word units in native and non-native speaker language and their role in language learning and use

Asserting the role of gestalt processes in language learning, Peters (1983) has further proposed that the process of extraction of 'units' from the speech stream is one of the central processes in early language acquisition and has stressed the dissimilarity between the units used by linguists for analyzing language and the ones employed by children during early language acquisition. The researcher explains her proposition by stating the following:

Since the child does not know the language, it is unreasonable to assume that the first units she or he extracts will coincide with the words or morphemes of the system. On the other hand, it *is* reasonable to assume that at the earliest stages there is a cognitive limitation constraining the child to processing (e.g., extracting and producing) what to the child is one unit at a time (Peters 1983: 5; italics in the original).

In a similar vein, Plunkett (1993: 44), who agrees with Peters on the segmentation problem being central in first language acquisition, states the following:

En route to solving the segmentation problem, children may ascribe structural properties to the speech signal which do not match those of the adult language user. For example, children may ascribe lexical status to syllabic sequences which in the adult tongue are considered parts of words. [...] Alternatively, children may ascribe lexical status to whole sequences of words and use these lexical chunks with the same distributional properties as adult lexical items. [...] The use of fillers and formulaic expressions would appear to be a natural outcome of children's attempts to solve the segmentation problem.

These propositions fall into place when one bears in mind that words in fluent speech just follow one another; many word boundaries are not marked by obvious acoustic cues such as a brief pause. As put by Saxton (2010: 110), words “flow together in a fast-moving river of sound” which, in turn, makes it difficult for the infant language learner to detect when a phoneme or a word stops and the next one begins, if not impossible (for a very clear summary of research on early infant speech perception see Saxton (2010: 108-132)). One thing worth underlining in this context is the fact that even though word boundaries are not enhanced and difficult-to-segment word boundaries are not avoided in infant-directed speech (Aslin et al. 1996),⁴⁸ it has been demonstrated by Jusczyk and Aslin (1995) that infants show some ability to detect the sound patterns of words in fluent speech contexts by 7.5-months; they can segment words from fluent speech and recognize words that they have heard in fluent speech when those words are later presented in isolation (see Thiessen and Saffran 2003). This finding may be accounted for by the assertion proposed by Saffran, Aslin and Newport (1996); according to the researchers, infants have access to a powerful mechanism for the computation of statistical regularities of language input and accomplish the segmentation of fluent speech mainly on the basis of the co-occurrence probability between adjacent syllables (see

⁴⁸ As found by Brent and Siskind (2001), single-word utterances are a regular occurrence in infant-directed speech but infants hear relatively few words in isolation. The researchers have found in their data only 9 percent of the maternal utterances to consist of isolated words.

also Thiessen and Saffran 2003). That is, the frequent co-occurrence of sounds in the input can be taken to be words or components of a word, whereas the low-frequency transitional probabilities between sounds can be posited as a word boundary. Bannard and Matthews (2008: 241) have insightfully commented on this assertion by stating that “[a] child performing segmentation in this fashion is likely to arrive at an inventory of segments containing not just conventional words, but also a number of multiword sequences.”

There is now a growing body of evidence showing that children’s language includes multi-word units, even at a stage where most of their speech is found to be consisting of single words, and that the children are sensitive to the properties of multi-word units. Pine and Lieven (1993), for instance, have found that nearly all children make use of at least some of the so-called ‘frozen phrases’ in their early speech, though to varying degrees (see Barton and Tomasello 1994 and Bates et al. 1988). These longer and complicated sequences in the child’s speech are not constructed from their constituents via the computation of the internal structure of these sequences but, as proposed by Brown (1973), their production results from frequent occurrence of a structure in input that is beyond the present linguistic level of the child. Brown and Hanlon (1970: 51) have similarly suggested, “any form that is produced with very high frequency by parents will be somehow represented in the child’s performance even if its structure is far beyond him.” Acknowledging the fact that some of the formulaic sequences include structures that are nonproductive in the child’s language at a given

stage, Olson (1973: 156) has further remarked that these sequences are used as a unit by the child for some specific semantic or pragmatic purpose without knowing the internal structure of the sequence (see also Tomasello 2003: 36-40). Support for these claims comes from various studies. Having investigated her son Adam's speech, Clark (1974), for instance, found that Adam would incorporate units from a preceding adult utterance into his own without fully analyzing them and use these units in similar situations without making any internal modification to them. For instance, copying from an adult saying 'Sit on my knee', he would say 'Sit my knee' when he wanted to sit on an adult's knee (Clark 1974: 4). Moreover, the researcher observed that the copied utterances in Adam's speech – 'unopened packages' as she calls them – were retained intact for several weeks without other lexical items being substituted, which, according to her, suggests that these copied utterances function "as units with limited internal structure" (*ibid.*). Drawing on her data, Clark (1974) concluded that "the process of modifying a practised sequence internally is psychologically more complex than the process of collocating linguistic units" (p. 7); so, only after having practiced a sequence well could the child begin to modify it and only in a limited way (p. 5).

Similar observations to Clark's were made in a larger scale study conducted by Braine (1976): The researcher investigated 16 corpora of 11 children at an early stage of syntactic development learning various languages including English and concluded that children's early multi-

word speech consists of what he termed ‘limited-scope formulae’ uttered in order to convey certain messages. According to his data, most of the children’s early utterances were produced using ‘constant-plus-variable type’ patterns in which some words were used in a constant position and combined with various words, e.g., *more*, *all*, *no*, and *other* recurring in a constant position (the so-called ‘pivot’ words) are followed by different words.

The claim that children learn not only single words but also multi-word sequences, which are taken directly from the input for later use has been experimentally tested in a study by Bannard and Matthews (2008). Their results show that young children (two- and three-year olds) repeated highly frequent four-word sequences they encountered in input (e.g., *sit in your chair*) faster and more accurately compared to their matched sequences in which the terminal word was substituted with a lower frequent word (e.g., *sit in your truck*). Similarly, Lieven et al. (2009) demonstrate that a large proportion of multi-word units found in the speech of four two year-olds can be classified as ‘frozen’; i.e. they could be traced back to utterances they had previously produced – either exact repeats of previous utterances – or to an utterance that required only one operation (allowed operations were a substitution into a semantically similar slot or addition of a word to the beginning or end of an utterance). Thus, the constituents of these sequences were not used productively, rather only in restricted combinations. Due to the high degree of repetitiveness they found in their data, Lieven et al. (2009: 503) conclude that “children are initially learning not only words but

strings as ‘big words’ which subsequently start to be internally analysed”, allowing for greater productivity in their later use.

Children’s error patterns have also been proposed as evidence for the use of multi-word units in their speech. For instance, having investigated the production of irregular plurals in children’s language, Arnon and Clark (2011) have shown that children produce many more correct irregular plurals inside frequent frames. Similarly, it has been shown that children are more likely to produce correctly inverted wh-questions for sequences (i.e. wh+auxiliary combinations) that occurred inverted with high frequency in the input (see e.g., Rowland and Pine (2000) and Rowland (2007)). Arnon and Christiansen (2014: 89) mention four-year-olds’ difficulty in changing a first person prompt (e.g., *I think*) into a third-person one (e.g., *he thinks*) for verbs that mostly appear with a first-person subject, e.g., mental-state verbs. Similarly, children’s me-for-I errors (pronoun case errors where accusative pronouns are used in nominative contexts, e.g., *me do it*) have been found to correlate with the proportion of preverbal uses of *me* (e.g., *let me do it*) in their input (Kirjavainen, Theakston, and Lieven 2009).

The use of multi-word units has also been found in children’s second language acquisition process. Having investigated the speech of five Spanish/Mexican children, aged 5 to 7, acquiring English as an L2 only through exposure in an American kindergarten in a year-long study, Wong Fillmore (1976) found abundant use of formulaic units in their

2.3 Phraseology and multi-word units in native and non-native speaker language and their role in language learning and use

speech and emphasized their vital function in the language learning process: Besides being useful in aiding learners to internalize syntax, “as a quick means to be communicative” initially (Schmitt & Carter 2004: 11) they also help learners in establishing and maintaining social relations with native speakers before sufficient competency in L2 has been gained, which in turn results in increased language input (Wong Fillmore 1976; Bogaarts 2011: 15; Schmitt & Carter 2004: 11f).⁴⁹ In a case study investigating the process of L2 acquisition of a child between the age of 1; 10 and 3; 0, Karniol (1990) also demonstrated the prominence of multi-word units in the early stages of child naturalistic second language acquisition. Unlike in emerging L1, the child’s early speech did not have the one- or the two-word stage but it included many imperatives and interrogatives acquired by rote learning in an unanalyzed manner and the transition from the formulaic combinations to productive speech was found to occur via the juxtaposition of elements extracted from known formulas, resulting in errors such as the move from correct to incorrect use of verb forms (Karniol 1990: 160).

⁴⁹ Wong Fillmore (1976: 670; quoted in Krashen and Scarcella 1978: 294) explains the heavy use of formulaic language by children as follows:

“[...] the desire to maintain contact and sustain social relations with members of the group is the motivation for the acquisition of formulaic speech. [...] what he (the acquirer) must do is to acquire some language which will give the impression of ability to speak it, so that his friends will keep trying to communicate with him. The use of formulas by the learners in this study played an important part in their being able to play with English speakers as they did. [...] This kind of language was extremely important, because it permitted the learners to continue participating in activities which provided contexts for the learning of new material.”

More recently, Perera (2001) investigated the transitional stage of four preschool Japanese children in two-way immersion programs learning English as an L2 from one-word utterances to multi-word utterances. Her results demonstrated how these learners became capable of socializing and linguistically creative with the help of multi-word units (in her terminology ‘prefabricated language’). Most of the learners’ novel utterances were found to be constructed from multi-word units or analyzed multi-word units, i.e. they were not constructed from free combinations of words. The first few analyzed multi-word utterances were found to have a tendency to resemble the original multi-word unit; learners first replace a word in a multi-word unit with semantically similar words and then expand the range of word selection (Perera 2001: 336). Perera (2001: 327) concludes that multi-word units have an important role “as a scaffold for linguistic creativity.” Other studies which point in the same direction but in the context of L1 acquisition include, for instance, Lieven et al. (1992) and Pine and Lieven (1993).

Although the use of multi-word units, the internal composition of which is yet not known to children, so as to discover the individual units of language and the ruling regularities of their combination is now being considered to be as important a process as the process of using smaller units to form larger ones, the role of the unanalyzed units in the early stages of L1 acquisition has been a wrangling argument. Some researchers have argued that using formulaic language is a “dead-end or side street on the road to productive language” (Hickey 1993: 29). Bates et al. (1988: 223), for instance, have emphasized that the acquisition of

unanalyzed multi-word units in early vocabulary development “leads nowhere.” Other researchers have argued a rather small role for the multi-word units in the language learning process. For example, Brown and Hanlon (1970) in the process of L1 acquisition and Krashen and Scarcella (1978) in SLA observed that prefabricated language resulting from the frequent occurrence in input tends to become automatized and in turn resists segmentation. Krashen and Scarcella (1978) assert that although prefabricated language (in their terminology, ‘patterns’ and ‘routines’) aids children in socializing, it develops independently of the creative construction process (p. 287) and does not directly contribute to it. In essence, prefabricated language does not “turn into” creative constructions but an internal creative process evolves independently and accounts for the actual SLA (Krashen and Scarcella: 286). So, too, there are other early studies that lend support to this position, arguing against the direct contribution of prefabricated language in L1 and L2 acquisition; namely, Bates et al. (1988) and Brown (1973), Bohn (1986) and Lightbown (1983). The reporting of Clark (1974), however, do not coincide with this proposition: The unanalyzed multi-word units (referred to as ‘prepackaged routines’ or ‘unopened packages’) in her nearly three-year old son’s language developed into patterns through gradual analysis of the internal structure, which became productive over time. Other early studies that claim a role for multi-word units in the learning of L1 and L2 include, for instance, Huang and Hatch (1978), Peters (1983), and Vihman (1982).

One further, conspicuous, argument for the central role of multi-word units in language development is pointed out by Bybee (2010) as follows:

it is not the case that in general the language acquisition process proceeds by moving from the lowest level chunks to the highest. Even if children start with single words, words themselves are composed of smaller chunks (either morphemes or phonetic sequences), which only later may be analysed by the young language user.

Furthermore, Pine and Lieven (1993) consider the segmentation of ‘frozen phrases’ from the exposed language as being an analytic process as the segmentation of single words.

As obvious from the preceding, there have been various ideas on the way children acquire their language. Snow’s (1986) argument can be mentioned to settle the debate: The researcher has argued that there is not only one route to language acquisition and the acquisition of multi-word units that the child later analyses into their components is just one of these routes.

The rising awareness that multi-word units form an integral part of the language acquisition process has further evolved in the framework of usage-based and emergentist/constructivist approaches to language acquisition where children acquire the grammar of their language(s) from linguistic experience (e.g., Bybee 1998, 2006; Goldberg 1995, 2006;

2.3 Phraseology and multi-word units in native and non-native speaker language and their role in language learning and use

Langacker 1987; Tomasello 2000, 2003, 2009). Usage-based approaches to language hold that language learning happens by employing domain-general cognitive mechanisms – not differing from other types of learning, and that children construct a language based on exemplars; that is, language acquisition is primarily based on the linguistic input children receive. By hearing exemplars, children extract the productive linguistic elements from the language they hear being used around them and then generalize on the basis of a common pattern among those exemplars stored via general cognitive processes. So, this account of language acquisition is based on the assumptions that children’s language production relies on multi-word units taken directly from the input and the cumulative experience of children with language results in its learning (see Behrens (2009) for a review of empirical evidence).

Tomasello (2000, 2003, 2009), one proponent of the usage-based approaches to language, asserts that children acquire language mainly by understanding how adults use language and the starting point in language development for children is not the word but the ‘utterance’ (Tomasello 2003; 2009: 72) - he defines an ‘utterance’ as “the smallest unit in which a person expresses a complete communicative intention – that is, an intention that another person attend [sic] to something within the joint attentional frame and so do [sic] something as a result.”⁵⁰ The child

⁵⁰ Tomasello (2009: 74) also states that “children do not try to learn words directly; they try to comprehend utterances and in doing so they often must comprehend a word in

is prompted to understand the overall communicative function behind the multi-word units used by the adult while s/he at the same time attempts to extract and determine the communication function of the particular constituents within the utterance. In other words, the child must resolve both the adult's intention in using the language and what the linguistic elements used refer to, which will, in turn, enable the child to use the linguistic elements creatively in the future (Tomasello 2003; 2009: 72-74). As a result, Tomasello (2003: 39; 2009: 76) suggests conceiving the child's early utterances, even if they are composed of more than one word, as "holophrases", which he defines as indivisible units of language with an intonational contour that convey a single and complete communicative intention – expressions such as *I-wanna-do-it*, *Lemme-see*, and *Where-the-bottle* (2003: 38). Tomasello further claims that children's early utterances are tied very closely to specific lexical items. Thus, although the early multi-word units of children soon develop into "schemas" or "constructions"⁵¹, they are highly concrete, i.e. based on particular words and phrases, not abstract (Tomasello 2009: 76) – "item-based constructions" as referred to in usage-based approach. As shown in many experimental studies (e.g., Tomasello 1992; Lieven et al. 1997; see Tomasello 2000 for an overview of these studies), children cannot generalize their existing syntactic knowledge,

the sense of determining the functional role it is playing in the utterance – and they see commonalities in this functional role across utterances."

⁵¹ "A linguistic construction is prototypically a unit of language that comprises multiple linguistic elements used together for a relatively coherent communicative function, with sub-functions being performed by the elements as well" (Tomasello 2009: 75).

which depends on how they have heard a particular verb being used, for new verbs until after their third birthdays. Tomasello refers to this verb-specific knowledge as the “Verb Island Hypothesis” (2000: 213). The idea captured in this hypothesis is that children learn grammatical relations on a verb-by-verb basis (Tomasello 1992). The reason of the emphasis on verbs results from their pivotal role in determining the format of the whole sentence. As Tomasello (2003: 122) observes, “the most abstract constructions characteristic of adult linguistic competence typically revolve around verbs in one way or another.” These verb-island constructions are considered prerequisite for the construction of grammar. According to Tomasello (2009: 77f), “[e]arly syntactic competence is best characterized as a semi-structured inventory of relatively independent verb-island constructions that pair a scene of experience and an item-based construction, with very few structural relationships among these constructional islands.”

What is also known today is that learning through the employment of unanalyzed multi-word units is not only a strategy of early language acquisition; as language learners get more experienced, both children and adults, they continually acquire and use more (longer) multi-word units that they build additional chunks (Newell 1990; see also Olson 1973). Referring to Pine’s (1990) study, Pine and Lieven (1993) accentuate the fact that the acquisition of unanalyzed phrases is not a strategy used only in the very early stages of language development, but rather its importance increases as vocabulary development advances and they further note that this is the case for all children, not for one particular

subgroup. Accordingly, Tomasello (2003: 7) posits that adult language competence consists of nothing other than “a structured inventory of linguistic constructions”. The fact that native speakers do not analyse the multi-word units if the need does not arise supports this assumption (see Wray 2008: 223; Wang 2016: 5).

Thus far a picture of the growing acceptance of the role of multi-word units in language acquisition and use in the recent accounts to language acquisition has emerged. As a result recent years have seen an increasing interest in the role of these units in foreign language learning. A substantial body of research now exists to show their potential contribution to foreign language learning and teaching. The next section deals with the use of multi-word units in learner language and provides, fairly briefly, an overview of learner-related studies in this field, which will help create a better framework of reference for the present study.

2.3.4 Multi-word units and the learner

The increasing recognition that a large part of the native speaker’s language consists of units larger than a word and it is the knowledge of these multi-word units that allows the native speaker to use the language fluently have important implications for foreign language learners and for foreign language education. As multi-word units make a given language highly idiomatic, in order to achieve native-like proficiency the learner is to know the idiosyncratic ways of expressing things in the target language, i.e. the conventionalized and preferred ways of encoding meanings by native speakers in a given culture, since without the

control of an appropriate range of multi-word units, it is not possible to perform at a level satisfactory to native speakers, in writing or in speech (Cowie 1992). As pointed out by Wilkins (1972: 128), “[e]ven the extremely proficient foreign language speaker is still likely to be marked out as a non-native speaker if in his speech or writing he seems to avoid the collocations that would be characteristic of the native speaker” (see also De Haan 1997: 226). The learner, however, has to possess not only a high level of awareness as to which lexical combinations are used by native speakers but also which ones are not – what Pawley and Syder (1983) referred to as “the puzzle of nativelike selection.” The notion of “nativelike selection” is nicely summed up in the following quotation:

The fact is that only a small proportion of the total set of grammatical sentences are nativelike in form – in the sense of being readily acceptable to native informants as ordinary, natural forms of expression, in contrast to expressions that are grammatical but are judged to be ‘unidiomatic’, ‘odd’, or ‘foreignisms’. [...] If a language learner is to achieve nativelike control, then, he must learn not only a generative grammar as this term is usually understood – a set of rules specifying all and only the sentences of the language. In addition he needs to: learn a means for knowing which of the well-formed sentences are nativelike – a way of distinguishing those usages that are normal or unmarked from those that are unnatural or highly marked. How this distinction is made is the ‘puzzle of nativelike selection’ (Pawley & Syder 1983: 193f).

This distinction, as Wray (1999: 468) observes, creates a great challenge even for the most proficient learners. The results of an idiom experiment conducted by Kellerman (1977) – reported in Kellerman

(1979) – demonstrated that advanced English learners did not display great ability to spot which expressions were incorrect although they knew which expressions were correct, which was taken as evidence for the fact that even advanced learners lack the kind of knowledge native speakers normally have in the area of idiomatic language (Kellerman 1979: 43). In order to attain idiomatic knowledge or collocational awareness in a language, the knowledge of grammar and of vocabulary alone are not enough. Skehan (1998: 39) asserts that learners who restrict themselves to language production based simply on a combination of grammar and lexicon “will be for ever marked as non-members of the speech community they aspire to.” However, without a great deal of exposure to the target language, as pointed out earlier, there is no way for the learner to distinguish the acceptable from the unacceptable (Wilkins 1972: 128; see also Adolphs and Durow 2004)⁵², let alone extending his/her active repertoire of collocation knowledge.

The successful learning and active use of multi-word units is not easy for learners due to various reasons (see Section 2.3.4). Besides the fact that many of these units have non-literal meanings, i.e., non-compositional, many of the phraseological tendencies are random – “purely linguistic” (Sinclair 2004: 29). That is, the choice of component lexical items in many multi-word units is largely a matter of convention in a

⁵² As we shall have an occasion to note later on, it has already been reported in the literature that mere exposure to target language is not enough for attaining (active) collocational knowledge (e.g., Marton 1977: 54; Nesselhauf 2005)

2.3 Phraseology and multi-word units in native and non-native speaker language and their role in language learning and use

given culture. These variable conventions make the prediction of permissible combinations based on the knowledge of the meanings of the words challenging and often impossible for foreign language learners (Wilkins 1972: 126), e.g. while *think on the feet* works in English, *decide on the feet** does not for no obvious semantic reason. In addition to the collocational selectivity, there is often no logic involved regarding the grammatical restrictions applying to multi-word units. Some idioms, for instance, do not have the full range of grammatical possibilities that a similar non-idiomatic phrase would have (Palmer 1981: 80f); for instance, whereas the idiom *kick the bucket* can be used in different tenses, if the idiomatic meaning of the idiom is to be retained it cannot be used in the passive while *lay down the law* can be. In some other cases the passivization of idioms is questionable; whereas **the beans were spilled* does not sound right at all, *the beans have been spilled* may sound all right, though only to some native speakers. Furthermore, multi-word units are more or less firmly fixed and allow little or no variation in form (see section 2.3.2), e.g. in the idioms *kick the bucket* or *blow one's top*, the nouns cannot be pluralized (Palmer 1981: 80), or the order of the lexical items in the collocations *safe and sound*, *ifs and buts*, *ups and downs*, *first and foremost* cannot be changed nor can their components be substituted by synonymous lexical items. Furthermore, the stylistic restrictions underlying the use of many multi-word units add to their difficulty. Learners need to know when a particular multi-word unit is appropriate to use, i.e. in which register to use it, e.g. *alight*

from a bus is a formal collocation used in notices and other official contexts whereas *get off a bus* is used in everyday language (O'Dell & McCarthy 2008: 6).

Apart from learning all these aspects of multi-word units, active and correct use of them appears to be the major hurdle for the language learner. A number of writers have commented on this: For instance, in the context of collocations Laufer-Dvorkin (1991: 20) states, "since the meaning of most collocations is transparent, the problem occurs mostly on the production level." Similarly, Nesselhauf (2003: 223f) asserts that comprehension of collocations does not normally cause problems for learners but their use does. Support for these parallel claims comes from a number of studies. The results of the experiment reported in Marton (1977: 45-47) is one of the early studies lending support to this claim; although advanced Polish learners were found to comprehend the reading texts containing the elicited English collocations (Marton adopts the term "conventional syntagm"), which they did not have in their productive repertoires before the experiment, they could not produce those same collocations in a follow-up translation task. Based on these findings, Marton's conviction has been that mere exposure to the target language is not enough to gain "lasting learning effect" in the productive use of collocations (Marton 1977: 54). In a more recent study, Nesselhauf (2003) has shown that the length of a learner's exposure to the target language has only a slight effect on collocational accuracy. One reason proposed to explain this observation is related to the aforementioned nature of collocations – their transparency. As they do not cause

any problems to the learner on the receptive level, i.e., in reading and listening comprehension, the learner does not pay conscious attention to collocations as a unit, which in turn prevents their storage as a whole in the mental lexicon. Another reason is related to congruence, i.e. word-for-word equivalence of a collocation in the learner's L1 and L2. As Laufer-Dvorkin (1991: 19) states, “[c]ollocations are problematic when their meaning is apparent at first glance but their constituent elements cannot be given their translation equivalents” – which is mostly the case as the pairing preferences vary from language to language to a great extent (Wray 2002: 73). For instance, whereas in English you *make* or *reach* a decision, you *meet* a decision in German (*eine Entscheidung treffen*), *take/give* a decision in Turkish (*karar almak/vermek*). Even though failures to observe such collocability differences between languages do not necessarily result in errors, they do result in odd expressions (Laufer-Dvorkin 1991: 20). It has long been shown that learners find incongruent collocations more problematic (e.g., Bahns 1993; Biskup 1992). More recently, as will be discussed in greater detail in the next section, the importance of congruence as a source of difficulty for the learner has been clearly demonstrated by a study conducted by Nesselhauf (2005). Investigating the factors correlating with collocation difficulty for advanced German learners of English, the researcher has found congruence as the most important intralinguistic factor (see also Nesselhauf 2003; Wang 2016; cf. Men 2015). The learning burden being heavier in the case of incongruent collocations, teachers

may need to spend some extra time on teaching them (Nation 2006: 449).

As mentioned in the preceding section, there is a consensus that unanalyzed multi-word units provide the “raw material” during early L1 acquisition, which is then segmented into smaller components and analyzed over time to further language development. A commonsense assumption is then that it may also be possible that multi-word units may serve the same purpose for L2 learners – that is, they may also facilitate further language learning for L2 learners by serving as a foundation. Although this assumption has received support for children’s naturalistic L2 acquisition (see Section 2.3.3), it has been controversial for adult L2 learners. Ellis (1996) is one of those who support this assumption: He argues that grammar knowledge in L2 stems from the implicit analysis of the lexical sequences that are acquired by chunking processes by the learners. A number of studies can be cited in support of this argument. Myles et al. (1998), for example, tracked the use of formulas in a classroom setting by 16 beginner learners of French in a longitudinal study with the aim of determining whether the learners could gradually “unpack” the initially unanalysed multi-word utterances and use parts of them productively in generating new utterances. Their results indicate that in the early stages the use of formulas enabled learners easy entry into communication and speeded up the production but later, the realization of learners – conscious or not – that the formulas provided for early classroom exchanges were inadequate for their developing communicative needs triggered the breakdown process of

the formulas. With the pressure of communicative needs, most of the learners gradually analysed their early chunks; rather than dropping the formulas from their speech repertoire as their creative rule-governed competence developed, they started modifying them in a number of ways, which suggests the use of formulas by the learners “as a database for hypothesis testing” (Myles et al. 1998: 359). Having found most of the learners not only gradually “unpack” their early chunks but also use parts of them productively in the generation of new utterances, Myles et al. (1998: 323) sum up their research in the following sentence: “rote-learning of formulas and the construction of rules are not independent processes, but interact and actively feed into one another”. This “dynamic two-way process” has been confirmed by the results of a similar study, focusing on the production of interrogatives by early classroom learners, conducted by the same researchers (Myles et al. 1999). In these two studies, the learners who made the most progress were the ones who both had an extensive databank of formulas which they did not discard and kept actively working on them, tackling their segmentation throughout the data-collection period (Myles 2004: 158f). On the basis of these findings, Myles (2004: 153) asserts that there is a clear correlation between use of formulas and the development of grammatical competence in instructed L2 learning, just as in children’s naturalistic L2 acquisition (Wong-Fillmore 1976; Karniol 1990; Perera 2001). In another longitudinal study, Bardovi-Harlig (2002) found that some learners use formulaic language in the acquisition of future expression prior to creative use (and also beyond the earliest stages). In the data

investigated, the learners break down the formula into smaller parts over time, from the full *I am going to write about* to the core *going to* where not only the verb but also person and number vary. The researcher takes this as evidence in support of the claims of learner production moving along the “formulaic-creative continuum” (Weinert 1995).

There are, however, studies results of which do not accord entirely with the contributive claims of unanalyzed utterances in the linguistic development. For instance, according to a longitudinal study of an adult L2 learner’s language development conducted by Hanania and Gradman (1977), adult language learning proceeds creatively and “memorized items” which are appropriately used in particular social contexts by the adult L2 learner, such as *thank you, I’m sorry*, are perceived as single units and resist segmentation for further analysis (see also Shapira (1978) and Schumann (1978) for similar findings even though their study does not focus on the acquisition of formulaic language). Yorio’s (1989) results point in the same direction and the researcher concludes by stating that “[u]nlike children, they [adult L2 learners] do not appear to make extensive early use of prefabricated, formulaic language, and when they do, they do not appear to be able to use it to further their grammatical development” (p. 68). In another longitudinal study conducted by Schmidt (1983), an adult Japanese learner of English named Wes was reported to demonstrate an extensive use of formulaic sequences in his language but they were mainly used for communicative functions and did not lead to further grammatical development. Schmidt

2.3 Phraseology and multi-word units in native and non-native speaker language and their role in language learning and use

(1983: 150) noted that, on some occasions, even minor changes in formulaic sequences were wrong, and the attempts by Wes to build novel utterances and hypothesis formation were more often than not amiss. For his subject's use of these sequences, Schmidt (1983: 150) commented as follows:

Wes has a rather rich repertoire of formulaic utterances, memorized sentences and phrases [...] which increase the appearance of fluency in English. [...] It is not always clear which of Wes's utterances are memorized wholes, except for those which clearly exceed the limits of his acquired grammatical system, but it is clear that he has chosen this as a major language strategy.

Accordingly to these inconclusive findings regarding the extent and role of multi-word units in adult L2 acquisition, opinions also diverge on whether and to what extent teaching should be based on them. Some express their concern arguing that extensive reliance on formulas in teaching may have a negative effect on learners' linguistic development (e.g., Felix 1981; Lightbown 1983). Still some others suggest that multi-word units should be taught from the very beginning since they provide the raw material for later analysis and segmentation (e.g., Nattinger 1988: 77; Ellis 1996; R. Ellis 2002). More recently Rod Ellis (2005: 210f) has asserted that instruction needs to ensure that learners develop not only a rich repertoire of formulaic sequences but also a rule-based competence.

Even if the assumption that multi-word units facilitate further language learning for adult L2 learners is disproved, one thing with regard to their

role is sure, as rightly pointed out by Schmitt and Carter (2004: 12): The automatic use of formulaic sequences will allow chunking, freeing up memory and processing resources of the learner. This ease of processing must be of a particular value for learners since they commonly necessitate more processing effort to express their intended message (Nesselhauf 2005: 2), mostly struggling with the correct lexical retrieval. Along the same lines, believing that formulaic sequences (“patterns” as they refer to them) are as essential to fluency as to accuracy, Hunston and Francis (2000) state, “[i]t is an unfortunate learner who has to think of every next word separately when uttering a sentence” (p. 271). Thus, learning multi-word units rather than individual words will promote fluency both in speaking and writing by relieving the learner of attending to each word and allowing them to devote instead more attention to the larger structure of the discourse and on the social aspects of the conversation (Nattinger 1988: 77) – as in the case of native speakers (see section 2.3.3). As memorized sequences the learner can utter them without hesitation, which will in turn enable the learner to be perceived as a more fluent speaker. Since, unlike the native speaker, whose hesitational pauses appear between long sequences of words and who uses his/her pauses to conceive the next thoughts, even a relatively fluent learner usually pauses every two or three words as a result of not having an automatic command of a large set of multi-word sequences, presumably using his/her pauses to a great extent for decisions “at fairly trivial word-structure level” (Kjellmer 1991: 124; see Pawley and Syder 1983: 200ff). In short, these sequences can act as, in Dechert’s (1983)

2.3 Phraseology and multi-word units in native and non-native speaker language and their role in language learning and use

terminology, “islands of reliability”⁵³, enabling more fluent language production (see Towell et al. 1996). The facilitating role of multi-word units in the development of fluency in L2 has been shown in a study conducted by Wood (2006): Analysing the spontaneous spoken narrative retellings of ESL learners for ways in which the use of formulaic sequences may have facilitated fluency growth over a six-month period, Wood has found that these sequences can indeed facilitate extended runs of speech by making pauses shorter and less frequent. Boers et al. (2006) have indicated that the use of formulaic sequences is especially beneficial to perceptions of learners’ fluency and range of expression; the number of formulaic sequences used by learners was found to correlate well with oral proficiency ratings.

Besides, as most of the standardized or institutionalized multi-word units cannot be predicted either by grammar rules or by the features of the individual words of which they are composed, having a command of a great number of ready-made phrases can help learners come across as more ‘native-like’ (Boers et al. 2006: 246f), which is, though clearly not an aim of all L2 learners, the ultimate aim of some. Being ‘native-like’ equals being more accurate. As formulaic sequences are retrieved

⁵³ In an investigation of the spoken output of a German learner of English as she narrated a story based on six cartoons, Dechert (1983) observed that although some of the learner’s output was marked with hesitations, fillers, and corrections, she also uttered completely fluent and smooth stretches, which seemed to be formulaic. Dechert called these fluent stretches an “island of reliability” – fragments of speech that the speaker was confident about. Moreover, Dechert believed that they supported the real-time delivery of speech.

from memory holistically, as ready-made chunks, and are believed to be ‘zones of safety’, Boers et al. (2006) hypothesized that the use of formulaic sequences would help learners come across as *accurate* L2 speakers. Nattinger (1988: 77) was of the same opinion when he claimed that the use of multi-word units would both prevent learners from violating some lexical restrictions and enable them to produce less incongruities of register (see also Hunston and Francis 2000: 271). However, the results for the positive influence of formulaic sequences on learners’ perceived accuracy was found not to be statistically significant (Boers et al. 2006: 256). Thus, further investigations are needed in order to determine the role of multi-word units with regard to accuracy.

Having a large repertoire of multi-word units also contributes to efficient comprehension, just as is the case for native speakers (Section 2.3.3). As Hunston and Francis (2000: 270f) point out, the knowledge of phraseological patterns would allow the learner to reconstruct the meaning of words if s/he happens to mis-hear some words in a dialogue. With regard to comprehension, one further gain of the use of multi-word units by learners is for their native speaker interlocutors: Skehan (1998: 38) points out that the target-like formulaic language produced by learners will be more easily understood by native speakers, creating less processing demands in their part. Millar’s (2011) results have given credibility to this claim (see Section 2.3.3). Howarth (1998b: 174) states that non-target-like formulaic language, i.e. lexical mis-collocations such as *suggest a proposal*, could draw the attention away from the core

of the message – away from content to form, which may in turn result in irritation of the recipient (Nesselhauf 2005: 2). In this context, Wray (2002: 143) argues that “[f]ailing to use a nativelike expression can create an impression of brusqueness, disrespect or arrogance”. Although Wray’s statement is a bit excessive, it is quite likely that the use of inappropriate formulaic sequences may lead to occasional misunderstandings and irritations (Cowie & Howarth 1996: 90f).

It should, however, be noted that the processing advantage of formulaic language for nonnative speakers is less clear (e.g., Schmitt 2004; Underwood et al. 2004). This is not surprising given the inherent variability of learner language (Ellis 1994: 22). As pointed out by Wray (2002), it is quite plausible to assume that formulaic sequences stored in the mental lexicon of nonnative speakers show great variety from those of native speakers because nonnative speakers are not “phraseologically virgin territory” (Granger 1998a: 158). That is, the large set of phraseological sequences one has in his L1 will unavoidably affect the acquisition of phraseological sequences in the L2 – in both a positive and negative way (Granger 1998a: 158).

By way of an interim summary, we can conclude that multi-word units are as important for learners as for native speakers with regard to fluency in language production and also to some extent in language processing. They allow communication, though limited, and in turn offer more input possibilities to learners at the beginning of their learning

process. As such, they may help advanced learners seem more ‘native-like’. However, their contribution to accuracy is still unclear.

Despite the given importance of multi-word units for learners, research in SLA has already revealed that foreign language learners have incomplete, in other words, deficient phraseological competence; their phraseological knowledge lags behind their grammatical and lexical knowledge. The language of foreign language learners is often characterized as non-formulaic due to several reasons. First of all, certain types of multi-word units have been found to be underused by learners compared to native speakers; for instance, stereotyped adverb + adjective combinations, such as *acutely aware* (Granger 1998a: 150) or lexical bundles, such as *in the context of* (Chen & Baker 2010: 30). In addition to insufficient uses of multi-word units, the learners tend to have a small stock of formulaic sequences at their disposal that they overuse (e.g., Hasselgren 1994; Granger 1998a; Chen & Baker 2010; Men 2015), make use of deviant lexical combinations (e.g., Granger 1998a; Nesselhauf 2005; Wang 2016), and often display lack of awareness of stylistic connotations and of register restrictions (e.g., Waibel 2007; Chen 2013b; De Cock 2000). The literature is full of claims about the deficient knowledge of language learners regarding the use of multi-word units, in both their speaking and writing. It has also been shown by Siyanova and Schmitt (2008) that even though advanced learners may be able to produce a large number of target-like formulaic sequences (adjective-noun collocations in this case), their judgements of

collocational frequency are not as accurate as native judgements. Moreover, compared to native speakers, they have poorer intuitions regarding (a)typical collocations and are slower in processing both frequent and infrequent collocations.

As mentioned earlier in this chapter (Section 2.3.2), there are various types of multi-word units in different lengths, and their internal syntactic structures vary considerably. While mastery of multi-word units is hard in general, some types have been claimed to be more challenging than others, for instance the mastery of collocations (especially those involving a verb), phrasal/phrasal-prepositional verbs and prepositional verbs – the multi-word verbs of interest in the present study.

The field of phraseology is vast and wide in range, and it has been expanding fast. The rapidly growing number of studies conducted on multi-word verbs with pure pedagogical concerns in recent years, especially in the last decade, can be seen as an indicator of the increasing awareness of attention they deserve in the classroom. Without attempting to be exhaustive, some studies dealing with the learner's knowledge of multi-word verbs will be addressed in subsequent sections of this chapter. The focus will be on the multi-word verb types investigated in the present study – namely, verb-noun collocations, phrasal and phrasal-prepositional verbs and prepositional verbs.

2.3.4.1 Previous findings from collocation studies in learner English

A large body of literature is available on the acquisition and use of collocations by L2 learners. The focus in most of these studies has been the difficulty experienced by learners in their acquisition and usage, attempting to pinpoint possible factors for the deficient and restricted collocational knowledge of learners, including the role of L1 and proficiency. Some studies have focused on the phenomena of over- and underuse, trying to explain how the collocation use of learners differs from that of native speakers. Meanwhile, others have had pure pedagogical concerns. The number of studies conducted in this field is expanding quickly. Therefore, it is important to stress that this paper shall not attempt to be exhaustive here. Instead, in the present context, only a brief review of some research on collocational knowledge of learners is intended, as far as is relevant to the present study.

It has been well documented in literature that even highly proficient learners have difficulty mastering collocations, especially in production of them, although their meanings are almost always transparent. This very feature of them – their semantic transparency – is indeed held responsible for the learner's not paying (enough) attention to the lexical and/or syntactic features of the preferred combinations in the target language. That is, since collocations do not cause any problems for the learner on the receptive level, they do not attract the learner's attention when encountered in the input. Hence, the learner does not make con-

scious efforts to memorize collocations as chunks or prefabricated sequences, often resulting in these collocations passing unnoticed (Biskup 1992: 87). This, in turn, hinders the learner from integrating collocations as wholes into his/her mental lexicon. Even the learners who can handle demanding grammatical structures may fail to use the appropriate verb with a specific noun. Howarth (1998b: 162) illustrates this case with a piece of writing from a student who successfully produced a demanding structure on the grammatical level but constructed an inappropriate verb + noun combination: *reach findings*. Howarth justly points out that although such phraseological errors are not grave in and of themselves and the extent of intelligibility problems they cause may vary, the errors can negatively affect one's writing in terms of effectiveness and "their cumulative effect can be a serious loss of precision" (ibid.). So, too, collocation accuracy has been shown to be a significant predictor in the evaluation of lexical proficiency (Crossley et al. 2014). What is more is that the learners in most cases are not aware of their deficient knowledge in this particular aspect of vocabulary learning (Howarth 1998b: 161; Nesselhauf 2005: 238).

One of the early studies that demonstrated the deficient collocational knowledge of L2 learners was conducted by Farghal and Obiedat in 1995. The researchers administered two questionnaires – one with a fill-in-the-blank task and one with a translation task. The questionnaires involved 22 common collocations on core topics such as weather, food, and color and were distributed to junior and senior English majors and

to English teachers with 5 to 10 years of teaching experience. The results showed that both groups had seriously deficient knowledge of collocations. Especially problematic for both groups was the “unpredictable” collocations, i.e., where there is a divergence between L1 and L2. Due to a lack of collocational knowledge, the subjects resorted heavily to the ‘open choice principle’ by using various strategies of lexical simplification like synonymy, paraphrasing, avoidance, and transfer (Synonymy was the most common strategy adopted, followed by the strategy of avoidance). Farghal and Obiedat (1995) accounted for the use of synonyms or near-synonyms for the targeted lexical item in a collocation (e.g., *steady* color or *stable* color instead of *fast* color) with the failure of recognizing the collocability of words (e.g., *fast*) in stable collocations. They further stated that using synonyms as a lexical simplification strategy was not a result of the absence of the target lexical items in the subjects’ lexicon; contrarily, they put forth that the phraseological failure of the subjects was a direct consequence of the language teachers’ tendency to teach words individually. Therefore, the researchers argued for highlighting of collocational aspects of lexical items in EFL classes alongside the ‘open choice principle’.

Similarly, the early study conducted by Gabryś-Biskup (1990) tested the collocational knowledge of L2 learners. Expanding upon the preceding study, Gabryś-Biskup examined both perception and production of lexical collocations by advanced Polish-speaking learners of English. The study investigated whether collocations posed a problem for L2 learners and, if so, which type of collocations (i.e., verb-noun/noun-

verb/adjective-noun collocations) were particularly challenging to master. Fifty learners were given a list of selected English collocations and asked to provide their Polish equivalents. Meanwhile, Polish equivalents of these English collocations were given to another fifty learners who were asked to provide their English equivalents. The result of the comprehension tests produced 100 percent correct answers. Simply put, there was, as expected, no difficulty for the learners. The results of the production task, however, indicated that learners required to provide English translation equivalents of L1 collocations were faced with significant problems in the production of both verb-noun and adjective-noun collocations. Gabryś-Biskup (1990: 39) noted that in some of the deviant collocations, L1 interference was observable.

In a later study, Biskup (1992) compared collocational knowledge of advanced Polish- and German-speaking learners. The two different language learner groups were asked to supply English equivalents of lexical collocations. The equivalents were then evaluated by three native speakers of English on a 4-point scale – from ‘unacceptable’ to ‘full equivalent.’ The number of correct answers was not significantly different for the two learner groups. However, the results indicated that Polish learners typically refrained from giving any answer at all if they did not know a given collocation while German learners in the same situation tried creative, alternative ways of rendering the meaning of collocations by using definitions or paraphrases, not caring much about the “well-formedness” of their answer (Biskup 1992: 88), e.g., *make the clock working* for *winding a watch* or *to break a nut open* for *to crack*

a nut (p. 89). The researcher explained the discrepancy between the two groups with the different emphasis in EFL teaching traditions in Poland and Germany. Polish education insists on accuracy while, in Germany, education prioritizes fluency and communication. This study also clearly found that the learners had the most difficulty with collocations that included verbs. Concerning this finding, Biskup (1992: 87) stated, “[c]ertain features that verbs impose on nouns such as [+HUMAN] and [+ANIMATE] determine the collocational system of a given language and make it different from that of other languages” and without the experience of a native speaker in lexical patterning in his/her language (which requires a great deal of exposure to the language), it is improbable, if not impossible, for a non-native speaker to be able to produce only the preferred, idiomatic collocations or to predict the co-occurrence restrictions between lexical items in the target language.

Later larger-scale studies focused on the collocation use of L2 learners and also provided evidence for the challenge collocations present to language learners (e.g., Granger 1998a; Nesselhauf 2003, 2005; Altenberg & Granger 2001; Gilquin 2007; Men 2015; Wang 2016; Alangari 2019 – to name just a few). Adopting a corpus-based approach to learner phraseology, Granger (1998a) compared advanced French learners’ use of collocational amplifiers such as *bitterly cold* and *deeply*

in love and the so-called ‘sentence-builders’⁵⁴, with that of native speakers, focusing on the concepts of under- and overuse. Her findings regarding the use of collocations demonstrate that the learners not only underuse native-like collocations, both in terms of the number of types and the number of tokens, but also in the use of “atypical” word combinations (Granger 1998a: 152). It must be mentioned that the researcher made some noteworthy observations: Some collocations used exclusively by native speakers were either “stereotyped” combinations like *acutely aware* or “creative” combinations like *ludicrously ineffective* – types of combinations which were significantly underproduced by the learners (p. 150). Some rare examples of creative combinations were found in the learner corpus, but they were not always successful. Moreover, the few stereotyped combinations found in the learner corpus were usually congruent with the learner’s L1, i.e., they had a direct translation equivalent in French (Granger 1998a: 150). Granger also observed that learners overuse some structures compared to native speakers (namely, active sentence builders such as *I maintain/claim that...*). Based on all her findings, Granger (1998a: 155) thus concluded that not only underuse, or lack of prefabs, but also excessive use can lead to the “foreign-soundingness” of learners’ productions. The findings of a later study conducted by Durrant and Schmitt (2009) agree with those of Granger. Durrant and Schmitt’s study focused on different collocation

⁵⁴ Granger defined these constructions as “phrases which function as macro-organizers in the text” (1998a: 154).

types than the ones investigated in Granger – namely, on directly adjacent premodifier-noun word pairs (including both adjective-noun and noun-noun combinations). In their study, learner groups with different language backgrounds used the same amount of collocations in their writings as native speakers, but they displayed an underuse of low-frequency collocations and “novel” combinations, over-relying on high-frequency collocations. Yoon’s (2016) study substantiated the underuse of low-frequency collocations by learners. The researcher found that native speakers displayed a tendency to produce three times more low-frequency verb-noun collocations in their essays than non-native speakers (Yoon 2016: 53). Alangari (2019) found that advanced Arab learners used more verb-noun collocations than native speakers, but the type/token ratio in their essays was lower, i.e., less lexical variety existed in their essays (p. 69). In Laufer and Waldman’s (2011) study, however, learners of three different proficiency levels displayed a general underuse of verb-noun collocations (see also Li and Schmitt 2009). The discrepancy between studies regarding the quantitative use of collocations can partly be related to different studies applying different methodologies to extract and analyze collocations, partly to different definitions of collocations adopted, and partly to the collocation type(s) investigated. It has already been shown that not all collocation types cause an equal degree of difficulty to L2 learners. Earlier research demonstrated that L2 learners experienced greater difficulty in producing verb-noun collocations than other types of collocations, as seen in Biskup’s 1992 study (see also Howarth 1996; cf. Namwar 2012). In a

cross-sectional study on the development of Greek ESL learners' collocation knowledge, Gitsaki (1999) discovered the existence of a developmental route in the acquisition of collocations; adjective-noun collocations were the "easy" and "early acquired" type of collocations, and verb-noun collocations were the "difficult" and "late acquired" ones. This finding may explain the underuse of verb-noun collocations observed in Laufer and Waldman's (2011) data.

Nesselhauf (2005) conducted one of the most comprehensive studies on the usage of collocations by learners. Basing her study on a corpus of writings by advanced German learners of English (namely, GICLE), Nesselhauf investigated the learners' use of verb-noun collocations qualitatively and quantitatively. Her results clearly showed that the advanced learners experienced great difficulty in producing verb-noun collocations, especially in selecting the correct verb – the most frequent deviant element in the attested collocations (Nesselhauf 2005: 71). Nesselhauf found that approximately one third of the collocations produced by the learners were unacceptable or questionable. A similar amount of deviant verb-noun collocations was reported by Laufer and Waldman (2011) and by Alangari (2019). The most difficulty-inducing factor in Nesselhauf's study was found to be non-congruence between what the learner wanted to express in the L2 and the corresponding L1 phrase; it led to deviation in about half of the instances (Nesselhauf 2005: 238). Later tests carried out by Wolter and Gyllstad (2011) concur with this finding (see also Gilquin 2007). The researchers found that congruent verb-noun collocations were not only processed faster by the learners

but were also recognized more successfully than non-congruent collocations – “L2 only” collocations as Wolter and Gyllstad referred to them. In addition, the researchers showed that learners needed a longer time to determine whether a collocation was correct or not – thus conforming to the suggestion of Siyanova and Schmitt (2008) that learners have poorer intuitions about (a)typical collocations. However, Wolter and Gyllstad noted that non-congruent collocations, once acquired, could be stored and processed as wholes independently of the learners’ L1 (2011: 442). This result is in agreement with Yamashita and Jiang (2010). They showed that both the congruency effect and the amount of L2 input interactively influence the acquisition of L2 collocations. The findings of Obukadeta (2019) add support to these findings. In his investigation of collocations in the writings of Yoruba-speaking Nigerian ESL learners, Obukadeta observed that, in their essays, the participants at all four proficiency levels (A2, B1, B2, C1) had difficulty producing incongruent collocations, using more congruent collocations. Yet, he noticed that frequently used incongruent collocations in the learners’ speech community were less problematic for the learners. He has, therefore, concluded, “frequency trumps incongruency” (Obukadeta 2019: 151, 261). In contrast, Nesselhauf’s study revealed that a large number of collocations identified as problematic for the learners were not specialized or technical collocations; they were instead frequent expressions. Similarly, the bulk of the verbs that were misused or attempted by the learners were high-frequency verbs (Nesselhauf 2005: 244). The

difficulty learners experience in the use of high-frequency verbs in collocations has already been demonstrated by, for instance, Altenberg and Granger (2001) and more recently by Wang (2016).

The repeated observation of the restricted and deficient collocational knowledge of L2 learners has raised the question of whether a learner's knowledge of collocations expands in parallel with his/her knowledge of general lexical words. To answer this question, Bahns and Eldaw (1993) employed a translation task and a cloze task investigating advanced German-speaking learners' productive knowledge of English collocations. The focus of their study was on verb-noun collocations. The results showed that the learners had difficulty paraphrasing collocational phrases and that their collocation knowledge did not improve with the advances in L2 proficiency. Bahns and Eldaw (1993: 108) stated that "learners are more than twice as likely to select an unacceptable collocate as they are to select an unacceptable general lexical word, and that EFL learners' knowledge of general vocabulary far outstrips their knowledge of collocations." Findings in a more recent study conducted by Laufer and Waldman (2011) substantiate Bahns and Eldaw's (1993) results. Comparing verb-noun collocations produced by Hebrew-speaking learners of English at three proficiency levels, Laufer and Waldman reported that collocational knowledge did not show any improvement with increased proficiency level. Although learners at all levels of proficiency produced deviant collocations in their writings, the learners at intermediate and advanced levels produced significantly more deviant collocations than beginner learners. To be exact, more

than one third of all recurrent errors in their data were produced by learners at an advanced level. According to Laufer and Waldman, the percentage of recurrent errors can be attributed to a presumed higher degree of learner confidence, this confidence resulting in attempts to use more collocations (2011: 665). The researchers concluded that “not only is there no decrease in the number of errors with growth in proficiency, but as proficiency increases, the frequency of errors increases” (ibid.). Obukadeta’s (2019) study lends support to the conclusions obtained by Laufer and Waldman. His findings revealed that the most proficient learners – who produced a higher number of acceptable collocations in their essays than other learners at lower proficiency level – also produced the highest number of deviant collocations (p. 262). This result goes hand in hand with the reported increase in the number of incongruent collocations relative to the increased proficiency of the learner (p. 261). Regarding this pattern in the data as a positive developmental process, Obukadeta (2019: 261) stated the following:

it means the learners buoyed by their increasing proficiency were willing to take risk in their output resulting in the production of more collocations – many of them acceptable and some unacceptable. On the contrary, the least proficient learners stayed in their comfort zone which means fewer collocations were produced and fewer collocational errors were made.

Moreover, as rightly interpreted by Obukadeta (2019: 174, 261), the fact that learners with increased language proficiency produced more incongruent collocations and less congruent collocations indicates that

their reliance on L1 lessens as they become more proficient learners. In the context of proficiency, Obukadeta's study offers another finding worthy of mention. Here, as the proficiency levels of the learners increased, they more frequently used verb-noun collocations of increased structural complexity (i.e., long span collocations) and collocations with figurative meaning. In summary, we can conclude that Obukadeta's study reveals a strong and complex link between proficiency and the production of collocations by learners.

Recently, Wang (2016) has provided a comprehensive overview of learners' use of delexical verb-noun collocations – the verbs on which learners over-rely and tend to use incorrectly. In her corpus-based study, Wang investigated two learner groups with different L1s – namely, Chinese- and Swedish-speaking advanced learners of English – and compared the learners' use of delexical verb-noun combinations with target language data. Yet, unlike most studies, the learner data was also compared with data representing the learners' L1s in efforts to uncover possible traces of L1 influence and to find out whether the problems caused by collocations are universal or are characteristic of a particular learner group due to L1 influence. Wang's results confirmed the repeatedly asserted L1 influence on the use of delexical verb-noun collocations (2016:124 and *passim*). However, Wang (2016: 47) noted that this influence is by no means a straightforward, systematic one, lending support to Yamashita and Jiang's (2010) findings. Several factors – such as, among others, learners' proficiency, learner perceptions of the typological distance between their L1 and the TL, and the frequency or

prototypicality of a given collocation – appeared to interact dynamically on a learner’s decision as to whether or not to rely on their L1 (Wang 2016: 125). Wang’s results also revealed an overuse of some delexical verb-noun combinations in the learner data. Each learner group showed different preferences regarding the high-frequency verbs, and the repeatedly used collocations by the learners were often among those that occurred fairly frequently in the TL data (cf. Juknevičienė 2008: 5). This finding indicates, as interpreted by Wang (2016: 162), the possibility that the learners may have acquired these common collocations as a result of sufficient input, resulting in their storage and processing as single units. In other cases, however, the learners applied the ‘open choice principle’ in some verb choices, especially in the case of less-restricted collocations. One particularly significant insight in Wang’s (2016) study is that learner groups with different L1s may experience different difficulties during the process of learning collocations. In this regard, different L1 learner groups may require special assistance and guidance.

Recently, one other extensive study on the collocational competence of L2 learners was conducted by Men (2015). Men explored the written performance of Chinese EFL learners at three different proficiency levels. He investigated three factors that are associated with the “stagnant development of collocational knowledge” in order to understand the process of collocation learning – namely, vocabulary growth, congruency, and L1. Three types of collocations were investigated in this study: verb-noun, adjective-noun, and noun-noun collocations. The

overall results are in line with the earlier studies on the collocational competence of learners. The results showed, among others, that the learners over-relied on a small number of common collocations, producing an overall lower number and less diversified collocations (Men 2015: 85 and *passim*). Significantly, nearly a quarter of all of the collocations evidenced erroneous uses, and these erroneous uses occurred in varying degrees at all three levels. Furthermore, with increasing learner proficiency, no sign of a decrease in the frequency of collocational errors was observed. This finding confirms what has repeatedly been asserted in earlier studies: Collocations pose problems to all learners, even those at advanced levels. However, it is worth mentioning that Men observed both qualitative and quantitative development in the use of lexical verb-noun collocations in correlation with increasing learner proficiency. Men declared, “with the learning of more verbs and nouns, the possibilities of combining them into well-formed collocations increase, but errors increase sharply as well” (2015: 96f). The study also showed that the learners had a poorer performance on verb-noun collocations, manifesting a preference for the ‘open choice principle’ in their production and that learner knowledge of verb-noun collocations did not show a parallel increase with learner proficiency level. This claim supported the earlier reported difficulty of the verbal element in collocations. Alangari (2019: i), for instance, reported that more than a quarter of the verb-noun collocations attested in the essays of advanced Arab learners were deviant. Men accounts for the difficulty of verbs in collocational constructions with the fact that verbs have more synonyms compared

to adjectives and nouns – a markedly different explanation than the one proposed by Biskup (1992). Men states, “for word classes where there is little increase in a synonym set, collocation errors are seldom made (as for adjective and nouns in AN [adjective-noun] and NN [noun-noun] collocations); where there are increases in words in synsets, chances of errors subsequently increase (as for verbs in VN collocations)” (2015: 138). This finding supports Laufer’s (1990b: 585) claim that synonymy is one of the difficulty inducing factors in learning words, creating “an unnecessary load” (see Webb & Kagimoto 2011). Moreover, Men asserted that learning new verbs in a particular semantic domain inhibited the learning of collocations. He explained this assertion as follows:

learners may only have an incomplete command of the semantics of the new verb, i.e. basic meaning of that verb is acquired but not its distinguishing features as distinctive from a set of semantically related verbs (Men 2015: 160)

In Men’s data, L1 transfer was most traceable in the use of verb-noun collocations compared to the other two types of collocations investigated (Men 2015: 162). Similarly, Alangari (2019: 70) reported that L1 influence could be traced in nearly half of the deviant verb-noun collocations produced by the learners. Men rightly believes that “collocation acquisition requires complete acquisition of the semantics of a word” (2015: 162).

Men's findings regarding congruency refute the findings of earlier studies; for the Chinese learners of English, congruent collocations were found to be more prone to error (cf. Nesselhauf 2005; Yamashita & Jiang 2010; Wolter & Gyllstad 2011). Men attribute this finding to the high level of lexical difference due to typological distance between the learners' L1 and the TL. In effect, few one-to-one correspondences and common partial congruence, i.e., differentiation (one-to-many correspondence) and coalescing (many-to-one correspondence), occur between the two languages (Men 2015: 161f). Men (2015: 162), however, asserts that once non-congruent collocations are acquired, they are seldom prone to error, confirming earlier claims (e.g., Yamashita & Jiang 2010; Wolter & Gyllstad 2011).

As the foregoing brief review demonstrates, the studies that explored the use of collocations in learner language and tested the learners' collocational knowledge differ to a great extent from one another with respect to their size/scope, nature, and design – they have ultimately focused on different collocation types and investigated differing data types in terms of both task and learner variables with varied methodologies.⁵⁵ Even the definition of collocation itself varied between studies, and, in some, as noted by Nesselhauf (2003: 224), the concept of collocation itself remained “hazy”.

⁵⁵ The table in Goulart (2019: 7-10) provides a clear overview of some studies conducted on the use of collocations, displaying the variety in the field.

Although it is hard to compare the findings of studies directly and to make a generalization with certitude, on the basis of all the above data one clear finding emerges: Even at advanced levels, mastery of collocations remains a major hurdle for learners, regardless of their L1 background. The degree of the difficulty they cause varies depending on both the type and frequency of a given collocation as well as learner-related variables, such as the typological distance between the learners' L1 and TL determining the level of congruency, psychotypology, and proficiency in the TL. Despite the differences across studies, recurrent findings have been surfaced. First, learners display a general underuse of collocations in their language compared to native speakers (e.g., Hasselgren 1994; Laufer & Waldman 2011; Howarth 1996; Granger 1998a; Li & Schmitt 2009; Wang and Shaw 2008), and they exhibit an over-reliance on a small number of common, high-frequency collocations. These learner tendencies go hand in hand with an underuse of low-frequency, but strongly associated, collocations (e.g., Hasselgren 1994; Granger 1998a; Bestgen & Granger 2014; Yoon 2016; Durrant & Schmitt 2009; Wang 2016). Second, the verbal unit in collocations causes the most difficulty to learners (e.g., Nesselhauf 2003, 2005; Biskup 1992; Men 2015). Some contradictory results have also emerged, which is not unexpected given the differences between studies outlined earlier. For instance, in some studies, L1 influence was reported as subtle and not straightforward (e.g., Wang 2016; Men 2015), while in others, the influence was found to be strong (e.g., Granger 1998a; Nesselhauf 2005; Alangari 2019; Obukadeta 2019;

Juknevičienė 2008). Moreover, some researchers suggested that congruent collocations proved, in general, easier to learn than other researchers claimed (e.g., Granger 1998a; Nesselhauf 2005; Wolter & Gyllstad 2011; Bahns 1993). Other researchers alleged that non-congruent collocations are less problematic (e.g., Men 2015). Although some studies showed language proficiency to be positively related to the qualitative and/or quantitative production of collocations (e.g., Laufer & Waldman 2011; Obukadeta 2019; Men 2015), others observed no relation between language proficiency and use of collocations (Bahns & Eldaw 1993; Howarth 1996). This discrepancy between the studies regarding the link between proficiency and the collocation use is not surprising given the fact that language proficiency of the learner has been variously defined in different studies – in terms of the number of years of study, the age of the learner, or the Common European Framework of Reference (CEFR) levels (Paquot & Granger 2012: 137). Such methodological discrepancies and contradictory results in the field call for systematic replication of earlier studies in order to broaden the data sets for further analysis.

2.3.4.2 Previous findings from studies focusing on phrasal and phrasal-prepositional verbs in learner English

Considerable research is available on phrasal verbs studied from an SLA perspective. Some of the early studies, which are empirical in nature, investigated phrasal verb use by learners with a special focus on the phenomenon of avoidance and the role of L1 in this avoidance behavior. With the availability of computerized learner corpora, which

provided a solid empirical basis to study learner language qualitatively as well as quantitatively, it became possible to conduct larger-scale studies, going beyond the issues of avoidance and L1 influence. Whereas the focus in some of these studies has been the difficulty experienced by learners in learning them, some had pedagogical concerns. In the following a selected number of studies and their main findings will be reported, without aiming at providing an encompassing overview of learner-related research in the field.

One of the early studies on phrasal verb use was conducted by Cornell in 1980 – reported in a later article published in 1985, in which the aim was to establish German learners' productive knowledge of 60 selected idiomatic phrasal verbs. The participants were given two sentences; one including a phrasal verb construction where the particle was missing, one with a paraphrase of the phrasal verb intended in the first sentence. The participants' task was to insert the appropriate particle(s) from a list of twenty-one particles supplied. The results of the test revealed that the average score of correct answers was approximately 25 percent with great individual variation for different combinations, the highest being 87 percent and the lowest 0 percent (Cornell 1985: 271). Even between advanced learners the confusion between phrasal verbs was very frequent, e.g., *look for* – *look after*, *make out* – *make up* (p. 273). Cornell concluded that the learning of phrasal verbs at school and university was generally not very successful and explained this failure with students' limited contact with phrasal verbs (p. 273).

The well-known study conducted by Dagut and Laufer (1985) was the first study dealing explicitly with the notion of avoidance in phrasal verb use. Based on Schachter's (1974) assertion that error analysis should examine not only what is actually produced by L2 learners in their attempts to express themselves, but also what is not produced, and on Kleinmann's (1977) observation that only those items can be avoided that are actually known to learners, Dagut and Laufer hypothesized that Hebrew learners of English would avoid the active use of phrasal verbs, while at the same time being passively familiar with them, since the lack of parallel structures in L1 would provide the Hebrew learners with no patterns for transfer. After having identified 15 phrasal verbs preferred by English native speakers over their one-word 'equivalents' (e.g. *let down* vs. *disappoint*), the researchers tested the learners' active use of phrasal verbs both in free expression and in elicited responses – namely, in a multiple choice test, a verb translation test, and a verb memorizing test. The results showed that the majority of the learners, despite being familiar with the phrasal verbs as such, exhibited a strong preference for one-word verbs where English speakers chose the phrasal verbs. Avoidance of phrasal verbs was highest for the figurative and lowest for the literal phrasal verbs. Dagut and Laufer attributed this finding to L1-L2 structural differences, i.e. the lack of phrasal verb equivalent in Hebrew. Interpreting their results as evidence for the dominant role played by the learners' L1 in the process of learning L2, the researchers concluded that the notion of avoidance should be found among other English learners of non-Germanic languages

other than Hebrew since the phrasal verb construction is a “peculiarity of Germanic languages” (Dagut & Laufer 1985: 78).

Its importance notwithstanding, there exist a number of caveats that need to be noted regarding Dagut and Laufer’s study. It is unfortunate that the knowledge of the learners with regard to the phrasal verbs under scrutiny had not been tested beforehand: Based on their teaching experience the researchers only assumed that the learners had encountered these verbs at some point in their education (Dagut and Laufer 1985: 75). That makes impossible to determine whether it was really a case of avoidance at hand; the underuse of phrasal verbs observed could just as well have been caused by the learners’ ignorance since only the features that are known to the learner can be avoided. Furthermore, apart from L1-L2 difference, no other possibly influential factor was taken into consideration in the interpretation of the results – not even the semantic difficulties of phrasal verb types even though less figurative phrasal verbs were used less often than literal ones by Hebrew learners. Despite these weaknesses – maybe because of them – the study has provided impetus for further research into the use of phrasal verbs by learners coming from various L1 backgrounds.

A corollary derived from Dagut and Laufer’s (1985) conclusion was that learners with a Germanic language would not avoid phrasal verb constructions. In a follow-up study, Hulstijn and Marchena (1989) tested this corollary on the grounds that one other, maybe more possible, reason behind the Hebrew learners’ avoidance behavior of phrasal

verbs might be the L2-inherent semantic difficulties rather than structural difficulties (lack of phrasal verb construction in L1) because phrasal verbs usually have a specific, sometimes idiomatic, meaning compared to their one-word equivalents, which have more general, multi-purpose meanings (p. 242f). Hulstijn and Marchena argued that the finding that Hebrew learners used literal phrasal verbs more frequently compared to figurative and completive phrasal verbs – in which the particle describes the result of the action, e.g. *cut off* – could be seen as an indication of the fact that the learners' avoidance behavior might have been related to the semantic difficulties of these verbs. The researchers, therefore, hypothesized that Dutch learners of English, despite the similar construction in their L1, would also avoid phrasal verbs as a result of their semantic difficulty and the tendency to avoid phrasal verbs would decrease with increasing proficiency. In order to test these hypotheses, Hulstijn and Marchena administered three tests (multiple choice test, memorization test, and translation test – the same test types used in Dagut and Laufer's (1985) study) – to intermediate and advanced Dutch learners of English. Each test contained 15 sentences, eliciting preference for either a phrasal verb or an equivalent one-word verb. Results showed that intermediate Dutch learners used fewer phrasal verbs than advanced learners but in comparison to Hebrew learners in Dagut and Laufer's (1985) study, both intermediate and advanced Dutch learners used more phrasal verbs. That is, neither advanced nor intermediate Dutch learners avoided phrasal verbs as a category. However, they avoided the idiomatic phrasal verbs that they

perceive as too Dutch-like, i.e. similar in form and meaning. Moreover, the learners, particularly intermediate learners, displayed a tendency to adopt a “play-it-safe strategy” by preferring one-word verbs with general meanings to phrasal verbs (Hulstijn and Marchena 1989: 241). The researchers, thus, concluded that English phrasal verbs as a morphological verb class did not constitute a learning problem for Dutch learners but some phrasal verbs might be avoided as a result of a perceived semantic difficulty (1989: 251).

A few years later, Laufer and Eliasson (1993) undertook a study to determine which of the three sources purposed by the previous research could best predict the avoidance behavior in the use of phrasal verbs by advanced learners: (a) L1-L2 difference, (b) L1-L2 idiomatic similarity, and (c) inherent L2 complexity. The participants, two groups of Swedish advanced learners of English, were given a multiple-choice test and a translation test – the two tests included the same 20 sentences. The overall results showed that advanced Swedish learners did not exhibit any kind of avoidance behavior in the use of phrasal verbs, not even when the verbs were figurative and similar to their L1 equivalents – unlike the Dutch learners in Hulstijn and Marchena’s (1989) study. Thus, according to the results, neither L1-L2 similarity nor inherent L2 complexity alone could predict the avoidance behavior in learners; only in combination with the L1-L2 difference, inherent L2 complexity would result in avoidance (Laufer and Eliasson 1993: 45). The researchers have thus stated that “[t]he only factor that in itself is conducive to avoidance is consequently L1-L2 difference” (ibid.), lending credibility

to the claim that structural differences between L1 and L2 can be impedimental in the active use of phrasal verbs. However, these results do not tally well with another study conducted by Linnarud (1986), whose participants also had Swedish as their L1. Linnarud compared compositions written by 17-year-old Swedish-speaking learners of English, who had been studying English for 9 years, with those written by native English speakers of the same age. The focus of the study was on lexis in general. The results displayed clear differences between the two groups and one of these differences was related to the verb usage: “One important area of verbal usage almost totally neglected by SL [Swedish learners] is that of multi-word verbs or verbs where an adverbial or prepositional particle forms a semantic and syntactic unit with the verb” (Linnarud 1986: 68). This discrepancy between the results obtained in Laufer and Eliasson’ study and those in Linnarud’s may be the consequence of different levels of language proficiency or it may have to do with the study design (elicitation data in the former, free composition in the latter).

Building on the previous three studies mentioned (namely, Dagut & Laufer 1985, Hulstijn & Marchena 1989 and Laufer & Eliasson 1993), Liao and Fukuya (2004) conducted a study focusing on the phenomenon of avoidance but this time the participants were Chinese-speaking intermediate and advanced learners of English. Most of the intermediate learners had not been in an English-speaking country and the ones who had been abroad stayed only for a short period of time (7 for 5 months and 3 for nearly 9 months). All of the advanced learners spent at least

more than nine months in the U.S. As in earlier studies, research interests lay in the role of the learners' L1, phrasal verb type as well as proficiency of the learners but there was an additional factor included, which had not been addressed previously: the influence of test type. The same three test types were used as in the previous studies mentioned – namely, multiple choice, translation, and recall tests. The results showed that proficiency level of learners, phrasal verb type and test type play an influential role in the learners' avoidance behavior. In all three tests, intermediate Chinese learners were found to display a preference for one-word equivalents of phrasal verbs, avoiding the use of phrasal verbs. Compared to advanced Chinese learners, who almost produced as many phrasal verbs as native speakers, intermediate learners produced a significantly low number of phrasal verbs. Liao and Fukuya attributed the intermediate learners' avoidance to the lack of phrasal verb structure in Chinese. However, since no avoidance behavior could be reported for the Chinese advanced learners, Liao and Fukuya (2004: 211) stated that “learning seems to have counteracted the effects of the L1-L2 difference for the advanced learners of English.” Therefore, their results lent only partial support for the idea put forward by Dagut and Laufer (1985) that the structural differences between L1 and L2 can predict avoidance behavior in learning L2. As for the factor of phrasal verb type, Chinese learners, as in the previous studies, were found to use literal phrasal verbs more frequently than figurative (i.e., idiomatic) ones, regardless of their proficiency level. The learners at intermediate level, however, used an even lower number of figurative phrasal verbs

than advanced learners. Incorporating their findings with the ones obtained in the previous studies, Liao and Fukuya (2004: 212) have argued for “a developmental manifestation of interlanguage from avoidance to nonavoidance.”

Liao and Fukuya (2004) also found an correlation between test type and phrasal verb type and this correlation occurred only in the translation test, where no cue with regard to a possible verb choice (neither phrasal verbs nor their one-word equivalents) was available to the learners. That is, only in the translation test Chinese-speaking learners (both advanced and intermediate) displayed a tendency to use less figurative phrasal verbs than literal ones. The researchers took this finding as evidence in support of the claims of L2 semantic complexity in the avoidance of phrasal verbs (Liao & Fukuya 2004: 216).

One other rather early study dealing with the use of phrasal verbs by learners was conducted by Sjöholm (1995). This study built on previous research but by exploring a wider range of factors likely to exert influence on the learners’ use of phrasal verbs, it aimed at explaining the underlying mechanisms in the SLA process. To put it more precisely, Sjöholm was interested in finding out the effects of different learner-internal and –external factors, such as the learner’s L1 (L1-L2 distance), the learner’s proficiency, other known languages (L3), the quantity and quality of input (i.e., natural and classroom input) and semantic properties of phrasal verbs (i.e., transparency vs. idiomaticity) on the learn-

bility and transferability. By means of a multiple-choice test that comprised 26 items, Sjöholm elicited empirical performance data from Finnish- and Swedish-speaking Finns at intermediate and advanced level. Sjöholm chose these two particular groups (namely, Swedes and Finns in Finland) to study phrasal verb use for two reasons. Firstly, both groups were considered to be fairly comparable with respect to their cultural and educational background (Sjöholm 1995: 131) and, secondly, the two languages differed structurally in terms of phrasal verb constructions – whereas Swedish has a phrasal verb equivalent, Finnish does not. Sjöholm's results validated the impact of the structural L1-L2 differences on learning of phrasal verbs, as proposed by Dagut and Laufer (1985) and Laufer and Eliasson (1993). Although both learner groups used fewer phrasal verbs than native speakers, Finns used significantly fewer phrasal verbs than Swedish-speaking Finns ('Swedes'), especially at the intermediate level. They also committed significantly more errors than Swedes, a finding attributed to the advantage of Swedes over Finns with the overt similarity between Swedish and English. Furthermore, although both groups showed a preference for transparent phrasal verbs over idiomatic ones, the Finns were found to underuse (or avoid) idiomatic phrasal verbs to a greater extent, especially at lower level proficiency. These findings converge with the previous studies that support the proposal of the structural and semantic

L1-L2 distance being a determinative factor in L2 learning. The distance impedes acquisition especially at lower levels of proficiency, as shown also in Liao and Fukuya's (2004) study.

Sjöholm's study also demonstrated the importance of input, regardless of whether it is formal or natural, in the development of competence in phrasal verbs: The more input the learners had received, i.e., the longer they had studied English, the more native-like was the performance. However, in the case of non-Swedish-based phrasal verbs, the Finnish learners who had only been exposed to classroom input avoided and underused phrasal verbs more than the Swedes who had been exposed to the same kind of classroom input, showing again the language distance factors (Sjöholm 1995: 227). It is worth noting that, despite the similarity between Swedish and English, Swedes also displayed a tendency to underuse or avoid phrasal verbs compared to native speakers. Sjöholm gave two possible reasons as an explanation for this observation: One is that "phrasal verbs may be less attractive because they are very often semantically opaque and thus perceived as language-specific" (1995: 227). The other explanation, which is equally reasonable, is that "the equivalent one-part verbs constitute a more plausible choice to many learners because they have a broader and more general meaning" (ibid.). Taken together, these results mean that experience with the phrasal verb construction in L1 is an influential factor in their learning

in L2 but their learning can only be explained by the interplay of input, cross-linguistic and semantic factors.

A commonality of all these early studies on phrasal verb use is the fact that their data was obtained by means of elicitation – a useful technique when one is interested in the avoidance phenomenon but not quite so for the investigation of learners’ active knowledge in an ‘unbiased’ way, unlike free production data (Waibel 2007: 27). A similar corpus-based study to Linnarud’s (1986) was conducted by Yorio (1989), in which free written productions of learners – essays written by 26 ESL learners who had spent between five and seven years in the United States – were taken as the database. The focus of the study was on the role of conventionalized language and the notion of idiomaticity in late second language acquisition in general and phrasal verbs were only one aspect of the study. The results with regard to the use of phrasal verbs by advanced learners showed that the difference between the learners and native speakers was rather qualitative. Although the learners used nearly as many phrasal verbs as the native speakers, contradicting Linnarud’s (1986: 68) results, the learners used a lot less idiomatic phrasal verbs than native speakers. However, only 59 percent of the phrasal verbs produced by the learners were correct despite the time spent in the L2 environment (Yorio 1989: 65). Unfortunately, there are no de-

tails with regard to which phrasal verbs were used or what types of errors in the use of this multi-word verb category were committed. Moreover, this study was limited by its small sample size.

Siyanova and Schmitt (2007) applied a combined approach to explore the choice made between multiword verb and one-word verb alternatives. By means of both elicitation data and corpus analysis, the issue of avoidance was investigated further, with a focus on the possible impact of exposure to L2 in its natural environment and the comparison between written and spoken modes of language – CANCODE (Cambridge and Nottingham Corpus of Discourse in English) and BNC were used as the database representing the native speaker English, the former providing spoken, the latter written native English data. The learner language data was obtained from the ICLE. In the empirical part of the study, a questionnaire including short dialogues in an informal tone with an option between a one-word verb and a multi-word verb was given to a group of native speakers and advanced English learners with a non-Germanic L1. Both groups of participants were university students or recent university graduates, who used English on a daily basis. The results of the questionnaires showed that native speakers displayed a statistically significant tendency to use the multi-word verb option of the verb pair than the one-word verb alternative in spoken, informal contexts. So did the learners but to a lesser extent, albeit still statistically significant. Compared to native speakers, the advanced learners were

more likely to use the one-word verb option. As for the impact of exposure to English in its natural environment, the findings reported by Siyanova and Schmitt (2007) were striking: An extended stay in an English-speaking environment (over 12 months) decreased the likelihood of using one-word verbs although it has no significant impact on the likelihood of using the multi-word verb alternatives, contrary to the results found by Liao and Fukuya (2004). As the learners who spent the longest time in the L2 environment made less use of one-word verbs compared to those learners who had no exposure to a natural L2 environment or less than 12 months, Siyanova and Schmitt (2007) have argued for the possibility that there may be “a threshold of input-rich experience which is necessary to begin relying less on one-word verbs” (p. 130). The data analysis of the same verb pairs in the native speaker corpora confirmed the well-established fact that multi-word verbs are more common in spoken than written discourse. At the same time, it also showed that one-word verbs were preferred to their multi-word verb alternatives in both written and spoken discourse, i.e. in the BNC and CANCODE corpora. As for the learners’ use of verbs in their written production, it was found that they showed a tendency to choose the one-word verb alternative in nearly half of the tested verb pairs, supporting the results of Dagut and Laufer (1985). Siyanova and Schmitt (2007: 132f) provided three possible reasons for the learners’ tendency to use fewer multi-word verbs: a) the complexity of multi-word verbs, b) a cross-linguistic reason, i.e. multi-word verbs are “alien” to the learners with non-Germanic L1, which may result in learners needing

more time to “overcome their discomfort” with these verbs, c) as one-word verbs were found to be more common than their multi-word verb ‘equivalents’ in both the BNC and CANCODE corpora, learners’ underuse of multi-word verbs might simply reflect the input they receive, rather than learners’ active avoidance of multi-word verbs – all of which undoubtedly have a role to play in learners’ behavior.

In 2007 another large-scale and comprehensive corpus-based study on phrasal verb use by learners was carried out: namely, the one by Waibel (2007). After having carried out a general survey investigating the eleven sub-corpora recorded on the first edition of the ICLE-CD (see Section 4.2) with respect to quantitative use of frequent phrasal verbs, based on a list of 72 phrasal verbs, Waibel focused on two groups of advanced learners of English (German- and Italian-speaking).⁵⁶ By approaching the data both quantitatively and qualitatively and following Granger’s Contrastive Interlanguage Analysis (1996b), Waibel compared these two learner groups’ performance with that of their native speaker counterparts (LOCNESS). The results of the general survey rejected the initial hypothesis of a universal underuse of phrasal verbs in advanced learner writing (Waibel 2007: 76). Although most learner groups used fewer phrasal verbs than native students, with strong variation between the learner groups, Dutch and Polish learners produced

⁵⁶ Some preliminary findings of this general survey were already reported in Waibel (2005).

nearly as many phrasal verbs as native speakers (p. 77). German learners, on the other hand, were found to use more phrasal verbs than native speakers (ibid.). When the performance of learners whose L1 stemmed from the same language family as English was compared, the results obtained showed positive L1 influence. While learners with a Germanic language background produced almost as many phrasal verbs as native speakers, Finnish learners and those with a Slavic language background used phrasal verbs to a lesser extent, i.e. about 300 phrasal verb tokens less. The learners with a Romance native language, on the other hand, only used half the number of phrasal verbs compared to native speakers. Waibel (2007: 78), thus, stated that “[a]lthough one can speak of an underuse by students with a Finnish and a Slavic background, the extent of underuse is much more pronounced in the writing of students with a Romance background.” The lack of similar verb type in Romance languages and the tendency to use more Romance-based verbs in English were considered to be the two possible reasons behind the overall low frequency of phrasal verbs spotted in the language productions of the learner groups with a Romance background (ibid.) As for the fact that Slavic learners did not display an underuse of phrasal verbs despite their unfamiliarity with this verb type in their L1, Waibel proposed successful teaching as a probable reason (p. 79). Based on the results of this preliminary survey, the researcher concluded that although individual items were used less frequently across all the learner corpora investigated, “variation within the individual learner groups is too strong to

speak of a universal underuse of specific phrasal verbs” (Waibel 2007: 82).

Waibel’s interests went beyond the extent of L1 transfer in phrasal verb use. Focusing on advanced German and Italian learners of English, she investigated other possibly influential factors in phrasal verb use and the features of unnaturalness in learner language in relation to the use of this verb type. Some clear findings emerged from her study: L1 influence – both positive and negative – was observed in both learner groups’ performance. As mentioned in the preceding paragraph, whereas German learners, familiar with a similar verb type from their L1⁵⁷, produced a larger number of phrasal verbs in their writings compared to even native speakers of English, Italian learners underused them, which was attributed to the learners’ greater reliance on Latinate verbs. However, this divergence between the two groups was only related to overall frequencies; the ratio of phrasal verb types and phrasal verb tokens in both learner data was similar. Whereas the Italian learners’ essays displayed topic-sensitivity to a greater extent (i.e. the most frequent phrasal verbs used in the Italian data were highly topic-sensitive), German learners’ essays displayed style-insensitivity – i.e., they used both formal and informal phrasal verbs within the same essay. Out

⁵⁷ Although German does not have phrasal verbs as such, it has particle verbs, which are similar (see Waibel 2007: 38-40).

of the possibly influential internal and external factors in the use of phrasal verbs, three factors were found to be determinative for German learners, but not for Italian learners: time pressure, the use of reference tools, and the years of learning English. As we will have the occasion to note in Section 5.3.4, the amount of time spent in an English-speaking country, however, was found to have a positive correlation with phrasal verb use for both learner groups; in fact about 20 percent more phrasal verb use (Waibel 2007: 105ff.). In terms of the qualitative use of phrasal verbs, no impact of L2-exposure on learner productions was found (Waibel 2007: 108f). Writings of the two groups shared some features of non-nativeness, such as collocational deviations, inappropriate phrasal verb choices in a given context, and the tendency of using verbs of general meaning with a particle instead of a more appropriate verb – what Waibel referred to as “simplified” use of phrasal verbs (see Section 6.2). Cautiously interpreting her findings, Waibel (2007: 162) has concluded that language distance between the learners’ L1 and L2 may impede the successful learning of phrasal verb use.

Gilquin’s (2011) study gave support to Waibel’s conclusion. Focusing on the phrasal verbs constructed with the particle *up*, Gilquin examined the spoken and written language of both ESL learners (learners of English in Kachru’s outer circle) and EFL learners (learners of English in Kachru’s expanding circle) in comparison to native speaker language. Results revealed differences among EFL learners coming from different

mother-tongue backgrounds – which Waibel (2007) had already noted in the general survey she conducted. Agreeing with Waibel (2007), Gilquin (2011) attributes the discrepancy between EFL learner groups with regard to the quantitative use of phrasal verbs to L1 influence and the degree of exposure to the TL. The results of Gilquin’s study also indicated that EFL learners are not sensitive to register variation; they tend to overuse phrasal verbs in writing while underusing them in speech to a great extent while their native speaker counterparts do the opposite (see also Siyanova and Schmitt 2007). A later study which focused on the use of phrasal verbs by higher-intermediate to advanced French-speaking learners of English by Gilquin (2015) yielded similar results; despite the general underuse of phrasal verbs in both their writing and speech, the French-speaking learners made more use of phrasal verbs in writing rather than in speech, in stark contrast to native speakers, who used considerably more phrasal verbs in speech than in writing (Gilquin 2015: 66). Moreover, French learners also displayed mixed uses of formal and more informal phrasal verbs in their writing, showing close similarity between their choice of phrasal verbs in their writing and speech (Gilquin 2015: 77). The researcher attributed the stylistic discrepancy between learners and native speakers to “the learners’ failure to recognise the spoken-like (and often informal) nature of most phrasal verbs and/or to their lack of automaticity in the production of phrasal verbs under unplanned conditions” (Gilquin 2015: 81). The learners’ lack of stylistic awareness regarding phrasal verbs has already

been reported and received support in other studies (e.g., De Cock 2005; Waibel 2007; Chen 2013b; cf. Hägglund 2001).

Gilquin (2011) found other issues with regard to the use of phrasal verbs by learners; for instance, non-standard uses. They were spotted both in ESL and EFL varieties – some of which were common in both varieties but some occurred only in learner Englishes. For instance, cases of misplacing particles were observed only in learner English while use of redundant particles was found in both varieties, such as *complete up* in Indian English, *increase up* in Chinese English. There were also uses of phrasal verbs in the data that were not listed in common dictionaries of phrasal verbs – what she referred to as “innovations”. While some of these innovations were simply incomprehensible (e.g., *flap up*), others were perfectly understandable, affording evidence for the creative skills of the learners (e.g., *meddle up* and *fashion up*). Moreover, some of these innovations were shared in ESL and EFL varieties, e.g., *to cope up (with)*.

Use of redundant particles and non-standard uses of phrasal verbs were also reported in Mazaherylaghab’s (2013) corpus-based study on Iranian learners of English. The essays in the data were collected from intermediate and advanced language learners who were mostly at year three or four of their undergraduate studies. Taking both a qualitative and quantitative perspective, Mazaherylaghab focused particularly on

the role of L1 and the notion of unnaturalness ('foreign-soundingness'). One thing worth pointing out with regard to this study is that the verb system of learners' L1 has been noted to show a rapid change towards multi-word verb formation, resulting in a decrease in the number of Persian simple verbs (Mazaherylaghab 2013: 167; *passim*). Based on this assumption, the Iranian learners were expected to display a tendency to produce more phrasal verbs than native speakers. The results obtained, however, did not accord closely with the predictions. Converging with most of the previous studies mentioned for other learner groups, Iranian learners also displayed a preference for one-word verbs, an underuse of figurative phrasal verbs as well as an overuse of some literal phrasal verbs. As part of his explanation for the results he obtained, Mazaherylaghab investigated the EFL textbooks used in the four years of high school in Iran so as to find out how the phrasal verbs were presented and taught. The results showed that phrasal verbs were left unattended for the most part, leaving the learners with a partial and ambiguous idea of phrasal verbs: They were formally introduced to the learners rather late – first in the third year of high school – and almost all of the phrasal verbs presented in the books had literal meanings (p. 72f). Therefore, the underuse of phrasal verbs by Iranian learners, especially the figurative ones, were accounted for by their avoidance in high school EFL materials used in Iran, in addition to other possibly influential factors – namely, the semantic and idiomatic complexity of these verbs. Given the fact that a learner's primary source of input in

the EFL context is the textbooks used in the classroom (see Schmitt and Redwood 2011: 186), this explanation seems quite plausible. Some of the errors committed by Iranian learners, such as additional use of prepositions as in *come across with* or *return back*, resulting in erroneous use in the former case and semantic redundancy in the latter, were accounted for by L1 influence (Mazaherylaghab 2013: 144). Other cases of semantic redundancy spotted in the data (e.g., *progress forward*) was explained by the researcher with the wish of the learners to be on the safe side with the meaning(s) they wished to convey (p. 159).

Related to the treatment of phrasal verbs in EFL textbooks are the notions of frequency and exposure, which were dealt with in Schmitt and Redwood's (2011) study. The researchers were interested in the extent phrasal verb frequency would affect the learning of phrasal verbs by EFL and ESL learners – both in a productive and receptive way. Being aware of the fact that “phrasal verbs are idiosyncratic in terms of their learning burden” and that “a purely frequency-based explanation can never fully explain their acquisition” (p. 187), Schmitt and Redwood (2011) also looked at some other possibly influential factors in their learning – namely, overall language proficiency, gender, age, the amount and type of exposure to TL, i.e., formal language instruction as well as incidental exposure to English outside of the classroom in the form of reading, watching films and TV in English, listening to English language music and social networking. Their results showed that the

frequency of phrasal verbs had a positive correlation with learners' productive knowledge of them to a considerable degree, but not in terms of their receptive mastery. Apart from phrasal verb frequency, exposure to the TL outside of the classroom in the form of reading and watching English language films and TV programmes was found to facilitate the learning of phrasal verbs. Contrarily, the amount of listening to English language music, social networking, formal-instruction-based variables (the type of instruction and hours of classroom input that the learners received), age and gender were found not to have an effect on the learners' knowledge of phrasal verbs. Overall, the findings revealed that the learners (whose proficiency level was intermediate to upper-intermediate) knew about two-thirds of the target 60 phrasal verbs receptively and about one-half productively (Schmitt & Redwood 2011: 191). Although these results still demonstrate an incomplete knowledge of the most common phrasal verbs, they are more promising than the results attained for advanced German learners by Cornell (1985) – the first study mentioned in this section. Cornell explained the lack of success displayed by advanced learners in using phrasal verbs with their limited contact with phrasal verbs. In Schmitt and Redwood's study, only two of the learners were German. The discrepancy between the two studies' results may be accounted for by the participants' L1 background but this is a very unlikely reason. One other, more plausible, reason could be the ever-increasing exposure of learners to English via various forms of electronic communication and media over the last few decades.

A quite recent study conducted by Garnier and Schmitt (2016), however, attained rather ‘disappointing’ results, confirming the fact that phrasal verbs are problematic for many learners. In their study, the polysemous nature of phrasal verbs was taken into consideration. To put it more precisely, unlike in earlier studies, which focused typically on the most frequent meaning sense of target phrasal verbs, Garnier and Schmitt tested multiple meaning senses of highly frequent phrasal verbs. Their participants – 128 students in BA English/TEFL courses, i.e., quite advanced English learners – were given a productive test in the form of a gap-filling task. Despite their high level of proficiency, the participants’ knowledge of phrasal verbs was found to be rather incomplete: The participants were found to know only about 40 percent of phrasal verb meaning senses on average, with only around a 20 percent chance of knowing all the different meaning senses of each phrasal verb tested (Garnier & Schmitt 2016: 37). The corpus frequency was found to be the best predictor of phrasal verb knowledge, a finding consistent with Schmitt and Redwood’s (2011) study. Time spent reading English and time spent in social networking in English per week were other influential factors in the phrasal verb knowledge. However, the factors of semantic opacity, previous L2 instruction, immersion in L2 environment, and year of BA study did not have any effect on phrasal verb knowledge, nor did listening to music or watching films in English (cf. Schmitt and Redwood 2011).

Coming back to the notion of frequency, its impact on the learnability of phrasal verbs was confirmed in another study. Chen (2013a, b), who conducted a longitudinal study on the quantitative use of phrasal verbs by Chinese learners of English in their written productions, found a positive correlation between phrasal verb frequency, in both novice and expert corpora, and their production by the Chinese learners. Therefore, Chen has argued against the claim that the learners whose L1 lacks phrasal verb structure struggle more in learning and using them. Chen has asserted that L1 is not necessarily the most influential factor in phrasal verb learning (2013a: 436) and her findings have lent support to this assertion: The participants in her study, despite the lack of phrasal verb structure in their L1, were found to produce a similar number of phrasal verbs in their writing. That is, the avoidance of phrasal verbs observed for most learner groups in previous studies was not found for the upper-intermediate Chinese learners of English in Chen's study; phrasal verbs were rather evenly distributed across individual learners, with every learner using roughly the same amount of phrasal verbs in their writing (2013a: 428). Chen has thus proposed an alternative explanation for the German learners' overuse of phrasal verbs as opposed to other learner groups in Waibel (2007): disparity between the English proficiency levels of different learner groups. In support of this claim, Chen (2013b: 433) refers to Thewissen's (2013) study in which essays in the German sub-corpus of ICLE were considered within the advanced range (C1 and C2) according to the Common European

Framework of Reference for Languages while essays in the French and the Spanish sub-corpora were primarily rated as upper-intermediate to advanced and lower to upper-intermediate, respectively (2013: 80). Acknowledging the fact that being able to produce a similar number of phrasal verbs does not necessarily mean that learners have achieved a native-like proficiency in their use (Chen 2013a: 433f), Chen investigated the semantic behavior of the phrasal verbs used in the data in a follow-study in 2017, viz. she calculated the number of meaning senses of each phrasal verb, arguing that as each sense of a given phrasal verb represents a different usage of it, the number may reflect how productive the learners were at different stages (during the three years investigated). The results have revealed that Chinese learners' semantic knowledge of phrasal verbs improved as their learning proceeded, i.e. after two years of undergraduate study. That is, the learners not only used phrasal verbs in a greater variety of senses as their learning proceeded, but also more often in figurative senses (Chen 2017: 142-147). One other way Chinese learners' knowledge of phrasal verbs improved was in relation to their stylistic aspect. While they made no use of academic phrasal verbs in their writing in the first year of their studies, they used some (e.g., *bring about*, *carry out*) in the following two years. Chen (2017: 149) stated, "the percentage of PVs [phrasal verbs] that are stylistically appropriate for writing increased from 37 per cent in Year 1 to 49 per cent in Year 3" (cf. Chen 2013b: 96f.) Chen's findings

clearly indicate the importance of increasing input and proficiency in phrasal verb acquisition.

Along the same lines as Chen, Mondor (2008) argued that proficiency level of learners might be more important than their L1. Mondor investigated the Swedish sub-corpus of ICLE, i.e. advanced Swedish learners, and compared the performance of this group of learners with that of their native speaker counterparts. Aims of the study were similar to the ones in previous studies; namely, identifying the similarities and differences between learners' and native speakers' use of phrasal verbs, focusing on the notions of over- and underuse as well as erroneous uses, with the ultimate aim of determining whether advanced learners were really as proficient in using phrasal verbs, especially idiomatic ones, as some previous studies claimed. Results showed that Swedish advanced learners produced not only literal phrasal verbs to the same extent as native speakers in terms of overall frequencies, but also figurative phrasal verbs (those in which there is an extension from a literal meaning, e.g. *bring in law and regulation* represents an extension from *bring in* in *bring in groceries*) and idiomatic phrasal verbs (where no literal interpretation is available, as in *figure out a solution to a problem*) (Mondor 2008: 212f). However, learners used fewer types of phrasal verbs compared to native speakers and displayed a large number of deviations, which are in most cases not completely erroneous but not entirely natural, either – what is referred to as “middle ground performance” errors by Lennon (1991b). The most common error type in the

learner data derived from the violation of collocational restraints, e.g. use of phrasal verbs with a wrong object as in **rub out thoughts*, followed by additional particle use with one-word verbs, e.g. *end up* instead of *end*. L1 transfer and intralingual analogy were found to account for most of the deviations. Although Swedish learners did not underuse phrasal verbs as a category, they underused transitive phrasal verbs in general and VOP [verb+object+particle] order with nominal objects, showing a general lack of knowledge of how different factors influence the word order choice in these constructions.

Similar results with regard to the discontinuous transitive phrasal verb use (i.e., VOP order) arose from a study conducted by Sung (2017), which is a large scale corpus-based study investigating the use of phrasal verbs by Korean learners of English, whose L1 lacks a similar structure. The learner corpus comprised of 300-word argumentative essays written by 3,286 Korean-speaking learners of English, whose proficiency levels covered the entire proficiency spectrum, albeit not evenly distributed. Focusing on the most frequent 150 phrasal verbs in COCA (Liu 2011), Sung compared the use of phrasal verbs by learners with that of native speakers, who were represented by the American part of LOCNESS. The focus of the study was on two factors: structural patterns of phrasal verbs and preference between phrasal verbs and their one-word equivalents in the learner corpus. The results showed that there was a striking discrepancy between learners and native speakers in terms of both token and type frequency of phrasal verbs. While every structural type of phrasal verb (i.e., VP, VPO, VOP) was significantly

underused by learners, greater underuse was observed with discontinuous transitive phrasal verbs (e.g., *bring it back*). The frequency analysis between phrasal verbs and their one-word ‘synonyms’ revealed that learners opted for one-word verbs more frequently, supporting the existence of a close link between the learners’ underuse of phrasal verbs and their preference for one-word synonyms, as suggested in earlier studies. Sung (2017: 21) attributed the general underuse of phrasal verbs in the learner data to three major causes: a) typological and systemic incongruencies between the learners’ L1 and the TL, b) syntactic complexity of the phrasal verb construction, and c) preference for one-word synonyms.

To my knowledge, only four studies have been conducted investigating Turkish learners’ use of phrasal verbs – the target learner group in the present study. The first study was Kayael’s unpublished Master thesis. In 2007, Kayael administered three elicitation tests to 400 ELT trainees – the same test types as in earlier studies; namely, a multiple-choice test, a fill-in-the-blank test, and a translation test, which consisted of the most frequent 20 phrasal verbs in BNC. Kayael ascertained their familiarity by examining various ELT textbooks and coursebooks. According to his results, the trainees (who are assumed to be advanced learners of English) did not display any avoidance behavior, in contrary to the expectations of the researcher. Test type was found to have an effect on the use of phrasal verbs; while all trainees used phrasal verbs less frequently in the fill-in-the-blank task, which required an active use of phrasal verbs, they all used phrasal verbs in multiple-choice tests more

often (at the level of recognition). The year of study, on the other hand, did not show positive correlation with the phrasal verb knowledge.

The second study conducted with Turkish EFL learners is Yıldız's (2016a). By means of a multiple-choice test comprised of 15 items⁵⁸, Yıldız aimed at determining whether Turkish and Norwegian EFL learners would avoid English phrasal verbs and if so, whether their proficiency level and the semantic complexity of phrasal verbs would be influential in their observed behavior. The proficiency levels of Turkish learners were B1 and B2, while that of Norwegian learners was B2. Yıldız interpreted the results he obtained as an instance of avoidance of phrasal verbs by Turkish learners, especially by lower level learners and when the phrasal verbs were figurative. Norwegian learners, however, did not avoid English phrasal verbs as a category, their performance being very similar to that of native speakers.

Continuing this line of research, Demiray Akbulut (2018) aimed at discovering Turkish learners' preferences between phrasal verbs and their one-word 'equivalent' by means of a multiple-choice test. The participants (112 participants in total), who were university students enrolled at the School of Foreign Language Department at a state university in Turkey, were also at two different proficiency levels, namely B1 and B2 level – as in Yıldız's (2016a) study. Based on Garnier and Schmitt's

⁵⁸ Although these verbs were provided in a list (p. 133), neither their one-word equivalents nor the sentences in the multiple-choice test were provided by Yıldız (2016a).

(2015) frequency list of phrasal verbs, 15 high-frequency and 15 low-frequency phrasal verbs were included in the multiple-choice tests because Demiray Akbulut was interested in the question of whether the learners' proficiency levels and frequency levels of phrasal verbs would affect the (avoidance) behavior of the learners. Although Demiray Akbulut (2018) claimed that Turkish EFL learners avoided phrasal verb use, this claim seems like a hasty conclusion to draw on the basis of the results provided in the article. While both B1 and B2 learners showed a (slight) tendency to use one-word verbs more often than the multi-word verbs – the former group displaying this preference more often than the latter one (p. 83f), they both used phrasal verbs, nearly as many times as their one-word equivalents (p. 82). With the increasing proficiency, the preference for phrasal verbs was found to be increasing as well. Regarding the phrasal verb frequency, the data revealed that in both groups the use of high-frequency verbs was more common than low-frequency verbs. One thing that needs to be noted about the results reported but was not explicitly addressed by the researcher is the fact that in the cases where the learners chose the one-word verbs, their choice was erroneous on quite a few occasions – 240 errors in total in the use of single-words. Whether this is due to learners' unfamiliarity of these verbs and their alternatives or due to the context provided which lacked enough information for the learners to make a successful choice should have been determined before judging the results as instances of avoidance.

To turn to one last study on the use of phrasal verbs by Turkish learners, Karakuş (2017), whose participants were 30 university students, employed two test types in her study. The most common 15 phrasal verbs in LOCNESS were taken as target phrasal verbs in both tests. The first test was a 5-point scale preference test, providing phrasal verbs with their one-word synonyms. The second test administered a week later was a translation test from Turkish into English. The results of the preference test revealed that some literal phrasal verbs (*slow down*, *stand up*, *speed up* and *come on*) were preferred over their one-word equivalents. This finding can be accounted for, as Karakuş (2017: 224) also states, by the low frequency of one-word verb equivalents, e.g. *decelerate*⁵⁹ and *accelerate*. There was, however, also one figurative phrasal verb which was preferred by more than half of the participants rather than its one-word equivalent, namely, *make up*. In the case of other phrasal verbs – *go out*, *bring about*, and *get away* – one-word equivalents were preferred by more than half of the participants. In the translation test, on the other hand, one-word verbs in general were used more often, which can be taken as an indicator for the difficulty of coming up with the phrasal verb on one's own. Karakuş concluded that although figurative phrasal verbs were preferred or used less than the literal ones

⁵⁹ Fletcher (2005) indeed notes that the verb *decelerate* is a very rare and specialized term, with fewer than 0.5 occurrences per million words in all registers of a large language corpus, unlike its equivalent phrasal verb *slow down*, which was found to register about 10 occurrences per million words in academic texts. He adds that while *slow down* can be used in a wide range of contexts, *decelerate* provides an alternative to *slow down* only in certain specialized contexts.

in both tests, one cannot talk of a general avoidance of phrasal verbs by Turkish learners. This study has shown that test type and semantic complexity of phrasal verbs as well as, although not mentioned by the researcher, the frequency of phrasal verbs with regard to their one-word equivalents play a role in determining the learners' performance.

In addition to other studies focusing solely on the use of phrasal verbs by learners with other L1 backgrounds (e.g., Kamarudin (2013) on Malaysian EFL learners, Uchida (2012) on Japanese learners, Wierszycka (2013) on Polish learners, and Fadanelli (2012) on Brazilian learners, just to name a few), there are studies dealing not explicitly with phrasal verbs but still directly relevant to the matter at hand; for example, Lennon's (1991a; 1996) studies, which have already been mentioned briefly in Section 2.2. Investigating the errors in advanced learners' oral productions, Lennon (1991a) found that almost half of the errors in his data were lexical or had a lexical element to them and 22 percent of the errors (162 out of 745 error types) were with prepositions and adverbial particle choice (pp. 41f). In a later article, Lennon (1996) examined the verb choice errors committed by advanced learners more closely, focusing on four high-frequency verbs, and concluded that even they struggled with the use of high-frequency verbs, lacking information as to their different meanings, grammatical use, contextual and collocational restrictions and phrasal verb combinations (p. 35). In some cases (for instance in the case of *put*) learners were found to have more trouble with the verb than the particle element in phrasal verbs. Not unrea-

sonably, Lennon (1996: 29) suggested putting more emphasis on differentiating the semantics of high-frequency verbs when teaching phrasal verbs rather than focusing on comparing the different particle combinations with a single verb – as is usually done.

One last thing to note is that in some of the aforementioned studies, phrasal-prepositional verbs were included under the phrasal verb type (e.g., Dagut & Laufer 1985; Hulstijn & Marchena 1989; Waibel 2007). To my knowledge, only one study has been conducted that focused exclusively on the use of phrasal-prepositional of nonnative speakers – namely, the study conducted by Ella and Dita (2017) on Philippine English – a second language variety. Their results have shown that Filipinos use a low number of phrasal-prepositional verbs in their speech and writing, more frequently in their speech, and are conservative in the choice of phrasal-prepositional verb structure – they prefer inseparable, monotransitive phrasal-prepositional verbs. Yet, they are proficient in their use. Taking the literature review above into consideration, one can conveniently propose that the results might have been different if the target had been a foreign language variety.

As it should be clear from the foregoing account of learners' phrasal verb knowledge, the results to date are somewhat inconclusive. One clear conclusion that can be drawn is that phrasal verbs remain a difficult aspect of English vocabulary despite the rising awareness with regard to their use – both by learners and native speakers. The difficulty with their use or preference of their one-word equivalents by learners

has been explained by a combination of various factors, including cross-linguistic differences between learners' L1 and English, the inherent semantic and structural complexity of phrasal verbs, quality and quantity of exposure, and the proficiency of learners. It may seem reasonable to most of us to suppose that all of these factors have a hand in guiding the learners' learning and use of phrasal verbs, but it is not very clear yet how they do so or how they interact.

There is rising awareness of the difficulties posed by the abundant phrasal verb use in native speaker language and an increasing acknowledgement of the marked significance phrasal verbs has to language learners. As a result, the number of pedagogically-centric studies conducted on phrasal verbs has been growing rapidly in recent years. A considerable number of these studies have yielded fairly positive results (e.g., White 2012; Farsani et al. 2012; Yasuda 2010; Kövecses & Szabó 1996; Karahan 2015; Lu & Sun 2017; Spring 2018; cf. Al-Otaibi 2019). Most studies have taken a cognitive approach to the instruction of phrasal verbs. Torres-Martínez (2015), on the other hand, has suggested taking a constructionist approach in teaching phrasal verbs. In approaches to their instruction, the focus differs: some focus on the verb as the starting point, some on the structural properties of phrasal verbs (e.g., Seidl 1990), some organizes phrasal verbs around a theme (e.g., Acklam 1992), whereas some others combine all these aspects (see Kurtyka 2001 for an early overview). Some make use of visualization in the presentation stage (e.g., Farsani et al. 2012) whereas some combine all these in one approach (e.g., Rudzka-Ostyn 2003).

As an area problematic for L2 learners, phrasal verbs are an aspect of language learning and teaching in need of further empirical research.

2.3.4.3 Previous findings from studies focusing on prepositional verbs

Prepositions are known as a chronic source of difficulty for foreign language learners and this difficulty has been widely expressed in the literature. McArthur (1989: 42), for instance, has noted that they are not as common as phrasal verbs but they are still common enough to trouble language learners. Drawing on findings from *Longman Grammar of Spoken and Written English* (LGSWE henceforth), Biber et al. (1999: 415) have shown that prepositional verbs are three to four times more common than phrasal verbs and in all four registers investigated – conversation, fiction writing, newspaper writing and academic prose. As they commonly occur in academic prose, they do not carry “informal overtones” as phrasal verbs (Biber et al. 1999: 415). Moreover, unlike in the case of phrasal verbs, for the formation of which a limited set of adverbial particles – all of which have locational or directional core meanings – are made use of, in forming prepositional verbs the full set of prepositions, including those denoting non-spatial relations, such as *as* and *of*, are employed (Biber et al. 1999: 415). Given their higher frequency compared to phrasal verbs, Leech (2001: 8) has suggested introducing prepositional verbs in ELT classrooms before phrasal verbs; according to him, prepositional verbs are “easier to handle” (ibid.).

Gilquin and Granger (2011) have underlined the fact that prepositions are troublesome not only for learners but also for teachers – “as the *bête noire* of both teachers and learners”, they are deemed impossible to teach and learn (p. 60). An abundant number of studies in the SLA field have demonstrated the difficulty they cause – both in their spatial and non-spatial uses – to learners coming from various L1 backgrounds – some of these studies will be mentioned in the brief literature review here.

Various explanations in different studies account for the difficulty experienced by learners in prepositional usage in English. It is indeed their very nature that does not make for easy learning – namely, their arbitrariness, polysemy, multifunctionality, semantic complexity, i.e., similar meaning senses making them hard to differentiate from one another, resulting in confusion on the learner’s side (e.g., *in* vs. *into* or *beneath/below/under/underneath*). Dealing with the semantic complexity of *between* and *through*, Kennedy (1991), for instance, has stated that it is not simple to explain how these two prepositions are used in spite of their high frequency, not even for adult native speakers of English – they will often refer to only the locative senses of these two prepositions and suggest that they are not “entirely discrete in meaning” (p. 95). Later in the same article, Kennedy (1991: 97) has added that “[i]t may even be that the higher the frequency of words and the more complex their semantic structure, the less reliable our intuitions about their most important functions in text”. Along the same lines, Kreidler (1966: 120) noted that native speakers of English have a tendency to think that each

preposition has its own separate meaning or groups of meanings. Kreidler then contradicted this tendency with an illustration of the anomalous usage of prepositions in English. Whereas different prepositions sometimes yield great differences in meaning (*look at* vs. *look for*), sometimes the difference is quite subtle (*concerned about* vs. *concerned with*). Furthermore, at times, there is no difference in meaning at all (*complain of* vs. *complain about*). Sometimes, the choice of preposition depends on the noun that follows it, i.e., whether the noun is animate or inanimate (*angry with* (or *at*) *somebody* vs. *angry about* (or *at*) *something*). Yet, in other cases, it is the part of speech which determines the preposition (*be fond of something* vs. *have a fondness for something*) (see also Rastall 1994). Moreover, contrary to native speaker intuition, prepositions are multifunctional. In addition to their spatial uses, they have non-spatial uses and they occur in various multi-word units. Moreover, they have a great variety of meanings. Fries (1945) noted polysemous nature of prepositions a long time ago in relation to nine prepositions – namely, *at*, *by*, *for*, *from*, *in*, *off*, *on*, *to*, *with*, which account for 92 percent of all occurrences of prepositions in Modern English. He stated that “the average number of separately numbered “senses” recorded and illustrated by the Oxford Dictionary for each of these nine words is thirty-six and a half” (Fries 1945: 30).

Cross-linguistic differences add to prepositions’ difficulty, which have been demonstrated by many studies (e.g., Jarvis & Odlin 2000 and the references therein; Arabski 1979; Özışık 2014). As exemplified by Ed-dine (2012: 105), *à* can indicate both direction and location/position in

French, whereas in English each sense can be expressed using one or more prepositions: *to* (direction) and *at/in/on* (location/position). Even between typologically similar languages, such as English and German, there is no total congruence with regard to their prepositional systems; for instance, while *in* can be used either for location or direction/goal in both German and English, English has another preposition similar in form used only for direction/goal – namely, *into*, which may result in the so-called underdifferentiation errors in learner language. Problems generally arise from multi-word units constructed with prepositions that are non-congruent between learners' L1 and TL, such as a German learner's producing of **dressed with a dark suit* or **You remind me at/on your father* (Swan 2009: 47). For the learner groups whose L1 lacks prepositions, such as the target group of the present study, prepositions may be more troublesome – as unfamiliar, 'alien' function words they may not be salient enough to stand out for their acquisition. As shown by VanPatten (1990), learners, especially those at the early stages, have great difficulty in attending to both form and content simultaneously in the L2 input – they attend to L2 input for meaning, focusing therefore mainly on content vocabulary. Functors with little meaning such as prepositions “may be left unattended, since they “escape” attention directed toward meaning or informational content” (VanPatten 1990: 294). Highlighting the anomalous and arbitrary nature of prepositions in English and the fact that one's L1 may be of no assistance during the process of their learning, Rastall (1994: 229) asserts that their learning involves “considerable costs in memorization

and storage of information”, very often resulting in “little or no compensating gain in communicational efficiency or usefulness.”

Regardless of their L1, even advanced language learners experience difficulty with prepositions, especially with their non-spatial uses (e.g., Lindstromberg 2010; Celce-Murcia & Larsen Freeman 1999). As was remarked by Kreidler (1966: 119f) long ago, “[a]nybody who has taught advanced foreign learners of English is aware that these abstract, chaotic functions of the prepositions remain as a stumbling block long after mastery of essentials has been achieved”. The difficulty experienced by learners at different levels of proficiency with regard to the use of prepositions has revealed itself in various ways in learner interlanguage. According to the prepositional errors he attested in the writings of Polish learners of English at three different levels (beginner, intermediate, advanced), Arabski (1979: 54f) concluded that the errors occurring in the use of prepositions fell into four major categories: (a) underdifferentiation errors caused by transfer of L1, (b) errors caused by transfer from acquired L2 constructions, i.e., errors of analogy (c) omission of prepositions caused by both transfer of L1 and other psycholinguistic reasons, and (d) mistakes (lapses) – which happen mostly at the early stage of learning due to lack of knowledge. Taking a more qualitative perspective, Arabski noted that while the errors of omission and lapses showed a tendency to decrease in the course of IL development, the errors in the higher level papers resulted mostly from transfer from L1 and L2 (errors of analogy) – the latter showing a tendency to accumulate with increasing proficiency (1979: 55f). The results of the studies

reported in recent years have given support to Arabski's observations to a large extent. For instance, Kao (2001), who focused on Japanese learners of English, has shown that communicatively redundant prepositions are likely to be omitted. The omission of a preposition where it is required has been observed in many learner varieties: For instance, Cowan (2008: 181) found instances of prepositional omission in the writings of Spanish and Korean learners of English; Hemchua and Schmitt (2006) in those of Thai learners of English. This phenomenon has also been reported for Turkish learners of English – the target group of the present study. According to the results of Kırkgöz (2010: 4354), who investigated 120 essays written by 86 adult Turkish learners of English at the early stage of their learning process, the highest number of errors was committed in the use of prepositions, omission being the most prevalent, occurring over three times more often than the other two prepositional error types – namely, addition of a preposition where it is not required and incorrect substitution of a preposition – what Kırkgöz referred to as “misuse”. Özışık (2014), who were particularly interested in the use of prepositions and the impact of L1 on their use, administered a gap-filling test to 30 Turkish EFL students at upper-intermediate level. All the test questions were based on interlingual errors in the phraseological uses of prepositions attested in student essays. That is, the test questions were designed to elicit uses of prepositions likely to be misused due to Turkish learners' L1, as illustrated in the following sentence (8) taken from Özışık (2014: 64):

(8) I thank _____ my family for their support.

The Turkish equivalent of the English verb *thank* requires a directional (dative) postposition – namely, –E, which corresponds to *to* in English – following its object. It was therefore expected that this inclusion would be carried over to English and the preposition *to* would be used in sentence (8) by the learners. The results obtained by Özişik verified this expectation: Turkish learners at this proficiency level still had great difficulty in using prepositions, committing a high number of errors, most of which could be explained by the influence of the learners' L1. While the most common preposition errors were found to be in the form of addition, omission errors were the least frequent ones. As for the part of speech of the words prepositions co-occurred with, it was found that nouns were the most difficult category (e.g. *the key to success*) compared to adjectives (*famous for*) and verbs (**marry with someone*) – the Turkish learners in the study were found to have the least problem with the latter category (i.e., prepositional verbs) (Özişik 2014: 68). The discrepancy between the results of Özişik (2014) and those of Kırkgöz (2010) – omission errors being the least frequent error type in the former, and the most frequent in the latter – could be accounted for by different proficiency levels of the participants. As mentioned earlier, errors of omission show a tendency to decrease with increasing proficiency (Arabski 1979: 55). One other possible reason could be task-related – gap-filling vs. free essay writing.

One other related study conducted with learners having the same L1 background was conducted by Yıldız (2016b). Investigating the spoken language productions of Turkish EFL learners at two proficiency levels

(intermediate and upper-intermediate), Yıldız has found prepositional errors to be the most common type compared to other lexical and grammatical errors attested in his data (cf. Elkılıç 2012; Erkaya 2012). Surprisingly, the omission of prepositions has not been attested in his data. However, the difficulty in use of prepositions revealed itself in two other ways: the use of an additional preposition (e.g., *marry *with a person, go *to abroad*) and a substitution of preposition – what Yıldız referred to as “misinformation” (*agree *about, shout *to students*). Notably, the number of substitution errors outnumbered the number of addition errors (Yıldız 2016b: 66f). Given the earlier mentioned fact that prepositions are communicatively not very ‘loaded’, omission errors are not unexpected or surprising (see Arabski 1979: 51). Cowan (2008: 181), however, simply explains the instances of preposition omission in prepositional verb constructions, as seen in the Spanish essays he investigated, with a lack of similar grammatical construction in the learners’ L1 (e.g. *listen to vs. escuchar*). As with all aspects of interlanguage, any possible explanation for omission of necessary prepositions must involve several factors. That is, it is more probable that interplay of less attention to form and impact of the learners’ L1, as well as some other factors, must be contributing to the non-use of prepositions.

Likewise, L1 has been commonly suggested as a possible source for the other two deviation types spotted in the use of prepositions in IL. Whereas Yıldız (2016b), for instance, explained the instances of both additional uses of prepositions and their substitutions, with L1 transfer (i.e., substituting a given preposition with L1 counterpart(s)), Cowan

(2008) mentioned L1 transfer as a source only for the former deviation – which he spotted quite often in Korean learners’ essays (e.g., *enter *to*). Given the fact that substitution errors occurred in the writings of students from a variety of L1 backgrounds, Cowan maintained that they could not result from L1 transfer and asserted that they occurred rather due to imperfect learning (Cowan 2008: 182). That is, lack of complete lexical knowledge and simple confusion of the verbs that are similar either in form or meaning (e.g. *result of* instead of *result from* or *look for* vs. *look after*) are other probable sources of this deviation type. Arabski (1979: 52) has claimed that substitution errors are the result of analogy based on acquired L2 constructions and they mostly occur in higher levels of proficiency. Having found this type of ‘error’ to be the most frequent one in the essays of Thai learners, Hemchua and Schmitt (2006: 19) stated that “the students were conscious of using prepositions, but they had not mastered which correct prepositions should accompany particular verbs, nouns or adjectives.”

All these aforementioned deviations in the use of prepositions and more (such as ‘unidiomatic’ usage) have been reported also for the indigenized L2 varieties (e.g., Nesselhauf 2009; Mukherjee 2009; Gilquin & Granger 2011; Schneider & Zipp 2013⁶⁰; Schneider & Gilquin 2018;

⁶⁰ Schneider and Zipp (2013) included two types of deviations under unidiomatic usage of prepositional verbs: (a) combinations that were used in contexts that do not match the interpretations given by dictionaries, and (b) combinations in which an existing particle verb is used where the simple verb would be the more appropriate choice.

Edwards & Laporte 2015; Mwangi 2004). Nesselhauf (2009), for instance, who investigated the phraseology of ESL and EFL varieties on the basis of the four subcorpora of ICE (namely, Kenyan, Indian, Singaporean, and Jamaican English) and of ICLE, has shown that the use of additional prepositions with one-word verbs is not constrained to learner varieties; they occur across varieties and in both ESL and EFL varieties. Referring to combinations including a redundant preposition (e.g., *comprise of*, *emphasize on*) as “new prepositional verbs” – as done by Mukherjee (2009) – rather than as ‘errors’, Nesselhauf has noted that such verb-preposition combinations indeed occur more frequently in several L2 varieties than in learner language (p. 19). The researcher has observed that there is a tendency to make the direction already expressed in verbs of movement more explicit in both ESL and EFL varieties (e.g., *enter into*, *approach to*) and remarked on these unrecorded combinations by stating that “it seems that the complex meaning of the verb ‘movement’ + ‘direction’ is broken up and each item assigned one part of the meaning” (Nesselhauf 2009: 20). Edwards and Laporte (2015), who refer to the redundant use of prepositions in combination with movement verbs as “hyper-explicitness”, have also attested this non-standard use in all three non-native varieties they investigated – namely Hong Kong English, Indian English and Dutch English – and stated that the case of *enter into* could be built by analogy with either *go into* or existing forms such as *enter into a partnership/discussions* (p. 158). Having observed some parallels in the formation of new prep-

ositional verbs, Nesselhauf (2009: 19) has called attention to the influence of the existence of a derivationally related noun which requires the preposition in question, e.g., the existence of *discussion about* leading to the creation of *discuss about* or *emphasis on* leading to *emphasize on* (see also Mukherjee 2009: 123f). Thus, acknowledging the impact of L1 or substrates in the use of prepositions, Nesselhauf (2009: 21) has concluded that the main reasons for the creation of new prepositional verbs in ESL and EFL varieties indeed “lie in constellations inherent in L1 English”, especially the irregularities existing in prepositional usage. Moreover, the “fine-grained” distinctions between closely related prepositions (e.g., *into* vs. *in*; *off* vs. *of*) have been mentioned, for instance, by Schneider and Zipp (2013) as a “pre-determined point of confusion”, resulting in unusual prepositional use, i.e., substitution (e.g. *result into*, *dispose off* attested in Fiji and Indian English). Failure to make such subtle semantic distinctions in some L2 varieties (e.g., Kenyan English), has already been considered as a simplification process (Mwangi 2004) – a common communicative strategy used by learners. In their study focusing on the phraseological aspects of the preposition *into* in learner varieties, in line with earlier studies, Gilquin and Granger (2011) have also mentioned confusion, L1 interference, analogy and blends as potential sources behind the non-standard uses of this preposition (p. 70ff). Considering the learner as creative in forming non-standard combinations based mostly on existing patterns, Gilquin and Granger (2011: 72) have noted that these non-standard combinations are entirely comprehensible, causing no difficulty in understanding the

intended message (e.g. *come into fruition/fiasco*), which in turn has made them reconsider whether these deviations from the standard English should be considered “real errors” or “new prepositional verbs”. As they rightly put it, “[t]he line is thin between errors and creative uses” and “non-native speakers are often denied the right to creativity” (Gilquin & Granger 2011: 72).

As the foregoing brief review might have shown, use of prepositions – in particular, that of prepositional verbs – is difficult to handle for foreign language learners. The claim that their use is particularly troublesome for learners whose L1 does not have a structurally equivalent form has been backed up by some studies. The difficulty experienced by learners in using prepositions reveals itself in different ways: omission of prepositions, substitution and use of a superfluous preposition in combination with one-word verbs, resulting in a lack of nativeness. These usage patterns are shared by some L2 varieties yet they are viewed more positively. Whereas the deviations from the norm in foreign learner varieties have been considered to be erroneous, mostly considered to result from L1 interference, the very same uses of prepositions by L2 speakers of English have been regarded as innovative or creative. Along with L1 interference, these uses are attributed to various other factors such as the irregularities of the English prepositional system, the parallels to existing semantically and formally related verbs and the existence of a derivationally related noun that requires the preposition in question.

2.3.4.4 Synopsis

What I hope to have shown in the foregoing review is that despite the differences across the studies, resulting in somewhat contradictory results, multi-word verbs remain a difficult aspect of English vocabulary for L2 learners, regardless of their L1 background. In spite of the environmental differences in the learning process of English, most learner groups display commonalities, i.e., shared difficulties, in their IL with regard to the use of multi-word verbs. The degree of the difficulty they cause varies depending on both the type and frequency of a given multi-word verb as well as learner-related variables, such as the typological distance between the learners' L1 and English determining the level of congruency, psychotypology, and language proficiency. The difficulty experienced by the learners in their use reveals itself in different ways, depending on the type of the multi-word verb. As a problematic area for L2 learners, multi-word verbs are an aspect of language learning and teaching in need of further research.

Chapter 3: Investigating multi-word verbs in a learner corpus

*Languages shape us
while we are busy thinking we are in charge of them.*
(Elif Shafak)

3.1 The notion of ‘multi-word verbs’ – definition and categories

Multi-word verbs consist of various combinations; either of lexical verbs followed by particles and/or prepositions or lexical verbs followed by other lexical words such as adjectives, nouns or verbs. These combinations are, however, considered multi-word verbs only where they operate as a single unit. For instance, according to Quirk et al.’s (1985: 1150) definition, a multi-word verb is a “unit which behaves to some extent either lexically or syntactically as a single verb”. The key phrase that needs to be highlighted in this definition is ‘to some extent’: The degree of fixedness in multi-word verbs, either syntactical or lexical, shows variety (see Section 2.3.2), resulting in a difficulty in differentiating them from free combinations. Therefore, Biber et al. (1999: 403) have suggested thinking of a cline – “a cline on which some verbs, or uses of verbs, are relatively free and others relatively fixed.”

Many different types of multi-word verbs have been dealt with in the literature. A number of linguists have attempted to categorize multi-word verbs, which, in turn, resulted in diverse classifications (e.g.,

Mitchell 1958; Fraser 1976; Vestergaard 1977; Denison 1981, 1984; Palmer 1974; Quirk et al. 1985; Claridge 2000; see also Cowie & Mackin, ODCIE, Vol. 1 (1975) and Cowie, Mackin & McCaig, ODCIE, Vol. 2 (1983)). The different attempts at their classification and the inconsistent use of terminology in referring to the verbs in the literature pose a general problem for their study. The aim of this chapter is not to provide the reader with an exhaustive description of the early treatment of the multi-word verbs and the classification of these verbs, but rather to provide the definition of *multi-word verb* as understood in this study (for a clear summary of the earlier treatment of multi-word verbs see Claridge 2000: 27-39). That will be followed by a clarification of how the four multi-word verb categories that will be investigated in this study are to be understood.

Before proposing the definition adopted in this study, we shall first discuss the characteristic features involved in the definition of multi-word verbs that distinguish them from free combinations. The example sentences given in the following (9-12) are taken from the Corpus of Contemporary American English (COCA).

(9) *A woman and child who were home when the men **broke in** weren't injured.* (COCA: 2015: NEWS)

(10) *Peter, opening the file, **glanced through** the pages with no real interest.* (COCA: 2012: FIC)

(11) *As much as you may be **looking forward to** retirement, turning from work on Friday to retirement on Monday can be an emotional shock.* (COCA: 2015: MAG)

(12a) *The numbers started rising sharply in 2009, when the first round of quantitative easing **took place**.* (COCA: 2015: NEWS)

(12b) *It was a mild, cloudless day, and she was glad of an excuse to **take a walk**.* (COCA: 2015: FIC)

(12c) *Those three words **put an end to** the old man’s talk, and his attention moved from conversation to the horses.* (COCA: 2015: FIC)

(12d) *In September, the firm presented a tweaked mine design that **took into consideration** local concerns about water.* (COCA: 2013: NEWS)

What the expressions in bold in the sentences above have in common is that (a) they consist of at least two lexical units, one of which is a verb, and (b) they represent a semantic unity. Despite the differences in their internal make-up, all the combinations exemplified by the sentences (9–12) function like a single lexical verb. That is, the parts of the combinations can be analyzed as constituents of a single unit.

One aspect in which these verbs differ is the extent to which the combination preserves the individual meanings of both the verb and non-verbal element, i.e., their (non-)compositionality or idiomaticity. In the case of some combinations, the meaning of the multi-word verb can be derived from the meaning of its constituents, e.g. *refer to*, *ask for*, *stay away from*. However, in the case of fused combinations such as *break in* in sentence (9), or *go into (a problem)* ‘investigate’, or *give in* ‘surrender’, or *walk away with* ‘steal and take away’, it is difficult, if not impossible, to derive the meaning of the multi-word verb from that of

its constituents. Although non-compositionality is a typical characteristic feature of multi-word units, as Wray (1999: 215) asserts, it is not a defining one (see Section 2.3.2). The borderline between literal and figurative combinations is in most cases rather fuzzy and permeable, partly owing to the polysemous meanings of some multi-word verbs, e.g. phrasal verbs, which fade into one another (cf. Claridge 2000: 47).⁶¹ Moreover, as Waibel (2007: 18) points out, the notion of transparency is restricted by a certain degree of subjectivity, especially where the learners are concerned: “a great deal hinges on language skill, the ability to detect and translate images and figurative language, and personal opinion” (Waibel 2007: 20). For instance, whereas the meaning of the verb *break in* in sentence (9) can be understood easily by native speakers, it may not be clear for a learner even though both of the elements in the combination are known to him/her. In that case this verb can be regarded as a true idiom for the learner. One other group of combinations, the meaning of which is not idiomatic but not as transparent either, the so-called “semi-idiomatic” phrasal verbs (Quirk et al. 1985: 1162) or “aspectual phrasal verbs” (Celce-Murcia & Larsen-Freeman 1999: 432), further complicate the transparency issue; in these verbs the verbal element in general keeps its original meaning while the particle

⁶¹ As pointed out by Waibel (2007: 16f), in the literature the term ‘literal’ is usually equated with ‘transparent’ and these two terms are used in opposition to ‘figurative’ and ‘idiomatic’. One other relevant term used similarly in opposition to ‘transparent’ is ‘opaque’. In this paper opaque multi-word verbs are believed to be idiomatic/non-compositional; i.e. their meanings cannot be derived from the combined meanings of its parts (cf. Quirk et al. 1985: 1163). For a full length treatment of the use of the terms ‘literal’, ‘figurative’, ‘transparent’, ‘opaque’, and ‘idiomatic’, see Waibel (2007: 116ff.)

specifies the verb by contributing aspectual meaning, such as signaling a beginning state (*take off*), continuity (*carry on, come along, work away*), repetition of an activity (*write over*) or completion of an action (*blow out, clean up*) (Celce-Murcia & Larsen-Freeman 1999: 432f).⁶² Thus, *up* in *use up* or *mix up* does not have its directional meaning – ‘to or in a higher position somewhere’; it rather adds the sense of completion (a detailed treatment of the aspectual meaning of *up* can be found in Bolinger 1971: 98-102). It turns an activity verb into an accomplishment, just as *down* in *burn down* (Celce-Murcia & Larsen-Freeman 1999: 433). Although aspectual particles signal certain meanings consistently, they cannot freely co-occur with any verb (*ibid.*). Moreover, their meanings may be interpreted differently. For instance, while *out* in *point out* conveys the meaning of completion for Quirk et al. (1985: 1163), it implies (figurative) direction from the inside (one individual) to the outside (another individual) for Waibel (2007: 19). Turning back to the transparency issue, the existence of polysemous multi-word verbs, such as *hold up* (one literal and one idiomatic meaning) and *take in* (several idiomatic meanings in addition to literal), complicates the problem further; they make it hard to determine objectively at which point the combination’s meaning is still transparent and at which point it becomes idiomatic.

⁶² These phrasal verbs are called differently in the literature: “completive” by Dagut and Laufer (1985), “semi-idiomatic” by Quirk et al. (1985: 1162), “aspectual phrasal verbs” by Celce-Murcia and Larsen-Freeman (1999: 432f) and “semi-transparent” by Waibel (2007) and Laufer and Eliasson (1993).

The current study adopts the definition as well as classification of multi-word verbs proposed by Claridge (2000).⁶³ Following Claridge (2000: 28f), a multi-word verb is here defined as a group consisting of two or more words including at least one real verb provided only that the meaning of combinations constitute a syntactic and/or lexical unit functioning like a single lexical verb regardless of whether they form an uninterrupted sequence or are spread discontinuously across the clause. All parts in multi-word combinations are treated the same, i.e. as parts of verbs. Thus, the indivisible unit of thought is taken as the most essential criterion in the definition of multi-word verbs in this study. The aspect of idiomaticity or transparency is not considered as a defining criterion for the multi-word verbs in this study. The decision was taken mainly on the grounds that, as mentioned earlier, a clear-cut differentiation between what is literal/transparent and what is idiomatic is hard to make in many cases, especially if one is dealing with learner language. A second reason behind this decision is practical: The number of multi-word verbs in the investigated corpora is too high to conduct a thorough semantic analysis. The term ‘multi-word verb’ is adopted since it conveys the multi-word character of these verbs and it is fairly transparent, making it more generally accessible. The frequencies and

⁶³ Claridge (2000) classifies multi-word verbs into six groups: phrasal, prepositional and phrasal-prepositional verbs and verb-noun, verb-adjective and verb-verb collocations. In her analysis, she omits verb-verb combinations from consideration because she is of the opinion that “they are certainly a minor category – probably slightly more uncommon than verb-adjective combinations, and as a group more fragmented than the latter.” (p. 40). In the present study, both verb-adjective and verb-verb collocations were excluded, mainly for practical reasons.

as a result the impact of the combinations exemplified above (example sentences 9-12) on the structure of English differ notably, with the type found in sentence (9) certainly being the most prominent one – namely, phrasal verbs (Claridge 2000: 26). The reason for deciding to treat all four verb categories together in this study is mainly their importance in the English language and in language learning as well as the difficulty they cause for learners. By restricting the analysis to one L1 group and including different categories of multi-word verbs, the present study aims to capture any difficulties that this group have in using multi-word verbs.

The proposed classification by Claridge (2000) with only brief information about the categories of multi-word verbs is given below. The precise definition as well as more detailed information on the individual categories will be provided in the following section.

(I) Phrasal verbs

Verbs followed by a particle of an adverbial nature in a non-prepositional use.⁶⁴ Phrasal verbs can be transitive (*look up, bring up*) or intransitive (*grow up, give in*), literal (*come in, sit down*) or figurative (*put off, rule out*).

⁶⁴ As has already been noted in the literature (e.g., Sroka 1972; Kaluza 1990; O’Dowd 1998), the distinction between prepositions, adverbs and intermediary categories is problematic. Claridge (2000: 39, fn. 4) states that her decision as regard to this distinction is based on a pragmatic surface-structure approach: “prepositions, however closely connected to the verb, need a nominal element to refer to, and whenever this nominal

(II) Prepositional verbs

Verbs followed by a preposition in its clear prepositional use (*look at, come across, refer to*). The noun phrase that follows the preposition is considered as its direct object. Thus, all prepositional verbs are considered transitive. The subtype of prepositional verbs with the sequence verb-free object-preposition (e.g. *turn N into N*) is also treated. Verbs with two prepositions are also included under this type of verbs by virtue of their first preposition, while the second one is ignored (*develop from N into N, struggle with N for N*).

(III) Phrasal-prepositional verbs

Verbs followed by an adverbial particle and a preposition. There are two types: monotransitive (*get away with, stand up for, check up on*) and ditransitive (*put N down to N*).

(IV) Verb-noun collocations

This group of verbs, despite the variance in their internal structure, contains a nominal element as a fixed part. Following Claridge (2000: 40), this group of verbs is further subdivided into three types: (a) simple verb-noun combinations (*commit a crime, attract attention, do the washing*), (b) verb-noun-preposition combinations (*catch sight of,*

element is missing, the adverbial interpretation is enforced.” The same approach is applied here.

catch a glimpse of, give rise to, put the blame on), and (c) verb-prepositional phrase unit (*come to a conclusion, bring to light, come (in)to (one's) mind, put on the spot*).

3.2 Multi-word verb categories investigated in this study

This section provides detailed information on the categories of multi-word verbs included in this study for the sake of clarification.

3.2.1 Phrasal verbs

This is the place to address two problems concerning the category of phrasal verbs – namely, its definition and the term itself. As the interest of the present study lies in actual learner performance, it is not intended here to enter into a detailed theoretical discussion of phrasal verbs. A few remarks regarding the problematic issues, however, are necessary for the following data analysis.

Not only the definition of a ‘phrasal verb’ differs in the extensive body of literature dealing with this group of verbs owing to the semantic and syntactic complexity of them but also the term itself is problematic.⁶⁵

⁶⁵ The term ‘phrasal verb’ was introduced by Logan Pearsall Smith (1925). Considering them one of the most striking idiosyncrasies of English, he defines ‘phrasal verbs’ as verbs which are “followed by prepositions, or prepositions used as adverbs.” As a footnote, Smith then adds that “[t]here are some phrasal verbs with two prepositions” (Smith 1925: 172). With regard to the term itself, Smith states the following as a footnote: “The term “phrasal verbs” was suggested to me by the late Dr. Bradley; not, as he wrote, that he was satisfied with it, or would not welcome any alternative that he could feel to be an improvement. But, as he said, one cannot write of these verbs without

The term is problematic firstly because this category of verbs has been referred to with different terms in the literature (e.g. “verb-adverb combination” (Kennedy 1920), “poly-word verbs” (Stevick 1950), “separable verb” (Francis 1958), “two-word verb” (e.g., Anthony 1954; Taha 1960; Meyer 1975), “verb-adverb locution” (Roberts 1936), “discontinuous verb” (Live 1965), “compound verbs” (Bywater 1969), “verb-particle constructions” (Lipka 1972), “verb-particle combination” (Fraser 1976; Cowie & Mackin 1975), “particle verb” (Johnson 1991), among others), stressing different aspects of these combinations, and secondly because the scope of the term has been different in the literature. That is, the term has not always been used to refer to so-called ‘proper’ phrasal verbs (i.e., idiomatic verb-particle combinations, Quirk et al. (1985)), but as a general term including prepositional verbs such as *cope with* (e.g., Sroka 1972), phrasal-prepositional verbs such as *come up with* (e.g., Bywater 1969; Waibel 2007; Mazaherylaghab 2013), or both prepositional verbs and phrasal-prepositional verbs (e.g., Dixon 1982; Cornell 1985; Sjöholm 1995), depending on the definition adopted. Similarly, most phrasal verb dictionaries, for instance, Sinclair (1989), Cowie and Mackin (1993), Cullen and Sargeant (1996), and the *Cambridge International Dictionary of Phrasal Verbs* (1997), also include prepositional verbs and phrasal-prepositional verbs under the

some workable description; and although the word “phrasal” is perhaps objectionable in formation, it fills a want, and is sometimes indispensable” (Smith 1925: 172).

term ‘phrasal verb’.⁶⁶ There are some others who include other combinations under this term; for instance, Courtney (1983) included verb-adjective (*lie low*), verb-pronoun (*kid oneself*) or other combinations such as the idioms that are formed from phrasal verbs (*be off the hook*, *take one’s breath away*). Most scholars, on the other hand, (“non-dictionary oriented linguists”, as referred to by Waibel (2007)), consider only verb-particle combinations as phrasal verbs (e.g., McArthur 1989; Lipka 1972; Biber et al. 1999; Greenbaum 1996; Palmer 1974) but some do not call these combinations ‘phrasal verbs’; Lipka (1972), for instance, uses the term “verb-particle construction.”

Regarding the problem with the definition of phrasal verbs, there are two main issues of concern: As Waibel (2007: 15) explains, these are the nature and grammatical status of the adverbial element and the phenomenon of idiomaticity, i.e. non-compositionality, of phrasal verbs.⁶⁷ The fact that the particle component of a phrasal verb has an adverbial status has been accepted by many linguists (Mitchell 1958; Bolinger 1971; Lipka 1972; Palmer 1974; Quirk et al 1985; McArthur 1989; Greenbaum 1996; Biber et al. 1999; among others). Particles of phrasal

⁶⁶ Defining ‘phrasal verb’ as “a form of idiom”, Cullen and Sargeant (1996: vii-viii) find it necessary to justify the inclusion of phrasal verbs that have a literal meaning with a related idiomatic meaning or a semi-idiomatic meaning in their dictionary even though they are not ‘strictly’ phrasal verbs. However, if one considers the fact that these dictionaries are to aid learners with verb-particle combinations in general, the inclusion of these phrasal verbs in addition to the so-called ‘proper’ phrasal verbs is already justified.

⁶⁷ For a brief survey of criteria used in earlier studies related to phrasal verbs, see Lipka (1972: 20-27).

verbs sometimes have their origins in prepositional phrase reduction and in many cases the missing prepositional phrase can be supplied by the listener/hearer thanks to contextual clues or sociocultural knowledge (Dixon 1982: 9), as in *she took the book out (of the bag)* or *she ran up (the stairs)* (cf. Bolinger 1971: 23f.). As the particle in these ‘reduced prepositional phrases’ can stand alone, it does not function as a preposition anymore but as an adverb (Lipka 1972: 175; McArthur 1989: 39; see also Quirk et al. 1985: 1155).⁶⁸ That is why, following Claridge (2000), these cases are included under phrasal verbs in the present study if the reduction is not a case of “context-determined ellipsis”, i.e., the ‘missing’ element should not be mentioned in the same or the immediately surrounding context (Claridge 2000: 48f.). However, since the particle can stand on its own as an adverbial in ‘reduced prepositional phrases’, that is, there is actually nothing missing, in the present study the particle in such phrases will be referred to as the ‘expandable particle’.⁶⁹

As for the aspects of transparency and idiomaticity, it should be noted that, as mentioned earlier, these notions are dealt with differently in the literature in the context of phrasal verbs. Whereas most phrasal verb dictionaries include both literal and idiomatic phrasal verbs (e.g.,

⁶⁸ Cf. Huddleston and Pullum (2002: 612f.).

⁶⁹ As noted in the preceding paragraph, Lipka (1972) has opted for “verb-particle construction” as a term rather than ‘phrasal verb’. He explains this preference with the inclusion of ‘reduced prepositional phrases’ in his study (p.17).

McArthur & Atkins 1974; Courtney 1983; Sinclair 1989; Cullen & Sargeant 1996), Cowie and Mackin (1993), for example, do not include non-idiomatic phrasal verbs in their dictionary. Similarly, only the combinations of verb and adverbial particle with idiomatic nature are considered to be ‘proper’ combinations by some linguists, i.e., where the meaning of a given phrasal verb cannot be understood solely through the meanings of its components. For instance, the literal verb-adverb combinations in which each component keeps its own distinctive meaning are considered ‘free combinations’ by Quirk et al. (1985: 1152), as well as by Biber et al. (1999: 403), arguing that “each element has separate grammatical and semantic status”. However, as noted earlier, it is anything but an easy task to make a clear distinction between free combinations and fixed multi-word verbs in practice.

In the present study, the literal-idiomatic division in the definition of phrasal verbs will not be applied, in contrast to, for instance, Quirk et al. (1985) or Dixon (1982); both completely literal verb-particle combinations (verb in its literal sense or both verb and its particle in their literal or adverbial sense, e.g., *come in*) and idiomatic verb-particle combinations (*put off* = *postpone*) will be considered as phrasal verbs provided that these two lexical items function like a single verb (similar to Claridge 2000; see also Bolinger 1971: 16, who refers to Fairclough 1965 as doing the same).⁷⁰ The verbal component of these structures

⁷⁰ Cowie and Mackin (1975: ix) state that semantic unity can be tested very easily by substituting the verb-particle combination with a single verb of equivalent meaning.

can be both an intransitive and a transitive verb. The terms ‘literal’ and ‘transparent’ are used interchangeably for the multi-word verbs whose meaning is implied by the meanings of the elements composing the verbs.

As noted in Section 3.1, the reasons for not applying the literal-idiomatic division in this study are mostly pragmatic: Firstly, the number of phrasal verbs in the investigated corpus is too large to conduct a thorough semantic analysis. Secondly, most phrasal verbs have polysemous meanings, which often shade off gradually into one another. Most importantly, it is often difficult to draw a clear-cut boundary between literal and idiomatic phrasal verbs (cf. Bolinger 1971: 36, fn. 12). Cowie and Mackin (1975: x) reasonably suggest that it is better to think “in terms of a scale of idiomaticity”. As Waibel (2007: 19) points out, “[a]lthough the end points of such a scale are fairly clear-cut with literal and opaque/idiomatic phrasal verbs, the intermediate stages consist of too many shades of grey which are impossible to define clearly.” Claridge (2000: 47) makes also a valuable point that is worth mentioning in this context:

Idiomaticity, after all, does not emerge out of nowhere, but is based on in some way or other on the regular patterns of the language. Literal phrasal verbs are the core from which figurative types are ultimately derived, and to which they are still connected by an identical, or in idiosyncratically frozen idioms at least similar, syntactical behavior. Thus, idiomatic phrasal verbs cannot be

understood without their literal background, and in the less idiomatic cases the connection, the underlying thought process of semantic idiomaticization, is still graphically clear.

To sum up, in this study, following Claridge (2000), those verbs followed by an adverbial particle in a non-prepositional use are considered to be phrasal verbs if these two lexical items function to some extent either lexically or syntactically as a single semantic unit, regardless of the degree of their transparency, i.e. including all combinations in the ‘cline’ of transparency from the completely literal to the totally opaque ones.

One other concern regarding these structures is the differentiation between transitive phrasal verbs (e.g. *look over a writing assignment*) and prepositional verbs (e.g. *look over one’s shoulder*), the forms of which are indistinguishable from one another at first glance. Various syntactic tests have heretofore been proposed to differentiate these two types of verbs by researchers (e.g., Bolinger 1971; Darwin & Gray 1999; Quirk et al. 1985; Claridge 2000). In this study Quirk et al.’s (1985: 1167) syntactic tests will be applied in order to distinguish them:

- (a) The particle of a phrasal verb can stand either before or after the noun phrase following the verb, but that of the prepositional verb must (unless deferred) precede the noun phrase.

- (b) When the noun phrase following the verb is a personal pronoun, the pronoun precedes the particle in the case of a phrasal verb, but follows the particle in the case of a prepositional verb.

- (c) An adverb (functioning as adjunct) can often be inserted between verb and particle in prepositional verbs, but not in phrasal verbs.
- (d) The particle of the phrasal verb cannot precede a relative pronoun at the beginning of a relative clause.
- (e) Similarly, the particle of a phrasal verb cannot precede the interrogative word at the beginning of a *wh*-question.
- (f) The particle of a phrasal verb is normally stressed, and in final position normally bears the nuclear tone, whereas the particle of a prepositional verb is normally unstressed and has the ‘tail’ of the nuclear tone which falls on the lexical verb.

The above criteria to distinguish transitive phrasal verbs from prepositional verbs are displayed in Table 1 (taken from Quirk et al. 1985: 1167). I can agree with Fraser (1976: 2) that the main criterion distinguishing these two types of verbs from one another is the position of the direct object – criteria (a) and (b) in Table 1. In the case of transitive phrasal verbs, the direct object may either follow the verb-particle combination or appear between the verb and the particle (*He turned on the TV* or *He turned the TV on*). This ‘particle movement rule’ is optional as long as the direct object is not a pronoun. If the direct object is a pronoun, the particle movement rule must be applied (*He turned it on* vs. **He turned on it*). One thought, however, needs to be added here on the particle movement rule: Although it is optional unless the object is a pronoun, this rule is not applied by native speakers if the transitive phrasal verb is followed by a long object noun phrase, as illustrated in

the following sentence (13) taken from Cowan (2008: 172), because the occurrence of long, complex object NPs between verb and particle would violate ‘the principle of end weight’ – a general tendency seen in English of having long, “heavy” elements at the end of a clause or sentence, rather than in the middle (ibid.; see Arnold et al. (2000) for the functions of end weight)

(13) John **looked up** *some information about an early religion in which forces of nature such as fire were worshipped.*

Moving the particle to the position following the long NP in the sentence (13), as shown in sentence (14), would not make the sentence ungrammatical, yet harder to process and understand for the listener (Cowan 2008: 176).

(14) John **looked** *some information about an early religion in which forces of nature such as fire were worshipped* **up.**

In contrast to separable transitive phrasal verbs, prepositional verbs, which have a preposition as its second element, cannot be separated by the object; separating the verb and preposition will produce an ungrammatical sentence, as exemplified with *call on* in Table 1.

Chapter 3: Investigating multi-word verbs in a learner corpus

Table 1. Diagnostic frames for phrasal and prepositional verbs.

Prepositional verb	Transitive phrasal verb
<i>call on</i> = ‘visit’	<i>call up</i> = ‘summon’
(a) They <i>called on</i> the dean. * They <i>called</i> the dean <i>on</i> .	They <i>called up</i> the dean. They <i>called</i> the dean <i>up</i> .
(b) They <i>called on</i> him. * They <i>called</i> him <i>on</i> .	They <i>called</i> him <i>up</i> . * They <i>called up</i> him.
c) They <i>called</i> angrily <i>on</i> the dean.	*They <i>called</i> angrily <i>up</i> the dean.
d) The man <i>on</i> whom they <i>called</i>	*The man <i>up</i> whom they <i>called</i> .
e) <i>On</i> which man did they <i>call</i> ?	* <i>Up</i> which man did they <i>call</i> ?
f) Which man did they <i>CALL on</i> ?	Which man did they <i>call up</i> ?

Table 2 below sums up the phrasal verbs investigated in the present study (adapted from Waibel 2007: 64). One other thing to note is that in this study only verbal uses of phrasal verbs will be taken into consideration. Nominal uses like *washing-up* in the following sentence (15) will not be considered:

(15) [...] *you are fed up with washing-up* [...] <TRCU1059>

3.2 Multi-word verb categories investigated in this study

Table 2. Types of phrasal verbs investigated in the present study

Phrasal verb consisting of	Meaning	Transitive phrasal verb
intransitive verb + particle	literal	<i>Come in</i> , Peter.
intransitive verb + particle	figurative	Susan finally <i>settled down</i> .
transitive verb + particle	literal	He <i>took out</i> the rubbish.
transitive verb + particle	figurative	They <i>took in</i> a homeless person.
transitive verb + expandable particle	literal	She <i>took</i> the book <i>out</i> .
intransitive verb + expandable particle	literal	He opened the door and <i>looked inside</i> . He <i>ran up</i> .

Phrasal verbs, which have been estimated by some to form about one third of the English verb vocabulary (Li et al. 2003: 513), are a highly productive category of multi-word verbs in English (e.g., Darwin & Gray 1999: 66; Meyer 1975: 3; Bolinger 1971: xi-xiii). Based on the 100-million-word BNC, Gardner and Davies (2007) have found that phrasal verbs occur on average in every 192 words in English, that is, nearly 2 phrasal verbs per page of written text (p. 347). The meanings of phrasal verbs are changing or new meanings are added, and new combinations come into existence (Sinclair 1989: iv), especially in slang (e.g. *boogie on down*) and in scientific or technical fields (e.g. *spin down* a sample) – often by means of analogy with existing phrasal

verbs (Side 1990: 146).⁷¹ In addition to ‘verbs of movement’ like *go*, *come* and *take*, McArthur (1989: 40ff) has listed three other sources in forming new phrasal verbs: adjectives (as in *freshen up*, *flatten down*), nouns (as in *board up*, *dish out*) and Latinate verbs (as in *contract out*, *measure off/up*).⁷² It looks like it is the unconscious knowledge of the meaning of particles that native speakers of English have that allows the creation of new phrasal verbs almost at will (Darwin & Gray 1999: 66). Sinclair (1989: iv) have commented insightfully on the new combinations as follows:

these new combinations are rarely made on a random basis, but form patterns which can to some extent be anticipated. Particles often have particular meanings which they contribute to a variety of combinations, and which are productive; that is, these fixed meanings are used in order to create new combinations.

According to Side (1990: 146), the particle is integral to the meaning of the phrasal verb. Acknowledging the fact that in some cases the particle adds little to the communicative value of verb (as in the case of *hang your coat here* vs. *hang your coat up here*), Side argues that in most cases it carries most of the meaning. He illustrates this argument by means of an example as follows (Side 1990: 146):

⁷¹ For a brief summary of the origin and development of phrasal verb patterns in English and the development of new meanings for the adverbial particles, see Meyer (1975: 5-8).

⁷² McArthur (1989: 41) has added that use of particles in combination with Latinate verbs, as completives or for emphasis, is regarded “pleonastic” by some, who in turn avoid or disparage them (see Bolinger 1971: xii).

if I tell somebody to *bog off*, that person is well aware that what I am saying has nothing to do with bogs and everything to do with beating a retreat. In other words, the main communicative function of this particular phrasal verb is carried by the particle. Whatever verb I put before it merely conveys the depth of my feelings (or depravity).

Others have expressed similar claims. Declerck (1977: 306), for instance, suggested that the particle in the phrasal verb construction is more central than the verb and stated that it “receives the semantic focus of the sentence”. According to Bacchielli (1993), the basic idea of the action is expressed by the particle: the particle “acquires a verbal force whilst the preceding verbs [...] are confined to a modal, or instrumental function and thus become premodifications of [the particle]” (p. 58).⁷³ Along the same lines, Bolinger (1971: 49) stated, “[a]t times it seems as if the verb proper and the particle had switched roles – the adverb becomes the verb and the verb the adverb” and illustrated his point with the following two sentences:

(16) *They scratched the mistakes off. They offed (erased) the mistakes by scratching.*

(17) *Johnny ran away. Johnny awayed himself (absented himself) by running.*

Accordingly, it was suggested that phrasal verbs could be better classified according to the meaning of the particle rather than according to

⁷³ Bacchielli (1993) illustrates his argument with the particle through in the following combinations: look through, to glance through, to run through, to flick through, to skim through, to leaf through, to riffle through and to thumb through (p. 58).

the verbal element (e.g., Bolinger 1971; Lipka 1972; Side 1990). For instance, in the case of the particle *off*, its meanings have been classified as (a) indicating distance in time or space (*put off a meeting, warn somebody off*), (b) departure (*take off*) or separation (*see someone off*), (c) removal (*strain off liquid, take off coat*), (d) disconnection (*cut off*), etc. Being systematic and logical, it is believed that such particle-oriented classifications would permit easier learning of phrasal verbs for the learners. Besides, the number of particles in these combinations is limited.

Although phrasal verbs were once thought to be colloquial, i.e., common only in speech and informal writing (e.g., McArthur 1989: 40), it is now generally accepted that they can be found in all registers. As already noted by Cornell (1985: 269), the situation with phrasal verbs is by no means static: “a phrasal verb hitherto felt to be colloquial or informal may, sometimes quite abruptly, become accepted in more formal contexts.” Acknowledging the fact that they are more common in less formal language, Darwin and Gray (1996) indicate that it should come as no surprise to find them in all registers of language since they are “virtually unavoidable without lengthy and often pretentious circumlocutions” (p. 66).

As regards the verbs that are commonly found in verb-particle combinations, it has been noted in the literature that there are clear preferences to be found. That is, not any kind of lexical verb could be a part of phrasal verbs. Live (1965: 430), for instance, points out that the verbs mostly found in these combinations are generally of the old, common

monosyllabic or trochaic “basic English” variety – many of them having irregular conjugation in modern English – such as *bring*, *send*, *take*, *set*, *go*, *come*, *look*, each occurring in combination with a great number of particles. Live adds, “[o]f the more learned (often poly-syllabic or iambic) verbs of classical or French borrowing, few co-occur with more than one or two of these particles, many with none” (ibid.). Kennedy (1920: 29) similarly claims that the verbs that enter into these combinations are largely monosyllabic and of Germanic origin (see also Bolinger 1971: xii and 175), and provides tentative numbers: “Out of 826 combinations specifically examined only ninety-seven have dissyllabic verbs” and only three of these disyllabic verbs have the accent on the last syllable (Kennedy 1920: 29). Supporting these observations on the phonological properties of verbs combining with particles, Fraser (1976: 13) adds that many of the monosyllabic verbs contain a final syllabic liquid or nasal [l], [r], [m], [n]. Being a monosyllabic verb, however, does not guarantee co-occurrence with particles; while this is a crucial prerequisite, it is not a guarantee because there are many monosyllabic verbs that never occur with a particle, e.g. *nod*, *rock*, *dive*, *fast*, etc. (Fraser 1976: 15). Moreover, there are few exceptions to the condition that a verb be either monosyllabic or bisyllabic and initially stressed, e.g. *divide up*, *partition off*, *balloon out*, etc. (Fraser 1976: 14). Therefore, despite the major role the phonological shape of the verb plays in determining the possibility of a combination with a particle, as pointed out by Fraser, it alone cannot determine the conditions for verb-particle combinations; it can only help specify a large amount of verbs

which do not combine with particles – almost all the polysyllabic verbs of the English language (Fraser 1976: 15).

One other feature shared by the verbs found commonly in phrasal verb constructions is the fact that they are the very common and multi-functional verbs (see Kirchner 1952; Bolinger 1971: xii).

Another observation in this context is that there is also a group of verbs entering into these combinations that seldom or never occur independently, i.e. without a particle, e.g. *even up* and *jot down* (Kennedy 1920: 29, see also Live 1965: 432) and those that are used intransitively only in these combinations, e.g. *calm down* and *brush up* (Kennedy 1920: 29).

Regarding the semantic features of the verbs in verb-particle constructions, it seems that stative verbs such as *know*, *want*, *see*, *hope*, *resemble*, etc. hardly ever combine with particles (Fraser 1976: 11) but other than this negative statement there is no basis – syntactic or semantic – for specifying which verbs can co-occur with a particle or which cannot (Fraser 1976: 11-13).

As it should be clear from the foregoing account, although there is some common thread for the verbs that are commonly found in verb-particle combinations, it is still not possible to provide a full list of all potential verbs that can be combined with an adverbial particle. The number of

possible adverbial particles that occur in phrasal verbs, however, is confined.⁷⁴ Thus, the 28 adverbial particles given in Waibel (2007: 65), which combine with verbs to form phrasal verbs, will be considered in the data analysis – 20 of these adverbial particles are taken from Johansson and Hofland (1989) and five of them from Quirk et al. (1985). Here is a complete list of all the adverbial particles to be examined in the study; those marked with an asterisk are the ones mentioned in Quirk et al. (1985: 1151):

*aback**, *about*, *across*, *ahead**, *along*, *apart**, *around*, *aside*,
away, *back*, *behind*, *by*, *down*, *forth*, *forward**, *in*, *off*, *on*, *out*,
over, *past*, *round*, *through*, *together**⁷⁵, *under*, *up*, *with*, *without*

Before moving on, a few notes are in order about the exclusion of some particles in this study. Apart from the particles mentioned by Johansson and Hofland (1989) and those marked with an asterisk, Quirk et al. (1985: 1151) list more particles: *above*, *astray*, *home*, *in front*, *on top*,

⁷⁴ According to Mitchell (1958: 105), the productive verbals in forming phrasal verbs are: *bring*, *come*, *get*, *go*, *keep*, *run*, *put*, *take*, *turn*, *send*, *fall*, *stand*, *look*, *set*. He, however, also notes that “although it should be possible to establish a closed system of particles, this would hardly be practical for the verbal component” (ibid.).

⁷⁵ Following Waibel (2007), *together* will be counted as adverbial particle only when it is not in opposition to *alone*. That means *live together* is not considered as phrasal verb because it contrasts with *live alone*. Similarly, *live apart* is also not included. Constructions like *bring together* or *mix together*, on the other hand, will be included because in these cases, the particle cannot be substituted by *alone*; it has intensifying meaning and furthermore passes the pronoun test given by Quirk et al. (1985: 1167):

(a) The minister is trying to **bring together** the two sides./The minister is trying to **bring** the two sides **together**.

(b) The minister is trying to **bring** them **together**./*The minister is trying to **bring to-gether** them.

under. The list of possible particles found in phrasal verbs provided in Claridge (2000: 46), which is based on Quirk et al. (1985: 1151), Cowie and Mackin (1975: lxxx), and Fraser (1976: 5) as well as her own data, includes also *aboard, above, ashore, after, astray, asunder, counter* and *forwards*. These particles have been left out in this part of the present study, firstly because they either did not occur in the corpora investigated (*aboard, ashore, astray, asunder*) or only in prepositional uses. *Counter*, for instance, did not occur in TICLE but it did as a verb, as a noun, as a part of an idiomatic phrase and as a prefix in LOCNESS (e.g., *counterarguments, counter-balance*, etc.). *Home* is excluded, following Waibel (2007: 65), since it is considered to be perceived as a noun by learners rather than as an adverb.

As for the productivity of adverbial particles, there have been some statements in the literature. According to Meyer (1975: 5), for instance, the number of the most productive adverbial particles (what he calls “second elements”) used to form phrasal verbs – with his term “two-word verbs” – are seventeen: *about, across, along, around, aside, away, back, by, down, in, off, on, out, over, through, under* and *up* – *up* being the most frequently used adverbial particle in present day American English.⁷⁶ Bolinger (1971: 175) similarly mentions the top-rank productivity of *up* and asserts that it “is the particle with virtually unlimited

⁷⁶ See Meyer (1975: 17-20) for the different meanings of these adverbial particles when used to form phrasal verbs and Meyer (1975: 7) for a brief history of the adverbial particle *up*. For a very clear treatment of aspectual meanings of some common adverbial particles see Bolinger (1971: 96-110).

freedom to attach, roughly comparable to that of the prefix *re-*” (p. 101) (see also Side 1990: 149f). McArthur (1989: 40) mentions the different uses and meanings the particle *up* carries in different combinations – namely, upward direction or approaching direction, completion in the sense that nothing is left or in the sense that something is done as fully as possible, literal or metaphoric emphasis – and notes the fact that it may even have several ‘nuances’ at the same time, as in the case of ‘*Drink up!*’, where it is both completive and emphatic. Biber et al. (1999: 413) show that *up*, *out*, *on*, *in*, *off* and *down* are particularly productive in forming phrasal verbs, combining with the largest number of different verbs – *up* being the most productive particle. The findings of a corpus-based study conducted by Gardner and Davies (2007) accord closely with Biber et al.’s findings: The eight most productive particles in the BNC are *up*, *out*, *back*, *down*, *on*, *off*, *in*, and *over*, in decreasing order (p. 346). Waibel (2007) has also found *out* and *up* to be the most productive particles both in the learner (German and Italian subcorpora of ICLE) and native speaker corpora (LOCNESS). As we shall have an occasion to note later on, the productivity of these two particles also receive support by the results of the present study.

3.2.2 Prepositional verbs

In these combinations, verbs are followed by a preposition in its clear prepositional use, as in *look at*, *come across*, *refer to*, *talk about*, *believe in*, etc. McArthur (1989) refers to this type of verbs as “fused” or “non-separable phrasal verbs” because “the preposition has, as it were, been

‘stolen’ from ^[VP]_[SEP] its own phrase and ‘fused’ with the verb in a new, unique idiomatic relationship”: [*he came*] [*across the street*] vs. [*she came across*] [*an old friend*] (p. 42). The noun phrase that follows the preposition is regarded as its direct object (cf. Quirk et al 1985: 1155ff). Thus, all prepositional verbs are considered transitive. The constituents of prepositional verbs cannot be separated. That means that we cannot put the direct object between the two constituents. The type with the sequence verb-free object-preposition (*turn N into N*) is regarded as a sub-type of prepositional verb – what Biber et al. refer to as ‘Pattern 2’. They are also treated here. Verbs with two prepositions are also included under this type of verbs by virtue of their first preposition, while the second one is ignored (*develop from N into N*, *struggle with N for N*).

In the present study the collocational fixity, i.e. the co-occurrence of one verb with one preposition consistently for a certain meaning, is considered an important criterion in identifying prepositional verbs. The verb-preposition sequences with the preposition having instrumental, spatial or temporal meaning have not been included – the sequences of verb and ‘meaningful’, i.e. lexical, prepositions (Claridge 2000: 62). Thus, the sequence of *play* + *with* in the following sentence (18) is excluded since the preposition in the combination functions as the head of a post-verbal prepositional phrase and has an instrumental meaning:

(18) *The boy is **playing with** the ball.* (instrumental)

Similarly, the underlined combinations in the following sentences are not included in the study, either:

(19) *We have **arrived at** the airport just in time.* (spatial)

(20) *We **arrived on** the morning of the fifth.* (temporal)

Yet the same word sequence in the sentence (18), namely *play with*, is included if its constituents form a single lexical unit, i.e. if the word sequence has a meaning of (a) ‘treating lightly, dallying with (someone; feelings, affections; sensitivities; susceptibilities)’ or (b) ‘considering, but not very seriously (notion, idea, scheme)’, as in the following two sentences (21 and 22):

(21) *She’s not a woman to tolerate being **played with**.*

(22) *He’s often **played with** the idea of emigrating to Canada, but that’s as far as it’s gone.*⁷⁷

As seen in the sentence (21), passive constructions are possible with prepositional verbs.

Prepositional verbs vary in the extent to which the sequence preserves the individual meanings of the verb and preposition. While some prepositional verbs have meanings that are not obvious from the verb alone or from the two constituents together, compared to phrasal verbs, meanings of most prepositional verbs can usually be deduced from the verb

⁷⁷ The sample sentences (21 and 22) as well as the meanings of the combination *play with* are from Cowie and Mackin (1975: 224).

alone (Cowan 2008: 175). In cases such as *talk about* and *wait for*, the meanings of the combination can be derived from that of its constituents. But in cases such as *go into (a problem)*, meaning ‘investigate’, *come by (the book)*, meaning ‘obtain’, it is not possible to deduce the meaning. In this study both idiomatic and non-idiomatic sequences are included, on the same grounds as explained in the case of phrasal verbs. As mentioned earlier on in this section, for the differentiation of prepositional verbs from transitive phrasal verbs, Quirk et al.’s (1985: 1167) syntactic tests are applied.

3.2.3 Phrasal-prepositional verbs

If simply described, phrasal-prepositional verbs are a combination of a phrasal verb with a preposition. The first particle is an adverbial, the second a preposition. Claridge (2000) refers to this group of multi-word verbs as the “offspring” of phrasal verbs and prepositional verbs. Although the preceding two categories are rather common in English, phrasal-prepositional verbs are rather rare and used mainly for physical activities (Biber et al. 1999: 424). According to Quirk et al. (1985: 1160), they are largely restricted to informal English. As they are considered to be colloquial, they are often avoided in formal writing – like phrasal verbs (Gelderen 2010: 93). Biber et al.’s (1999) study has revealed that these verbs are more common in conversation and fiction compared to news and they are particularly rare in academic prose (p. 424). Their “relative absence” in the semantic domains of occurrence,

existence and causative is considered to elucidate their rarity in academic prose.

The underlined structures in the following sentences, which are all taken from COCA, can serve as prototypical examples of this group of verbs:

(23) *I **look forward to** trying your recipe, but it will be months before I can find what I need to create it.*

(24) *They've got ideas and recommendations but have yet to **come up with** a solution.*

(25) *Her managers **put this down to** shift patterns which can mean certain staff have little contact with those that work office hours.*

(26) *Carpooling **cuts down on** carbon emissions by reducing the number of vehicles on the road, but it also means sticking to someone else's schedule and giving up some personal freedom.*

(27) *While this remains a mystery, in the summer of 1986s, thanks to Barbie's party preparation, I was **let in on** one of the greatest female fashion secrets of the modern era.*

(28) *Back in the 60's, bilingualism was **looked down on**, not promoted.*

(29) *There are a few tweaks to this year's rankings, including a deemphasis on the class rank of admitted students (which many high schools are **doing away with**), and on student selectivity overall, and more emphasis on graduation rates.*

(30) *In perhaps the ultimate irony of this story, Kristen is now the coach everyone **looks up to**, the coach she had always hoped Julie would be.*

The first particle following the lexical verb in each sample sentence above (23-30) is an adverbial (*forward, up, down, away, in*), the second a preposition (*to, with, to, on*). In the definition of phrasal-prepositional verbs adapted here the focus is, as in the case of phrasal and prepositional verbs just discussed, on the lexical unity of the combination. That is, the three words occurring together need to really form a lexical unit and have an internal cohesion; they should not just happen to occur next to one another. As Claridge (2000: 64) has noted, it is almost always possible for a phrasal verb to be followed by a prepositional phrase, as illustrated, for instance, in the following sentence from COCA:

(31) *He looks up to the night sky and roars with anger.*

The meaning of the same word sequence in sentence (31) as opposed to the one in sentence (30) is very different – in the former phrasal verb is followed by an independent prepositional phrase, the combination together lacking an internal cohesion. Thus, the semantic unity is the most important reason to regard them as a class of their own.

The semantic unity of most phrasal-prepositional verbs can be manifested, albeit not always possible, in replacement by a one-word verb: For example, *do away with* is equivalent to ‘exterminate’ or ‘abolish’, *come in for* to ‘incur’, *look up to* to ‘admire’, *put up with* to ‘endure’ or ‘tolerate’, *run up against* to ‘encounter’, *look forward to* to ‘anticipate (eagerly)’, *come up with* to ‘produce’ or ‘invent’, *come down with* to ‘develop (an illness)’, and *look down on* to ‘despise’. The existence of a one-word equivalent, according to Quirk et al. (1985: 1160), is a common sign of their idiomatic status.

There are some sequences of verb + particle + preposition that do not display an internal constituent structure. Denison (1981: 29f), for instance, illustrates this by contrasting the sequences *keep on + at* and *put up with*. Regarding the first sequence, Cowie and Mackin (1975) have two entries, namely *keep (on) at* and *keep on (at/in)*. In both cases the constituent in the parenthesis can be omitted, resulting in either a prepositional verb (*keep at*) or a phrasal verb (*keep on*), but with the same or only slightly changed meaning. The definition provided for *keep (on) at* by Cowie and Mackin is as follows: ‘worry, pester, harass (continuously) with suggestions, requests, complaints’, i.e. having an idiomatic meaning. The meaning of *keep on (at/in)* is much closer to the literal end of the cline: ‘not remove (from), continue to employ (at/in)’. In the case of *put up with*, however, none of the constituents can be omitted without resulting in a completely different meaning; omitting *with* from the sequence, for instance, would result in the phrasal verb *put up* for which Cowie and Mackin (1975) have ten different entries.

As for the distinction between their idiomatic versus non-idiomatic status, there is a semantic cline, as in the case of phrasal and prepositional verbs. Whereas the meaning of some phrasal-prepositional verbs can be easily understood from their individual elements (e.g., *keep out of*, *stay away from*), many have figurative meaning (e.g. *stand up for* (‘support’)) and in some cases it is difficult, if not impossible, to derive the meaning of the combination from that of its constituents (e.g., *walk out on* (‘desert’), *put up with* (‘endure’ or ‘tolerate’)). However, as noted

by Claridge (2000: 64), unlike phrasal verbs, phrasal-prepositional verbs do not reach to the entirely literal end of the semantic cline.

Regarding the syntactic features of phrasal-prepositional verbs, as they end with a preposition, they are all followed by a direct object, hence are transitive. In contrast to Fraser (1976: 5), who regards phrasal-prepositional verbs as intransitive verb-particle combinations followed by a preposition, these verbs are regarded here, following Claridge (2000), as a class of transitive lexical units. Some phrasal-prepositional verbs do not allow the object to occur in any other position but only immediately after the preposition, e.g., *We all **looked up to** her* versus **We all **looked her up to*** or **We all **looked up her to*** (what Cowie and Mackin 1975 refer to as ‘intransitive, A3 pattern’). In some cases the direct object is placed before the particle, e.g., *They **filled me in on** the latest developments*, and if the direct object is a noun or short noun phrase, in some cases it can follow the particle, e.g., *They **filled in** their colleagues **on** the latest developments* (Cowie & Mackin 1975: lvi). Still some phrasal-prepositional verbs can take a direct object after the verb as well as an object of the preposition, what Quirk et al. (1985) refer to as ‘Type II’, and Cowie and Mackin as ‘pattern B3’: *fix N up with, put N down to, let N in on, put N up to, take N out on, fob N off with* – N identifying the object noun phrase. These phrasal-prepositional verbs are regarded here as ditransitive verbs.

Other than being intervened by objects, constituents of phrasal-prepositional verbs are not generally separated. In some cases, however, an adverb or an adverbial phrase may occur between the particle and the

prepositional phrase, even when the preposition is not deletable and the sequence is idiomatic: *We **put up** cheerfully **with** these interruptions* (Cowie & Mackin 1975: xlii). Besides, though generally only by the adverbs *right* and *straight*, the verb may be divided from the particle: *The family **came straight up against** fresh problems* (Cowie & Mackin 1975: xlii).

As with prepositional verbs, stranding as well as passivation is possible although passivation of the prepositional object in these verbs is not very common, resulting in most cases in awkward sounding sentences such as **That noise cannot be put up with*. There are, however, some acceptable sentences, e.g., *The death penalty has been recently **done away with*** [‘abolished’] or *Such problems must be squarely **faced up to*** [‘confronted’] (Quirk et al. 1985: 1160). As for the passivation of these verbs, Cowie and Mackin (1975: xlii) note that sentences in which the prepositional object is part of a complex idiom are generally not passivized, e.g., **A good start was **got off to***.

3.2.4 Verb-noun collocations

This group of verbs has been referred to variously in the literature, including too general, broad terms such as “complex verbal structures” (Nickel 1978) or “expanded predicate” (Algeo 1995), and terms that may lead to confusion such as “take-have-phrasal” (Live 1973), along with more transparent terms such as “verb-noun combinations” (Nesselhauf 2005) and “verbo-nominal combinations” (Claridge 2000).

Given the variety of these combinations, different researchers chose different features of the combination as classificatory criteria. For instance, while Live (1973) took the character of the noun as the basic criterion in her three-way classification of verb-noun combinations, Claridge (2000) took the absence or presence of a preposition in the combination and in the case of the latter, its syntactic preposition as the basic classification criterion, which is adapted in this study.

As already mentioned earlier (see Section 3.1), verb-noun collocations are subdivided into three types in the present study: (Group I) simple combination of a verb and a noun – with a possible inclusion of determiners (*make sense, attract attention, lose control, commit a crime, do the washing, take (one's) time*), (Group II) verb-noun-preposition combinations (*make use of, catch a glimpse of, pay attention to, give rise to, put the blame on*), and (Group III) verb-prepositional phrase unit (*come to a conclusion, bring to light, keep in sight, come (in)to (one's) mind*). The following sentences taken from COCA illustrate the investigated verb-noun collocations:

- (I) (32) [...] *time constraints and the limiting effect on the opportunity to **give instruction** were consistently mentioned.* (COCA: 2014: ACAD)
- (33) [...] *so playing it safe financially does **make sense** in general.* (COCA: 2017: MAG)
- (II) (34) *Although some women's groups **took exception to** the call for women to stay at home [...]* (COCA: 1995: SPOK)
- (35) [...] *I do **give credence to** a few reviews that made a related complaint.* (COCA: 2016: ACAD)

(III) (36) *Ken explained that he always **took into consideration** his student's immediate academic well being because [...]*
(COCA: 2014: ACAD)

(37) *All good things **come to an end**.* (COCA: 2019: NEWS)

Despite the variety in the internal make-up of verb-noun collocations, as exemplified in the sentences (32-37), it is still possible to point out a number of fundamental characteristics shared by them all: In all these combinations, it is the second, the nominal part (so-called ‘deverbal (conversion) noun’) that bears the lexical load of the phrase – the ‘action’ or process (e.g., Howarth 1996: 94-97). Considering this nominal part as semantically an extension of the verb, Quirk et al. (1985: 750ff) refer to it as the “eventive object” (e.g., *argue – have an argument*). Cowie (1994: 3169) adds that in verb-noun combinations, “it is the noun which constrains the choice of the verb, rather than the other way about [sic]”. The verbal part is mostly a common verb of general meaning, such as *do, give, have, make* and *take* (see Quirk et al. 1985: 751f for some examples). Although the first part of the combination is usually considered not to play a significant role from a semantic perspective, it is syntactically necessary; it fulfills the syntactic requirements of a verb. As put by Live (1973: 31), “[t]he first part is almost devoid of lexical meaning but embodies the associated grammatical information, being the bearer of the inflectional endings (thus indicating tense, number and person).” Having lost most of its lexical meaning, the verbal part is weakly stressed compared to the deverbal part in verb-noun combinations, making it resemble to an auxiliary – so-called ‘light

verb' (Live 1973: 33; cf. Stein 1991).⁷⁸ In support of this claim, Live (1973: 33) mentions the existence of the "virtually synonymous" pairs of verb-noun combinations including different light verbs – the choice of the light verb being only a matter of dialect (social or regional): *have* or *take a stroll*; *give* or *make a promise* (see also Quirk et al. 1985: 752). However, the claim that the verbal parts have a weakened, delexical meaning has been challenged, for instance, by Stein (1991): Focusing on the verbs *have*, *take* and *give*, Stein has argued against the claim that verbs in verb-noun combinations are semantically 'empty', 'light' or 'weak'; Stein maintains that the verbs in these combinations make very specific contributions to the meaning of the whole combination. She backs up her argument by pointing out how the use of a single verb or a verb-noun combination in the case of imperatives results in a different illocutory force: *run* vs. *have a run*, the first, namely, the bare imperative form, being acceptable in an emergency situation, the latter (*have-constructions*) being more a suggestion rather than a command, expressing personal attention and care for another person (Stein 1991: 25ff). Given this connotational difference, frequent use of delexical verb-noun constructions in informal English (Quirk et al. 1985: 751) can only be expected. In this context, Live (1973: 34f) has noted that verb-noun combinations in general imply "a casual or single occurrence of the action rather than the continuous or the iterative or the unmarked." According to Live (1973: 34), this aspectual connotation of these combinations

⁷⁸ In the present study, the verbs that are used in the formation of delexical verb-noun collocations are referred to as 'light' and 'delexical' verbs interchangeably.

may be the reason for their widespread use. Live mentions two further characteristics of these combinations making them preferable over their equivalent simple verbs: First, as verb-noun combinations can stand alone, not in need of any kind of complement, Live asserts that when complementation is unnecessary and undesirable, these combinations would be opted for instead of single verbs, e.g., *He walked for an hour* vs. *He took a walk*; *He ordered the groceries* vs. *He made an order*, etc. (ibid.). Secondly, verb-noun collocations allow easy injection of an adjectival modifier before the deverbal element, e.g., *have a good laugh*, *take a long chance*, *make a great success of*, etc. (ibid.; see also Stein 1991: 19).

The foregoing represents a general account of verb-noun collocations. Coming back to the three subcategories of these collocations, a few general comments on their characteristic features will be offered in the following.

Some of the combinations in the first sub-type of verb-noun collocations (Group I) – the simple combination of a verb and a noun, without any preposition being involved – can be used both transitively and intransitively (*give instruction* as in sentence (32) vs. *give someone instruction/give instruction to someone*) (see Claridge 2000: 79f). Some of the combinations do not have any articles in their make-up and if there is one, it is in most cases the indefinite article (Nickel 1978: 66; Live 1973: 36) – Live (1973: 36) considers article-less combinations as a sub-pattern of her Type I category. The nouns are usually abstract in nature and have all the formal characteristics of a noun, i.e. they can

take determiners, they can be premodified or postmodified, and they can be pluralized. Yet, as pointed out by Claridge (2000: 78), each combination behaves idiosyncratically. For instance, pluralization is used very rarely in these combinations; whereas it is acceptable with some nouns, with some others it is not allowed at all, e.g. *have little chats* vs. *have a little sit* (Stein 1991: 13).

The combinations in Group II – verb-noun-preposition combinations – are all transitive. A special problem in this context concerns the preposition: As any simple verb-noun combination can theoretically be followed by a preposition, the problem is to decide in which cases the preposition (and in turn the object) is obligatory in these combinations (Claridge 2000: 73). The following sentences from COCA illustrate this problem:

(38a) *As royal governors **took charge**, the New England colonies lost the freedom [...]* (COCA: 2015: MAG)

(38b) *After the Federal Communications Commission **took charge of** regulating the Internet, [...]* (COCA: 2015: MAG)

(39a) *The cab driver **took his time**, but finally appeared.* (COCA: 2015: FIC)

(39b) *They **took their time over** dinner, drinking a bottle of wine and talking [...]* (COCA: 1993: FIC)

As can be seen in the above sentences (38a and 38b), the combination of *take charge* can be used without the preposition *of* resulting in no change in the meaning; the combination does not necessitate a complement. Similarly, *take one's time (over)* can occur with or without further

complement – the existence of the preposition *over* causing no change in the meaning (see also Claridge 2000: 73ff). Following Claridge, in this study the preposition is considered obligatory only if the combination never occurs without it and if the nominal element following the preposition can be regarded to be affected in the way the direct object of a simple verb would be (2000: 73f). Thus, combinations like *take charge (of)* and *take one's time (over)* are classified as Group I.

There are other verb-noun combinations to which addition of a preposition results in a change in the meaning, as seen in the sentences below (40a and 40b), which are taken from COCA.

(40a) [...] *the cigar shops on Broadway that **gave way to** an Apple store and [...]* (COCA: 2017: NEWS)

(40b) *The second plank **gave way** as easily as the first.* (COCA: 2016: FIC)

The meaning of *give way to (something)* in the sentence (40a) is ‘to be replaced by something’, whereas the same verb-noun combination means something different without the preposition: The meaning of the combination in (40b) is ‘to break’. Another meaning of *give way* is ‘to stop arguing or fighting against someone or something’, as exemplified in the sentence (40c) below, taken from BNC:

(40c) [...] *it was not to his liking and I was not prepared to **give way**.* (BNC: 1987: Interview)

Combinations like *give way (to)* belong to two groups (Group I and Group II) and classified accordingly depending on the meaning and existence of the preposition or lack thereof (see also Claridge 2000: 75f).

Along with such borderline cases, there are some clear examples belonging to Group II, such as *take exception to* (35), in which the preposition is an integrated part of the combination. Having said that, as pointed out by Claridge (2000: 76), compared to Group I, Group II is smaller since the number of combinations with a completely integrated preposition is not very high.

Group III includes both transitive and intransitive verb-prepositional phrase sequences, e.g., *keep in sight* and *come to an end*, respectively. In this group it is the verbal part of the combination that determines the (in)transitivity. Some of the transitive combinations require an additional preposition at the end, either to connect the object of the whole unit (e.g., *be in want of N*, *put in touch with N*) or to add another object or complement (e.g., *bring N into contact with N*) (Claridge 2000: 76).

(41) [...] *while the starving multitude **are in want of** the bare necessities of life* [...] (COCA: 1996: ACAD)

(42) [...] *farm labor and other jobs that **bring them into contact with** dirt and dust.* (COCA: 2017: NEWS)

The direct object of the combination usually appears between the verb and the prepositional phrase, e.g., *call N into question*, instead of *call into question N*. The latter option is, however, preferred when the noun phrase is heavy, i.e. long (compare the sentences 43a and 43b below):

(43a) *For an important interpretation that **calls** the traditional view **into question**, see John Cooper, [...]* (COCA: 2016: ACAD)

(43b) *The findings by scientists have **called into question** the traditional downplaying of emotions and intuitions in comparison to reasoning in deliberation.* (COCA: 2016: ACAD)

These three subcategories of verb-noun collocations are to be investigated in the present study, with no pre-defined list. As for the decision of differentiating these collocations from free combinations, and thus for the decision of inclusion in this study, the most important criterion is considered to be the semantic unity – as in the case of other multi-word verbs under scrutiny in the present study. That is, the combination should exhibit a unitary verbal character. Moreover, the noun should be carrying the meaning of the combination. To this end, substitution with a simple verb of the same meaning is used as an aid despite the fact that a simple verb equivalent is not always available, e.g., *make an effort* (see Claridge 2000: 78). The only criterion for the noun in these combinations is that it is abstract in nature, which depends on its use in a given context: For instance, *ear* in *give/lend (an) ear to N* does not represent the biological entity, but metonymic faculty of hearing. That is, a combination like *have an apple* is not included in the study whereas *have a bite* is. Combinations with the verb *to be* are excluded from the analysis of collocations in the present study. The decision was taken mainly on the grounds that the verb *to be* is exceedingly common in the data and, as noted by Nesselhauf (2005: 281), it is highly likely that it is entirely “unrestricted” with regard to its complements.

3.3 Summary

The present chapter laid the foundation by introducing the notion of multi-word verbs, followed by the definition as well as classification of multi-word verbs adopted in this study. Subsequently, detailed information on the criteria used to identify each of the four multi-word verb categories investigated was provided – namely, phrasal, prepositional and phrasal-prepositional verbs and verb-noun collocations. The next chapter will address the data and the method of analysis employed, with one section being devoted to the clarification of how some terms used are to be understood in this study.

Chapter 4: Data and Methodology

*Language is the blood of the soul into which thoughts
run and out of which they grow.*
(Oliver Wendell Holmes)

4.1 Preliminaries

A corpus can be approached in two different ways. It can be investigated with the traditional “hypothesis-based approach” (Granger 1998b: 15) in which a researcher aims to test a particular hypothesis or it can be approached in a more exploratory manner – with the so-called “hypothesis-finding” approach. According to Granger (1998b: 16), the latter approach is particularly suitable and effective for the investigation of computerized learner corpora since it allows gaining new insights into learner language by not limiting the investigation to a particular hypothesis or assumption. In this approach, the data is at first randomly investigated and after getting the first insights into the data, the researcher draws a conclusion as to which aspect of learner language to investigate further. The patterns of data obtained by initial analyses of the chosen aspect of learner language can then be inspected for unusual features, which may then be used to generate hypotheses about learner language. The method adopted for the corpus analysis shapes the results of research, as rightly pointed out by Mair (1991: 67). Thus, the choice between the “quantitative-statistical” and “qualitative-textlinguistic” method (Mair 1991) is a methodological decision that has to be made.

Both approaches have their merits as well as their limitations: The quantitative approach, which is considered to have as its main purpose the quantification of the data, is essential for providing a broad base of insight into, for instance, the frequency of specific structures, in turn enabling the researcher to draw a conclusion on their importance in language use. It cannot, however, explain why a structure is used very frequently or very rarely – unlike the qualitative approach, which is exploratory and investigative in nature and contributes to the in-depth understanding of a context. In addition, a quantitative analysis does not say anything about the appropriateness of a linguistic item in a specific situation. These two approaches are not necessarily polar opposites. In fact, rather than deciding on one method, Mair suggests combining both approaches to gain a deep understanding of language use: “The role of the corpus, after all, is not only to provide a limited and representative data-base for statistical analysis, but also to provide authentic and realistic data, the close reading of which will allow the linguist to approach grammar from a functional and discourse perspective” (Mair 1991: 77). Moreover, writing on methods for detecting formulaicity in language, Read and Nation (2004: 24) claim that “an adequate account of formulaic units as they function in language acquisition and language use can come only from a combination of quantitative and qualitative analyses.”

In the following data analysis, the corpora will be investigated from different perspectives. With the initial hypothesis in mind that learners make less use of some types of multi-word verbs in their writing than

their native-speaker counterparts, due to the various inherent difficulties and complexity of these verbs and the influence of a different L1 verb system (see Section 1.2), the Turkish ICLE component will first be approached from the quantitative point of view in order to determine patterns of under-representation. Afterward, an in-depth qualitative analysis of the Turkish ICLE component with respect to multi-word verb usage will be carried out. That is, collocational aspects of multi-word verbs will be investigated in order to determine whether there are any unusual cases or deviations in their use and if so, whether they can be explained with L1 influence or any other variables at hand, i.e. extralinguistic factors on the learners' written output such as length of exposure to English, other known (foreign) languages, circumstances of text production, essay topic and years of English learned. As learners are expected to operate more according to Sinclair's 'open choice principle' rather than the 'idiom principle' – in Kjellmer's (1991: 124) words, as learners' "building material is individual bricks rather than prefabricated sections" – it is also hypothesized that learners will display idiosyncratic use of multi-word verbs in their writings. Thus, in this study quantitative and qualitative aspects of corpus-linguistic research will be combined in order to gain in-depth insights into Turkish learners' inter-language regarding multi-word verb usage.

Statistical tests will be applied, wherever possible, in order to validate the significance of findings. The results of the quantitative and the qualitative analysis of the data will be presented in chapters 5 and 6, respectively.

4.2 Corpora

The nature of the learner sample on which the study is based may influence the characteristics and the distribution of the errors (Ellis & Barkhuizen 2005: 57f). Claridge (2000: 5) accordingly states that “[t]he more one knows about a corpus, the easier it is to put data derived from it into the proper perspective.” Therefore, this section will be devoted to providing detailed information on the two corpora investigated in this study.

As the database for the present study, the Turkish component of the *International Corpus of Learner English* (henceforth called TICLE) will be used. The *International Corpus of Learner English* (ICLE) was initiated by Sylviane Granger in 1990 and as a first large-scale collection of learner language, it aimed at encompassing countries in which English is learnt as a foreign language (the “expanding circle” in Kachru’s (1985: 12) terminology) – that is, at representing EFL context.⁷⁹ The ICLE corpus is the largest learner corpus that is publicly available,

⁷⁹ As mentioned earlier (see fn. 1), in SLA literature, there is a traditional distinction made between the terms ‘foreign language’ and ‘second language’ – the former refers to “the learning of a nonnative language in the environment of one’s native language” and the latter to “the learning of a nonnative language in an environment in which that language is spoken” (Gass and Selinker 2008: 7). Gilquin and Granger (2011: 56) draws attention to the fact that the situation in reality is more complex since the degree and type of exposure vary to a great extent in the two learning contexts. Although ICLE was collected in such a way as to represent EFL, some subcorpora of it contain texts written by learners who studied English in a context closer to ESL (Gilquin & Granger 2011: 56). On the basis of data from ICLE, the researchers suggest viewing the EFL and ESL distinction as a continuum rather than as a dichotomy, with many in-between categories corresponding to a variety of learning contexts (ibid.).

i.e. non-commercial, and contains argumentative essays produced by higher intermediate to advanced learners; the third, current version of the ICLE corpus (ICLEv3) includes data from 25 mother tongue backgrounds, amounting to over 5.5 million words.⁸⁰ Moreover, it is the only available corpus for the learner group under investigation, i.e. Turkish learners of English.⁸¹ Following the Contrastive Interlanguage Analysis (CIA) methodology (Granger 1996b) briefly touched on in chapter one, the TICLE will be compared to the native speaker control corpus, the *Louvain Corpus of Native English Essays* (LOCNESS). The benefits of L1-L2 comparisons from a pedagogical point of view have already been noted by Granger (2015: 14): “they [L1-L2 comparisons] provide language teaching professionals with precious information on what learners do right or wrong or partly wrong in a particular skill or task, which can be used to inform a wide range of pedagogical applications.”

⁸⁰ All the other existing learner corpora either focus on one learner group or they are commercial, i.e. not available for academic research. Cf. the Hong Kong University of Science and Technology (HKUST) Learner Corpus (25 million words of L1 Chinese learner English) and the TeleNex Student Corpus (nearly 3 million words of Chinese learner English). Besides ICLE, to my knowledge, there are only four other learner corpora that contain more than one learner group: the Longman Learner Corpus, the Cambridge Learner Corpus, Montclair Electronic Language Learners Database and the Corpus of Academic Learner English (CALE) – the latter is not commercial but the compilation of the Turkish sub-corpus has not been finished yet. See the regularly updated list of learner corpora around the world provided by the team working at the Centre for English Corpus Linguistics of Université catholique de Louvain: <https://www.uclouvain.be/en-cecl-lcworld.html>

⁸¹ The Corpus of Academic Learner English (CALE) contains a sub-component of Turkish learners of English but at the time of conducting this study, it was still being gathered.

Both TICLE and LOCNESS include argumentative written productions.⁸² As rightly pointed out by Nesselhauf (2005: 46), “any investigation based on a corpus has to be careful with generalisations that go beyond the text type(s) represented.” As argumentative essays do not generally contain any specialized vocabulary and display medium level of formality, they are rather neutral in register and style and that is why it is probable that the difficulties or problems identified would also occur in most other text types (Nesselhauf 2005: 46). In the following, two corpora will be presented in more detail, along with some points of importance as I go along.

In the compilation process of the ICLE project, several universities collaborated in order to reach the target size of 200,000 words per national corpus. As for the data collection process of TICLE, three universities collaborated: Essays were collected at the universities of Çukurova, Mersin and Mustafa Kemal University; from second and third year undergraduate students studying at the Language and Linguistics department in the Faculty of Science and Letters (Mersin University) and at the department of Foreign Language Education in the Faculty of Education (Cukurova University and Mustafa Kemal University) (Kilimci & Can 2009: 209).

⁸² Lorenz (1999) is of the opinion that the characteristic patterns of advanced learner language are best investigated in written production. For his sound reasoning behind this opinion, see Lorenz (1999: 11).

TICLE consists of 280 essays, totaling 199,532 words (Granger et al. 2009: 37).⁸³ All of the essays are argumentative in nature. The average length of essays in TICLE is 713 words (Granger et al. 2009: 6). All of the essays were written untimed; that is, they were not part of an examination. Slightly more than half of the students (52.1 percent) were allowed to use reference tools during the writing process. Stated more precisely, 134 essays (47.9 percent) were written without the help of the reference tools and 146 essays (52.1 percent) were written using reference tools (Can 2009: 25). However, neither these reference tools nor what was actually looked up by the learner was recorded. In terms of this study's concern, this information would have been of great value. Thus, this aspect of the corpus design of TICLE can be seen as a limitation for a lexical study like this one.

The average age of learners in TICLE is 22.08 years and 81 percent of the learners are female (Granger et al. 2009: 8f). As undergraduate learners, the years spent at the university studying in English range between 3 and 6 years (the variable 'years of English at the university').

⁸³ The WordList function of AntConc gives a different number than the one provided by the compiler: 203,745. The default definition of 'word' in AntConc is as 'letters' in the broadest sense, meaning numbers, dashes, etc. were not included in the overall number of tokens. This might be a possible reason for this discrepancy found in the overall counts. As the data was searched via AntConc tools, the total number of tokens provided by its Wordlist function was taken as the base in this study, which, in turn, does not affect the results in any way.

Languages spoken at home other than Turkish were also recorded and are listed in decreasing order of use as first, second or third ‘language spoken at home’. 98 percent of the participants (276 learners out of 280) have Turkish as their mother tongue – four participants stated other languages to be their mother tongue but did not specify what these languages are – for these learners, Turkish is the second language spoken at home. Regarding the first language spoken at home, 271 of the learners named Turkish and nine learners ‘others’, again not specifying which language is spoken. 255 (91.1 percent) of the learners were recorded to state that no second language was spoken at home whereas eight participants were recorded to state that Turkish was the second language spoken at home. Only one learner stated that English was the second language spoken at home (cf. Table 1 in Can 2009: 21 for more details on the distribution of languages in TICLE).

The majority of the learners – 276 of the learners – had not been in an English-speaking country. Four learners were recorded to have spent 4 months, 5 months, 8 months, and 144 months, respectively, in an English-speaking country.

In the data, the variable ‘years of English learning spent at school’ differs the most between the participants; it ranges between 4 and 13 years. However, all in all, it can be stated that the data investigated in this study is quite homogenous.

As for the essay topics, although it cannot be stated that a single topic dominates TICLE, the most favored topic is ‘Sex equality’, followed by ‘Most university degrees are theoretical and do not prepare students for the real world. They are therefore of very little value’ – both topics were chosen by 57 learners for their essays. The least favored topics in the data were ‘violence on TV’, ‘air pollution’ and ‘premarital sex’, all being chosen by only one learner (see Table 3 for the details of the distribution of all essay topics in TICLE).

Table 3. Distribution of essay topics in TICLE

Topics	Frequency	Percent
1 Capital punishment	7	2,5
2 Suicide	4	1,4
3 Sex equality	57	20,5
4 Cheating in colleges	17	6,1
5 Violence on TV	1	,4
6 Air pollution	1	,4
7 Great inventions and discoveries of 20th century and their impact on people's lives	23	8,3
8 Premarital sex	1	,4
9 In the words of the old song	30	10,8
10 Most university degrees are theoretical and do not prepare students for the real world. They are therefore of very little value.	57	20,5
11 Abortion	15	5,4
12 Euthanasia	21	7,6
13 Nuclear Energy	6	2,2
14 Divorce	12	4,3
15 Freedom of Press	9	3,2
16 Human rights	2	,7
17 Animal testing	11	4,0
18 Values and consequences of school interaction	4	1,4
Total	278	100,0

Regarding the proficiency levels, the very fact that there are differences in the proficiency levels of a learner group is natural and unavoidable. As claimed by Carlsen (2012), proficiency level has been a fuzzy variable in learner corpus research. The ICLE is, as referred to by Carlsen (2012: 164f.), a ‘one-level’ corpus, the underlying claim of which is that all texts are at the same level of proficiency. It was designed to sample the argumentative writing of advanced foreign learners of English and in order to ensure that all learners were advanced, essays were mostly collected from third or fourth year undergraduate students of English (Granger et al. 2009: 11).⁸⁴ That is, for the level assignment in ICLE project, external criteria, i.e. non-linguistic factors, were opted for (Atkins et al. 1992: 5). In Carlsen’s (2012: 165) terminology, a ‘learner-centered’ methodology was applied: Learners were considered to be advanced because of their institutional status (cf. Thomas 1994: 317; Tono 2003: 801). However, although one would justifiably expect Turkish learners to be also quite advanced learners of English due to

⁸⁴ Lorenz (1999: 10) comments on the ambiguity of the term ‘advanced’ learner as follows: “Unfortunately, it is not entirely clear what exactly qualifies a learner to be ‘advanced’. There seems to be a methodological paradox inherent in this notion: if we are seeking to determine the exact characteristics of advanced learners’ infelicities, we cannot yet determine the advanced level on strictly linguistic grounds.” He therefore bases his definition on external factors and inductive reasoning: “advanced learners are learners who have to meet advanced foreign language requirements, i.e. learners who are *generally expected* to have mastered the basic rules and regularities of the language they are learning” (ibid.).

their institutional status⁸⁵, the proficiency level of this group of learners is quite low compared to that of other learner groups in ICLE, e.g., Swedish, Bulgarian, Dutch or Finnish, which is perfectly understandable given the differences in learning environments. As stated by Aas (2011: 48), differences in educational systems and the quantity as well as the quality of exposure to English are significant factors determining the learners' language output. Besides the formal instructional input learners get, learners rarely associate or communicate with native speakers of English outside the classroom in Turkey (see Section 1.3). They are of course exposed to the English language through the Internet, English movies and other media but it should be noted that although the Internet today as an integral part of our daily routines can function as a "source of potential contact" with English, the ICLE data were collected before its use became so widespread and popular (Gilquin & Granger 2011: 57). That is why most probably the learners from whom the TICLE data was collected had not been exposed to much natural English via the Internet and other media.

A quick look at the essays from different subcorpora was sufficient for Granger et al. (2009) to notice that ICLE includes differences in proficiency level, both across and within subcorpora. Therefore, with the aim of getting a better understanding of the proficiency differences existing

⁸⁵ In Turkey someone who wants to study at a department of Foreign Language Education or Language and Linguistics Department needs to take, i.e. pass, a language test in addition to national written university entrance exam (see Section 1.3).

across subcorpora in ICLE, Granger et al. (2009: 11) submitted a random sample of twenty essays from each ICLE-subcorpus to a professional rater to rate the essays on the basis of the *Common European Framework of Reference for Languages* (CEFR) descriptors for writing. Granger et al. note that these results should be considered with caution due to the reduced size of the sample as well as the fact that only one rater rated the essays (ibid.). Regarding TICLE, the results show that out of 20 essays rated, sixteen were found to be B2 or lower level and the remaining four were rated as C1 level, none of the essays reaching the C2 level. Thus, it would indeed be more accurate to state that the proficiency level of learners represented in TICLE actually ranges from intermediate to advanced. This fact about the level of learners under scrutiny makes the study more interesting because in studies of writing that focus on lexicon, as Engber (1995: 150) asserts, it is this population of learners – intermediate to high-intermediate level of proficiency – who are the ones that “can provide unique insight into the language learning and writing process” because at this stage learners have not yet reached the point at which their active vocabulary growth stops – what Laufer (1991: 445) calls “active vocabulary threshold” – but they have sufficient language knowledge to maximize hypothesis testing and this should be seen in their language productions (cf. Lorenz 1999: 10; Ringbom 1987: 57). Engber (1995: 150f) compares the intermediate and advanced writers in terms of lexical usage as follows:

at the intermediate level, lexical variation is a productive strategy for expressing content, whereas linguistically advanced

writers may rely more on precision. It may also be the case that the intermediate writer is still trying out and adding to his or her stock of lexical items, resulting in higher lexical variation in conjunction with higher quality writing. The advanced writer, on the other hand, may be retrieving items that meet precise specifications from an already adequate lexical base, resulting in less variety but perhaps relatively high quality writing.

This is a position with which I can relate to and it proves the data at hand suitable for the present study since one of the interests here is to investigate how learners compensate for their lack of lexical knowledge.

As must be evident by now from the preceding chapters, multi-word verbs are closely connected with the idiomatic aspect of language, which is characteristic of native speaker language. That is why native speakers are included in the present study as a control group; learners' use of two multi-word verb categories which have been noted to be avoided and underused by some learner groups in the literature (see Sections 2.3.4 and 2.3.4.2) – namely, phrasal and phrasal-prepositional verbs – will be compared to that of native speakers so as to see whether and how they differ.

In the case of comparing the language of native speakers and that of non-native speakers, a vital issue, as stated by Granger (2002: 12), is the choice of a comparable control corpus. Along the same lines, Barlow (2005: 345) states, “[i]f a corpus is to be contrasted with an NS

corpus, then a variety of issues arise, as they always do when corpora are compared”, such as dialectal variety, genre, register or text type. Granger (1996b: 45) warns against the temptation to use existing computer corpora of English since “in view of the genre sensitivity of many linguistic features, comparison may be flawed.”⁸⁶ Besides the choice of a dialectal variant and a diatypic variant, Granger (2002: 12) adds that the level of proficiency of the native speakers should also be taken into consideration during the selection process of the control corpus “to avoid inadvertent comparisons between novice and professional writers” (Hasselgard & Johansson 2011: 38). Some researchers have already voiced their criticism about the use of native expert writing as a reference variety in learner corpus research. Lorenz (1999), for instance, considers comparing learners’ argumentative essays with journalistic and academic writing found in commercially available corpora to be unfair and “descriptively inadequate” for the reasons of text-type and writing proficiency.⁸⁷ McCrostie (2008: 101) expresses a very similar view:

Direct comparisons of published academic articles with non-native (and perhaps even native) English student writing are unfair. The authors published in academic journals might be best viewed as

⁸⁶ This danger is indeed shown very clearly in a study of connectors in native and learner writing by Granger and Tyson (1996: 23f).

⁸⁷ That is why, in his study investigating the intensification of adjectives, Lorenz (1999) designed corpora of learners and native speakers to include more and less mature writers in them; the written productions of intermediate and advanced learners of English with L1 German were compared to native pupil and college student writing. See Lorenz (1999) for the advantage of using a ‘graded corpus.’

highly skilled Formula-1 racecar drivers and few people prove capable of handling such sophisticated and impractical vehicles. Highly advanced non-native English writers, such as graduate students writing theses in English, may feel comfortable driving the family sedan but not yet confident enough to get behind the wheel of a high performance sports car. Meanwhile, beginning and intermediate non-native English writers are drivers who just received their learner permits and in some cases do not see the point in learning how to drive at all.

Taylor (1986: 144f) and Ringbom (1987: 73-78) suggested that many problems faced by novice native speakers in the process of writing were comparable to nonnative writers' problems (cf. Chen & Baker 2010: 44). Hyland and Milton (1997: 184) are also of the same opinion and show that native and nonnative students share a number of novice writer characteristics. On the basis of these proposals and findings, it has been suggested using a corpus of native student writing as a reference rather than native expert writing in learner corpus studies. More recently, addressing this very issue, Gilquin and Granger (2011) have indeed demonstrated that novice native writers share features with both expert native speakers and non-native writers, EFL learners displaying a number of unique characteristics that are not found in novice native writing.

For the ICLE project, the issue of choosing a comparable reference corpus was overcome with the compilation of LOCNESS, which fairly closely matches ICLE for text type, age and experience; LOCNESS contains 300,000 words in total and it comprises both argumentative and literary essays written by British and American students (Granger

1998b: 13).⁸⁸ The essays in the American part of the corpus were written by students from Marquette University, Indiana University at Indianapolis, Presbyterian College in South Carolina, the University of South Carolina, and the University of Michigan; the British part consists of British A-level essays (60,209 words in total) and university student writing – written by students of the University of Surrey in Guildford. For the above-mentioned similarity found between the language productions of novice native writers and that of L2 learners, but especially for its closeness to the learner corpus, LOCNESS will be used as the native reference corpus in the present study. One thing worth noting is that the A-level argumentative essays of LOCNESS are excluded in the present study in order to keep the native speaker and learner data as close to one another as possible. That is, the part of the native speaker corpus taken into consideration for the present study is composed exclusively of university student writing. Excluding A-level essays, LOCNESS is composed of a total of 322 essays of which 90 were written by British university students, 232 by American university students.

⁸⁸ Hasselgård and Johansson (2011: 38) note that in LOCNESS there is less information available on the contributors compared to ICLE and the texts in LOCNESS are more heterogeneous regarding the essay topics as well as contributors (both university students and A-level pupils). Nevertheless, the researchers add that “[st]ill, LOCNESS remains the best available comparable corpus to match ICLE and continues to be widely used.”

Regarding the learner corpus, it is important to note that two of the essays in TICLE – TRKE2043 and TRKE2069 – are exactly the same, although the provided background information and the date of the essay submission on the ICLE-CD were different. They are, therefore, excluded from the data. After the tags are hidden and the two texts excluded, the number of the data is 201,013 words. Four of the remaining 278 essays in TICLE were written by learners whose native language is unspecified; Turkish was indicated as the second language spoken at home by these learners (TRCU1039, TRCU1052, TRCU1096 and TRCU1145). Nevertheless, it was decided to include their data in the study as they live and study in Turkey – regardless of the school type and medium of instruction, all schools in Turkey are obliged by law to teach the Turkish language.

Following Waibel (2007), in order to determine the total number of tokens in the two corpora under investigation, WordList statistics was applied. Due to the fact that file headers/essay codes were not removed from the essays and that the corpora are minimally annotated with respect to deleted quotes (<*> or <quote>), deleted bibliographic references <R>, and illegible words <?> (cf. Granger et al. 2009: 12-13), the option ‘hide tags’ was selected. That is, anything between angle brackets was not counted. However, as the tag for illegible words is not separated by a space from the illegible word in question, this tag was analyzed manually in order to include those words (which resulted in two words in LOCNESS). The total number of words in LOCNESS is

264,903⁸⁹ and in TICLE 201,013. Table 4 outlines the quantitative information on the two corpora.

Table 4. Corpus information

	Tokens total	Number of essays	Average lengths of essays (tokens)
LOCNESS	264,903	322	823
TICLE	201,013	278	723

One advantage of ICLE is, as mentioned earlier in this section, that circumstances of language production (whether the essays were written timed and reference tools were allowed during writing process) and learner characteristics, such as age, sex, years of English at school/university, other (foreign) languages known to the learner, or exposure to the target language in an English-speaking country, have been recorded which in turn allows the researcher to investigate whether or not these variables affect the use of the linguistic feature in question. As for the corpus size, TICLE can be thought to be rather small compared to currently available native corpora but it is relatively large as far as learner corpora are concerned. Besides, as manual analysis is needed in the present study, dealing with a larger corpus would hamper the precise analysis required. However, it should be admitted that the size is not big enough for a study of lexical analysis. That is why, the results obtained,

⁸⁹ Due to programming differences between WordSmith and AntConc, the total number of words for LOCNESS given here differs from the one in Waibel (2007: 62).

especially those regarding the individual verb combinations, should be seen as preliminary.

As the two corpora compared in this study have different sizes, frequencies are normalized to one million words in order to guarantee the comparability of data. The reason for choosing the figure of one million words as a reference point was practical: For the analysis of two sub-categories of investigated multi-word verbs, namely phrasal and phrasal-prepositional verbs, the results of Waibel (2007) for the control corpus LOCNESS were used. Thus, all frequency figures given in this study are frequencies normalized to this standard, except where indicated otherwise. If the raw frequencies attained from corpora are low, they are always indicated.

4.3 Tools

In this study various tools have been used to retrieve and analyse the data and to test the research hypotheses.

4.3.1 Software for Retrieval and Analysis

In order to carry out the data analysis in the present study, AntConc Tools version 3.4.3 was used. This software package consists of a suite of tools for lexical analysis, two of which are especially useful for the present study: Word List and Concordancer. Word List provides the researcher with fundamental textual statistics of a corpus such as the overall number of tokens, types, and type/token ratio. Furthermore, it generates a list of all the words that appear in a corpus, sorting them into

alphabetical or frequency order. While this tool is very beneficial for the quantitative analysis of a corpus, it is less useful for the qualitative analysis because the list of words appearing in the corpus are listed in isolation, without their surrounding words in the text, and for the analysis of language use with regard to each of the listed words, context is necessary. As stated by Barnbrook (1996: 65f), “[t]he frequency list is very useful as a means of isolating words from the surrounding detail of the text so that they can be surveyed in this way, but the lack of this detail also prevents us from seeing precisely how these potential labels are actually used.” Therefore, in order to analyze language use in detail regarding each of the listed words in their original context, the Concordancer Tool of AntConc was used. Oakes (1998: 149) defines a ‘concordance’ as “a list, arranged in an order specified by the user, such as the order of appearance, of the occurrences of items in a source text, where each occurrence is surrounded by an appropriate portion of its original context.” The Concordancer Tool of AntConc has a wide range of features and its main purpose is “to show *how* a search term is used in a target corpus” (Anthony 2005: 731, emphasis in the original); it displays the words or phrases selected in their original context and in turn allows an in-depth analysis and conclusions about their usage. Hunston (2002: 9) states that by means of bringing together many instances of use of a word or phrase concordance lines allow “the user to observe regularities in use that tend to remain unobserved when the same words or phrases are met in their normal contexts.”

These two tools – Word List and Concordancer – serve well for re-searching language use at both a quantitative and qualitative level. This combination thus allows a comprehensive insight into the frequency and the use of multi-word verbs under investigation in their discourse context.

Microsoft Excel (2013) software was used as a tool to create databases of multi-word verbs and perform the functions of computing and graphing. SPSS (Version 17.0) was used to carry out the statistical analysis.

4.3.2 Reference Tools

In the identification of multi-word verbs in the learner corpus, in addition to general dictionaries of English, dictionaries of collocations and of phrasal verbs were consulted: *Cambridge English Dictionary Online* (2020), *Oxford Advanced Learner's Dictionary of Current English* (2000), *Collins COBUILD Dictionary of Phrasal Verbs* (1989), *Oxford Dictionary of Current Idiomatic English. Volume 1. Verbs with prepositions and particles* (1975), *Chambers Dictionary of Phrasal Verbs* (1996), *Macmillan Phrasal Verbs Plus* (2005), *Oxford Collocations Dictionary for Students of English* (2002) and *LTP Dictionary of Selected Collocations* (1997). Those multi-word verbs attested in the learner corpus but not in the dictionaries consulted were subsequently investigated in two general corpora of English in order to assess their acceptability – namely the *Corpus of Contemporary American English* (COCA) and the *Corpus of Global Web-Based English* (GloWbE).

COCA, a balanced corpus of American English, contains over one billion words of text and cover eight genres: spoken language, fiction, popular magazines, newspapers, academic texts, TV and movie subtitles, blogs and other web pages. The time frame the corpus covers is 1990-2019 and it is updated regularly – which makes it a valuable benchmark for measuring appropriateness of language use considering the ever-changing nature of language. GloWbE is based on web pages (general pages and blogs) and comprises 1.9 billion words of written English. Unlike COCA, which consists of one variety of English, GloWbE is composed of regional varieties of English from 20 different countries around the globe: Australia, Bangladesh, Canada, Ghana, Great Britain, Hong Kong, India, Ireland, Jamaica, Kenya, Malaysia, New Zealand, Nigeria, Pakistan, Philippines, Singapore, South Africa, Sri Lanka, Tanzania, and United States. The reason for referring to these two corpora – COCA and GloWbE – in addition to the dictionaries is not only their large size, which adds to the reliability of the study. As for COCA, its design allows studying the current and continuing lexical changes in English. As mentioned in Section 3.2, some categories of multi-word verbs – namely, phrasal verbs – are creative and progressive. One other related notion in this context justifying the inclusion of GloWbE is ‘New Englishes’ – it refers to new varieties of language that have become localized. These new Englishes which have their origin in the colonial era are considered varieties of English in their own right and as Jackson and Amvela (2007: 37) state “[v]ocabulary is the area in which these new Englishes best assert themselves.” The fact that GloWbE

consists of regional varieties, rather than only inner varieties (the “inner circle” varieties in Kachru’s (1985: 12) terminology), makes it a good point of reference because it allows an investigation of a given multi-word verb in many different varieties of English, given the expected creativity in learner language.

Whenever a multi-word verb extracted from the learner corpus was not attested in the aforementioned sources, it was considered ‘unrecorded’.

4.4 ‘Acceptability’ or ‘Appropriateness’ and the Role of the Native Speaker

It is at this point useful to clarify one methodological aspect. Although general-language reference corpora, which are based on factual, authentic evidence, are specifically beneficial for non-native researchers of English by enabling them to make reliable statements about Standard English, they have their limitations in judging nonnative English (Lorenz 1999: 18). If a combination of a verb and its collocater produced by a learner could not be attested in the aforementioned reference tools, it has been considered ‘unrecorded’ but it does not necessarily mean that such a combination is unacceptable. As Lorenz (1999: 18) states, “[i]t can be rightfully classed as ‘unlikely’, or perhaps even ‘unnatural’ or ‘unidiomatic’, but the ultimate authority on acceptability still rests with the native speaker”⁹⁰ and the validity of this statement does not change despite the inevitable fact that the opinions of native speakers differ to

⁹⁰ See Lewis (2002: 53) for ‘dreaded Edna’ (educated native speaker).

4.4 ‘Acceptability’ or ‘Appropriateness’ and the Role of the Native Speaker

a great degree, especially as regards the acceptability of formulaic language. Drawing attention to the “middle-ground” performance of advanced learners, i.e. language use which is neither obviously erroneous nor fully nativelike, rather “infelicitous” in a way, Lennon (1991b: 185) reported that native speakers’ judgments for many of the ‘doubtful’ cases obtained from the speech of advanced learners expressed disagreement. Similar observations were made by Mondor (2008: 174f).

Some subcategories of multi-word verbs present a more complicated area in terms of their acceptability; for instance, in contrast to prepositional verbs, where the preposition following the verb is either correct or incorrect (e.g. **care at*)⁹¹, in the case of, let’s say, phrasal verbs, one has to deal not only with the phrasal verb itself, but also with its frequently polysemous meanings and the context in which it can be used, i.e. its appropriateness. As Corder insightfully (1981: 40) comments, “[j]udgements about the appropriateness of an utterance require that we interpret it in relation to its context and the situation in which it is uttered. Appropriateness has many dimensions and cannot [...] be reduced to rules. Judgements about appropriateness must therefore be largely subjective.” However, it is important to note the well-known discrepancy between native speakers’ acceptability judgments and their

⁹¹ Indeed, the non-standard use of prepositions, their omission as well as their simplified uses are between the commonly cited features of various New Englishes, such as Kenyan English (Mwangi 2004), Pakistani English (Rahman 1990), Indian English (Mukherjee 2010) and Malaysian English (Percillier 2016), to name just a few.

actual language use: As put by Nesselhauf (2005: 53), “there is not necessarily a one-to-one relation between what native speakers find acceptable or unacceptable when explicitly asked about a certain language phenomenon and what they themselves produce frequently.” Being aware of these facts, which can be seen as a pitfall by some, in the present study that has been conducted by a nonnative English speaker, all judgements of acceptability, unless found in the aforementioned reference tools, have been subsequently evaluated by four native speakers of different English varieties in order to increase the validity of the conclusions. Three of the native-speaking judges are females and one male; two Americans, one British and one Irish. They are aged between 28 and 72, and they have all had tertiary education. One of these native-speaking judges is an experienced EFL teacher and one grew up bilingually and has had foreign language learning experience but is not professionally concerned with analyzing or teaching language(s). In this context, one further thing to note is that native speakers vary in their use of some multi-word verbs. Therefore, the acceptability judgments should be considered as a matter of degree, as suggested by Nesselhauf (2005: 39).

4.5 The Extraction of Multi-Word Verbs

The major methodological concern at hand, as is always the case in the investigation of discontinuous strings, was the question of how to extract multi-word verbs that are of interest, i.e. the general issue of precision and recall. Maximizing the precision may end up not including

some relevant data whereas trying to maximize the recall would mean the inclusion of irrelevant data. As mentioned earlier in section 4.2, the corpus used in this study is only minimally annotated with respect to quotes, bibliographical references, and illegible words, but not for parts of speech. Moreover, the automatic error recognition has not yet reached a point where it can deal with human language, which is, in Mair's (1991) words, "extremely supple, flexible and context-bound." A fully automated approach regarding especially some groups of multi-word verbs investigated in this study, i.e. phrasal verbs, is not feasible anyway since a particle might be tagged as a preposition or particle and as prepositions they are used highly frequently in any English corpus. Disambiguating all occurrences would therefore be remarkably tiresome and time-consuming. On the basis of these facts, a semi-automatic approach had to be opted for in this study. What follows is a specific application to this study of the general problem of optimizing precision and recall in corpus searches.

The first step was to extract all verbs in TICLE. To this end, the word list of the learner corpus was generated by using the Word List tool of AntConc computer software; this tool lists each word occurring in the corpus either according to its frequency or in an alphabetical order. This list obtained was then analyzed in order to extract all possible verbs. In the verb list all items that had potentially been used as verbs were first included although some of them could just as well had been used as nouns, adjectives, or adverbs, e.g., *even, right, reason, power, couple, long, side, rule, back, perfect, single, lower*, etc. Both BrE and AmE

spellings of words were included in the list to be further investigated in their context, e.g., *ageing and aging, favor vs. favour*, etc. As modal auxiliaries are not relevant to the interest of the present study for the reason that they do not combine with any other item to form a multi-word verb, they were not included in the analysis. All forms of the verbs *be, do* and *have* were considered. There were many wrongly spelled words in the word list. All typos were checked in their contexts in order to determine for which word they might stand; to see whether they were used as a verb. As some nouns were found to be used wrongly as verbs, e.g. *suicide*, most of the words in the word list were indeed investigated. It was sometimes not easy to differentiate a verb form of a word from its adjectival form at first glance in the list, e.g. *impending*. In those cases, the word was analysed in its context of use. With the completion of these steps, the verb list of the learner corpus was also finalized. This list was then used as a basis to extract the multi-word verbs that are of interest to the present study.

Before moving on, one thing that is worth noting at the outset is the fact that in order to assure that all occurring multi-word verbs in the learner corpus were accounted for, no pre-established list for the process of extraction was used. Basing the study on a pre-established list, based on existing lists, such as the one proposed by Biber et al. (1999: 410) or by Waibel (2007) for phrasal verbs⁹², would not allow an in-depth analysis

⁹² After investigating different registers (conversation, fiction, news, and academic prose), Biber et al. (1999) lists 31 highly frequent phrasal verbs. Waibel (2007), on the

of multi-word verb use by the target group of learners. This decision was taken on the grounds that learners, as known, display phraseological deviation from the natural use of language; that is, some multi-word verbs may be used by learners which are not included in any such lists (see Wang 2016: 26). During extraction of multi-word verbs false positives are deleted.

The extraction of prepositional verbs and verb-noun collocations was done fully manually. In the following the details of the steps taken in the semi-automatic analysis applied for the extraction of phrasal and phrasal-prepositional verbs are outlined, followed by a few remarks on the extraction process of prepositional verbs and verb-noun collocations. Details on the patterns of each multi-word verb category investigated in the present study can be found in Section 3.2 under the relevant subcategory of multi-word verbs. Further methodological aspects crucial to the quantitative and qualitative analyses of the data will be discussed in the preliminary sections of Chapters 5 and 6.

other hand, in order to get a first overall impression regarding the usage of phrasal verbs by learners, identified the ten most frequent main verbs likely to form part of a phrasal verb in three corpora she investigated (LOCNESS, German and Italian subcorpora of ICLE), which mostly correspond to the productive verb list proposed by Biber et al. (1999: 413). After having added the other verbs listed by Biber et al. (1999: 413) to the verbs found in her data, Waibel combined all the productive verbs with the particularly productive particles for the formation of phrasal verbs – namely, *down*, *in*, *off*, *on*, *out*, and *up*, which are also listed by Biber et al. (1999: 413). This resulted in a list of 72 phrasal verbs (see Waibel 2007, appendix 2 (p. 185-187)). Please note that Waibel includes the phrasal-prepositional verbs under the term ‘phrasal verb’.

4.5.1 The extraction of phrasal and phrasal-prepositional verbs

The extraction of phrasal and phrasal-prepositional verbs in the data was done together and semi-automatically. The previously created verb list from TICLE via the use of the Word list tool of AntConc was used as a basis for the extraction of these multi-word verb categories: all the verbs in the list were to be examined in their context by using the Concordancer tool so as to extract all phrasal and phrasal-prepositional verbs occurring in the corpus. In order to avoid going through all concordance lines for all the verbs listed in the verb list, the search was limited by specifying an adverbial particle within the range of five words on the left side and the right side of the search word, i.e. verb (see section 3.2 for these adverbial particles).⁹³ The next step was to manually examine the concordance lines to extract phrasal and phrasal-prepositional verbs. The extracted verbs were subsequently analyzed in the light of the research questions (see Section 1.2). Although this semi-automatic approach is very laborious, it is still less time consuming and requires less effort than conducting the search with the selected 28 adverbial particles alone (see Section 3.2.1) or only the verbal elements.

⁹³ Sinclair (1998: 15) notes that the optimal distance for two co-occurring words in English seems to be not more than four intervening words. However, albeit seldom, in the data investigated, cases like the following exist: “*Consent then, is the issue that I will **base** the moral permissibility of euthanasia **on***” (TRCU1173).

There are a few further remarks to be made on the process of extraction of these multi-word verbs. As noted earlier in chapter 3, some categories of multi-word verbs have fuzzy boundaries, making it sometimes hard to categorize them. In the case of verb + *out of* sequences, for instance, it is practically impossible to conclusively determine the status of *out* in individual cases. Therefore, following Claridge (2000), all the combinations of verb + *out of* were excluded in this study, e.g. *keep out of*, *put out of*, *get out of (patience)*, etc.

During the extraction of phrasal verbs, there were a few cases where it was not possible to determine whether the attested phrasal verb was an intentional choice or a typo, as can be seen in the following sentence (44).

(44) *As it is known, in youth period people have got too much problems and they **seek away** to solve their problems.*
<TRKE1036>⁹⁴

Albeit not ideal, this and similar attested word sequences were considered as they were found in the data since there was no way to determine

⁹⁴ A large corpus of different varieties of English, namely GloWbE, was investigated to see whether any uses of *seek away* as a phrasal verb exist: 6 occurrences of *seek away* were found – one hit in Australian English, two hits in Kenyan English, one hit in Irish English, one hit in Singaporean English and one in Hong Kong English. Whereas some of these occurrences can be read as *seek a way* in their context, some cannot. The facts that more than half of GloWbE (about 60 percent) consists of informal blogs and it is not proofread are known to the researcher but the main reason of consulting to this corpus was the lexical variation and creativity found in the outer circle varieties of English.

the intended word sequence by the learner – this problem could only be overcome by means of elicitation, i.e. looking for an ‘authoritative interpretation’ by asking the learner what s/he intended to say (Corder 1974), leaving no room for doubt but this was unfortunately not possible in the present study due to the nature of the data used (see section 4.2). As Arabski (1979: 26) points out, “[e]rror analysis should be based on a corpus of errors elicited by a technique which guarantees that [...] the error analyst knows not only what a subject said (or wrote) but also what he wanted to say (or write).” Excluding such questionable cases, on the other hand, would mean not fully investigating the learner language productions at hand. As for the differentiation between phrasal and phrasal-prepositional verbs, although there were some clear cases such as *put up* vs. *put up with*, there were other cases in the data where it was not easy to differentiate these two verb types, especially at first glance. In some studies these two verb types are grouped together, as the line between them is not always very clear-cut. However, in this study it was thought that it should not be the case as in-between cases are seen in all aspects of language. In order to differentiate these two categories, Cowie and Mackin’s (1975) dictionary was consulted since it is very conclusive and provides information on the first three multi-word verb categories investigated in the study, namely phrasal verbs, prepositional verbs and phrasal-prepositional verbs. For instance, according to Cowie and Mackin (1975) the definition of *go back to* as a phrasal-prepositional verb is to “have its origin in; date from/back to”. The meanings of *go back* as a phrasal verb provided by Cowie and

Mackin (1975), on the other hand, are as follows: “return [to school, to the library, etc.]; return to an earlier point in space, time, a discussion, etc.; (of clocks and watches) be set to an earlier time in order to allow for changing hours of daylight”. On the basis of this information, out of the eight attested occurrences of *go back (to)*, seven occurrences were counted as phrasal verbs, and only one (the second hit in Figure 1) was counted as a phrasal-prepositional verb as it was the only one with the meaning of ‘having its origin in’. Thus, during the classification of the extracted multi-word verbs, the context played a determining role in this study.

Hit	KWIC	File
1	eds are very different .If you go back to the early years ,inv	TRCU1033.txt 32 1
2	he root of lack of self-esteem goes back to the childhood. If	TRCU1037.txt 36 1
3	look for responsibles, when we go back in history, we will see	TRCU1088.txt 87 1
4	ndividual and so that they can go back into society and be ber	TRCU1104.txt 103 1
5	r one. At this point, let\x92s go back to our question. How mu	TRCU1115.txt 114 1
6	e is gone away and you want to go back. Absolutely impossible!	TRCU1122.txt 121 1
7	to go. Many women still had to go back to the cradle of violer	TRKE2010.txt 186 1
8	his better, it is important to go back to primary, secondary c	TRME3027.txt 273 1

Figure 1: Concordance lines retrieved for *go back (to)* in TICLE

The differentiation between transitive phrasal verbs and prepositional verbs was mainly based on the Quirk et al.’s (1985: 1167) criteria (see Section 3.2.1). As the extraction of phrasal verbs was semi-automatic – with a list of 28 adverbial particles (see Section 3.2.1), all concordance lines for the entire verb list were gone through manually at the end in

order not to exclude any prepositional verb. Thus, no prelist of prepositions was used. By means of a careful manual filtering process, all instances were eliminated in which the prepositions did not occur within the verb phrase or constituted false positives.

4.5.2 The extraction of verb-noun collocations

As mentioned earlier, the extraction of verb-noun collocations in the data was done entirely manually in order not to exclude any collocations. The previously created verb list from TICLE via the use of the Word list tool of AntConc was again the basis for the extraction of the collocations and the concordance lines retrieved were examined manually to filter the irrelevant instances.

During the extraction differences in determiners in verb-noun collocations were disregarded. Premodifiers other than determiners such as adjectives and nouns were not included for further analysis. That is, only the head noun was taken into consideration. For instance, the instances of *affect + the/your/human + health* were extracted as *affect (one's) health* or the instance of *make a hard and merciless journey* was extracted as *make a journey*. If the existence of the premodifier resulted in meaning change (e.g. *take place* vs. *take one's place*), they were then extracted and listed as two entries.

In addition to the most unambiguous pattern in which the head verb is followed directly by its object NP, other syntactical patterns in which verb-noun collocations occur were also extracted from the data – as exemplified in examples (45) to (47). In the first two sentences the verb-

noun combinations (*threaten one's life*, *take care*) occur in a passive clause whereas the latter sentence exemplifies a pattern that involves a zero relative clause.

(45) *If her **life** is **threatened** physically the ideal is to save both of them.* <TRCU1091>

(46) *Following the same idea as before, great **care** must be **taken** in distinguishing when someone is choosing death [...]* <TRCU1075>

(47) *At the end of the lesson most of the **knowledge** they **acquired** is staid [sic] on the notebooks.* <TRCU1965>

The instances consisting of a verb and a pronoun (e.g. *it*, *one*, *them*, etc.) were included only if it was straightforward which antecedent the pronoun in question stood for, as in examples (48) and (49), from which the combinations *answer a question* and *cure an illness* were extracted, respectively. The instances where it was questionable or not possible to pinpoint the antecedent were disregarded.

(48) *This is a difficult question, and **it** must be **answered** in historical terms.* <TRCU1115>

(49) *Also the x-rays are used to understand the cause of illness or to **cure them**.* <TRCU1019>

One last remark before we move on is the occasional long distance between verbs and their object NPs, as exemplified in (50) and (51). These instances clearly benefit from a manual analysis of the data.

(50) *Just think of a sick person who goes to a hospital to feel better and he learns that this **illness** is very bad and too difficult to **cure**.* <TRCU1122>

(51) *In this section, the universities **enter** and play an important role in **our life** for adaptation.* <TRCU1125>

4.6 Some words on terminology

In the investigation of learner language, there are some terms that one cannot dispense with, namely, ‘norm’, ‘native speaker’, ‘non-native speaker’, ‘overuse’, ‘underuse’, ‘error’, ‘mistake’ and ‘(in)correct’ use. With no intention of an exhaustive theoretical treatment of these issues, a few words should be added here on the use of these terms in order to clarify how they are to be understood in this study.

The traditional strict adherence to a native speaker norm in the context of ELT has been challenged by the debates on World Englishes and English as a Lingua Franca in recent years. The issue of defining the ‘norm’ or the question of whether a ‘norm’ is acceptable at all has become controversial. As a result of the rapid globalization of English, the strict division made between English varieties in Kachru’s framework⁹⁵ has recently proven to be too simplistic to reflect the complex reality of English use in different environments. Gilquin and Granger (2011), for

⁹⁵ In Kachru’s framework of the concentric circles of World Englishes, each circle represents the types of spread patterns, acquisition and functional domains of use of English.

instance, have demonstrated that there are many in-between categories between ESL (“outer circle”) and EFL (“expanding circle”) varieties due to the complexity of the language learning process and its many contextual determinants such as the varying degree and type of exposure to TL (see also Edwards and Laporte 2015; Edwards and Lange 2016). Hong Kong English as a variety, for instance, has been documented both in learner corpora and in ESL corpora. Accordingly, recent years have seen the emergence of unifying theoretical approaches to World Englishes such as Mair’s (2013) “World System of Englishes” model.

The common sense idea of the concept of a native speaker refers to “people who have a special control over a language, insider knowledge about ‘their’ language”, knowing what can be said in a given language and what not; thus, “[t]hey are the models we appeal to for the ‘truth’ about the language” (Davies 2003:1). Accordingly, in most of the SLA research, the native speaker has often been taken as the ‘touchstone’ for learners and as a means of evaluating learners’ proficiency (e.g., Cook 1997: 38) – both in the era of Error Analysis and, to a large extent, in the current Learner Corpus Research. As Flowerdew (2015: 469) notes, “[a] key facet of learner corpus research is that the learner corpus is usually compared with a native-speaker control corpus”. As mentioned earlier, one component of the CIA is a comparison between learner data

and native speaker data – the comparison type employed in this study.⁹⁶ Given the position of English as a world language, apart from the two major native varieties of English – namely, British and American English – which have been prevalent in ELT as standard native-speaker norms, there are now other, new Englishes that have developed or are developing their own standards that could also serve as ‘norms’ or ‘models’ in the context of ELT.⁹⁷ For instance, Indian and Nigerian English have developed their own set of norms. However, as far as the teaching of English in Turkey is concerned, where English has no official status but is increasingly taught and learned as a foreign language, the British and American English varieties have traditionally functioned as the standard norms in the education system and they will most probably remain the most widespread norms in the field of ELT in Turkey (cf. Dogancay-Aktuna 1998: 31). Therefore, these two major varieties are chosen as the points of reference in the current study (this choice

⁹⁶ Bley-Vroman (1983), in his paper demonstrating the detrimental effect of describing a learner’s language by a concern with the target language norms, introduced the term ‘comparative fallacy’; he states that “work on the linguistic description of learners’ languages can be seriously hindered or sidetracked by a concern with the target language” (Bley-Vroman 1983: 2). As a result of the criticism leveled at CIA over the years and in order to be more in line with the current state of foreign language theory and practice, Granger has broadened the definition of CIA to include the notion of expert variety alongside the native variety (see Granger 2012a, 2012b) and has revised the CIA method which now includes a wide range of “reference language varieties” (Granger 2015): In addition to the traditional inner circle varieties, the new version of CIA method now “also incorporates the possibility of using outer circle varieties as well as corpora of component L2 user data” (Granger 2015: 17).

⁹⁷ See Nesselhauf (2005: 37f) for a brief discussion of the question of the norm in English language teaching.

implies no value judgment). Another reason for this decision is the very fact that the control corpus chosen for the study, namely LOCNESS, consists of the writings of British and American undergraduate students. However, it should be born in mind that ‘norm’ is used here only as a point of reference, not as a set of rules which must be followed at all costs. As novice writers, the native speakers in the reference corpus themselves still make mistakes when they write, especially when it comes to academic writing.⁹⁸ Furthermore, as mentioned earlier (see section 4.3.2), in addition to language reference tools and LOCNESS, two other corpora – COCA and GloWbE – will be referred to should the need arises.

Similarly controversial are the terms ‘native speaker’ and ‘non-native speaker’. In learner language studies, ‘native speaker’ is commonly used to refer to those who were born and brought up in an English-speaking country and ‘non-native speaker’ to refer to those who are learning or have learnt English as a foreign language. However, from a sociolinguistic and purely linguistic point of view this native/non-native distinction is open to debate (Medgyes 1992: 341).⁹⁹ Rajagopalan

⁹⁸ Bolton et al. (2002) challenge the assumption that the best ‘target model’ for non-native students is the writing of native-speaker students. They maintain that “[t]he target norm in academic writing, for both ‘native’ and ‘non-native’ students is better defined as academic writing itself, and the best texts for comparison are clearly those already published in international English-language academic journals” (2002: 173).

⁹⁹ Medgyes (1992: 341) illustrates this controversial issue by an exemplification as follows: “Indeed, countries where English is a second language break the homogeneity of the native/non-native division. The trouble is that this division does not always apply in so-called native English-speaking countries either. Let us take Juan, for example,

(1997), for instance, refers to the concept underlying this distinction as the “myth of nativity” and by pointing out the racial and discriminatory stance behind the notion of ‘native speaker’, he states that it is “a potentially dangerous ideological agenda” (p. 229; cf. Davies 2003, 1991). Ferguson (1982: vii) argues that “the whole mystique of native speaker and mother tongue should probably be quietly dropped from the linguists’ set of professional myths about language.” Rampton (1990) and Paikeday (1985) argue along the same lines. As pointed out by Medgyes (1992: 341), “[e]fforts to define native competence or native-like proficiency have yielded inconclusive results at best”, resulting in attempts to eliminate this distinction, as well as in alternative terms and concepts to replace it. However, the alternative terms suggested, e.g. ‘educated English speaker’, ‘more or less proficient users of English’ (Paikeday 1985), ‘more or less accomplished communicators’ (Edge 1988), ‘experts’ (Rampton 1990), or concepts such as language expertise, language affiliation and language inheritance (Rampton 1990), “are no less spurious than the concept of the native versus non-native speaker” and they have not been clearly explained (Medgyes 1992: 342). In the present study, therefore, the terms ‘native speaker’ and ‘non-native speaker’ shall be kept, intending no imperialistic assumption about the

aged 9, who has been living in the United States for five years. His father is a Mexican immigrant, his mother comes from Norway. They both speak to Juan in their own mother tongue. Which is his native language, English, Spanish, or Norwegian? All three of them? None of them?”

ownership of English (cf. Tan 2005). The terms ‘non-native speaker’ and ‘(foreign) language learner’ will be used here interchangeably.

Other terms commonly used in learner language research which may need explanation on how they are to be understood in this study are related to the concept of ‘error’: ‘mistake’, ‘erroneous’, ‘incorrect’, ‘wrong’ and ‘unacceptable’. They are all used in this study to mean a linguistic form that is ‘nonstandard’, i.e. unlikely to be produced by a native speaker. With the use of these terms no other implication, such as errors or erroneous uses are bad or they need to be eradicated or avoided, is intended here. No distinction is made in this study between ‘errors of performance’ (what Corder (1967) refers to as ‘mistakes’) and ‘competence errors’, which are systematic errors of the learner due to lack of language knowledge. As this distinction is hard to make in practice, and impossible with the nature of the data at hand¹⁰⁰, this distinction is thus not adopted here; the terms ‘error’ and ‘mistake’ are used interchangeably and will be used to refer to clearly unacceptable combinations, just as the words ‘incorrect’, ‘wrong’, ‘unacceptable’. Accordingly, combinations will be considered acceptable (or ‘correct’, ‘right’ and ‘appropriate’) if they are acceptable in either one or both varieties taken as norms. Whether or not a learner mixes features from both varieties in her/his essay is not taken into consideration. Closely

¹⁰⁰ Referring to Kellerman’s (1976) study, Arabski (1979: 19) notes that only by means of special elicitation techniques it is possible to determine whether a given deviation is a mistake or error.

related to the notion of ‘error’ is the negative influence of the first language (or other known languages) on the second language production. This influence is commonly referred to as ‘interference’ and ‘negative transfer’. These terms will be retained in the present study.

In some cases, words like ‘inappropriate’ and ‘deviant’ will be necessary to use; for example, if a learner uses a phrasal verb with an incongruous context word, e.g. **carry out knowledge/occupation*. Albeit understandable, such combinations are nonstandard and unnatural. This unnaturalness will be verified by means of standard dictionaries – the reference tools mentioned in Section 4.3.2 and native-speaker judgments (see Section 4.4).

A further controversial issue is the use of the terms ‘overuse’ and ‘underuse’, which are relevant for the present study. Although they are now well-established terms in learner corpus research (LCR), meaning technically “containing more or less than”, they have been criticized in the literature (Granger 2015: 19). These two terms are retained in this study and are used neutrally as descriptive terms, without any negative connotations. They simply point to the differences between the two corpora investigated, referring to the fact that learners use a particular multi-word verb more or less frequently compared to their native counterparts. Thus, these terms do not carry any notion of inappropriateness or wrongness.

4.7 Summary

This chapter provided information on the methodology employed in this study. The learner corpus TICLE and the reference corpus LOCNESS – the two corpora used as the database in the present study – were presented in detail. Issues of comparability between corpora were discussed in terms of text type, size and learners (proficiency levels in TL and writing experience). Reference tools as well as tools used for the data retrieval and analysis were presented. Some explanations on the relevant terminology were provided in order to clarify how the terms used are to be understood in this study. The notions of ‘acceptability’ and ‘appropriateness’ and the role of NS judgments were briefly discussed. Having established the necessary background on the methodology and data, I will now proceed to the results of the investigation.

Chapter 5: Multi-word verbs in learner writing – a quantitative approach

*One forgets words as one forgets names.
One's vocabulary needs constant fertilizing or it will die.*
(Evelyn Waugh)

5.1 Preliminaries

The data-based, analytical part of the present study is divided into two major parts as quantitative and qualitative analyses of the data respectively. Before starting with the quantitative analysis of the data, some points require clarification. As already mentioned in chapter 4.2, the two corpora under investigation have different sizes. Throughout the paper, all frequencies found in the learner corpus were therefore normed to counts per million words (pmw), which is a common procedure in corpus linguistics allowing comparability between corpora of different sizes. The decision in favor of the number chosen was taken mainly on the grounds that the figures found by Waibel (2007) for the control corpus regarding the phrasal- and phrasal-prepositional verbs will be used as a comparison. Normalized figures are rounded to the nearest full percentage for easy reading. Absolute figures are mostly provided in parentheses. Some analyses necessitate absolute numbers; in the cases where only absolute figures are provided it is indicated.

As mentioned earlier (Section 4.2), on the ICLE-CD all learner data, including the setting associated with each individual learner's language production, is recorded and can be evaluated by means of a search screen. That is, the CD allows researchers to see whether a given learner used a reference tool during the writing process, the time of 'exposure' to the target language, the time spent on learning the target language, other known languages to the learner, etc., which, in turn, enables the investigation of the possible impacts of these variables on learners' productions – in the case of the present study, it allowed me to investigate their possible impacts on the (non-)use of multi-word verbs by Turkish learners of English under scrutiny. To this end, first Excel spreadsheets are generated recording all learner essays with variables provided on the ICLE-CD, which are then expanded with the verbs found in the data. During the analysis, wherever possible and needed, findings are corroborated by the chi-square (χ^2) test.¹⁰¹ In the present study, as for the level of significance, a distribution is assumed to be significant if p (the probability of error in rejecting the null hypothesis) is smaller than or equal to .05 ($p \leq 0.05$).

Multi-word verbs in the data are quantified as a whole, not according to their different semantic meanings although most of them, especially phrasal verbs, are polysemous (see Section 3.2). This decision was

¹⁰¹ It should be noted that although in some tables normalized figures are to be seen, results from the chi-square test are based on absolute figures as this is necessary for this test to be valid.

taken mainly on the grounds that in the context of learner data the quantification according to semantic meaning is not very feasible since there are many cases in which the learners' use of a multi-word verb deviates from one of the various dictionary meanings or the meaning intended by the learners cannot be reliably established. Besides, the number of some multi-word verb categories investigated is just too high for a semantic analysis, as pointed out earlier in chapter 3.

During the analysis of phrasal and phrasal-prepositional verbs, common alternative one-word equivalents to these words, if there are any, will also be investigated in an attempt to explain the cases of over- and underuse. Similarly, whether the common single verbs used by learners have a multi-word verb equivalent will be investigated. That will allow the determination of whether the learners used phrasal or phrasal-prepositional verbs rather than their single-word equivalents when they had the option to use either one of them. However, it should be borne in mind that the data emerged from essay writing. That means, learners would have actively chosen multi-word verbs or one-word verbs if they had been provided the option, for example, in a multiple choice test. Whether this would be the case can only be established by means of elicitation data, as has been shown, for example, by Dagut and Laufer (1985), Hulstijn and Marchena (1989), and Liao and Fukuya (2004). Therefore, the investigation of the use of one-word equivalents of phrasal and phrasal-prepositional verbs should be considered only as an indicator of preference of one over the other. Although learner production is considered to provide the clearest evidence of what a learner has

acquired, the process of developing explanations for the language produced by the learner is “necessarily one of inference” and one way of ascertaining whether an inference has any explanatory value is by collecting verbal report data (Ellis & Barkhuizen 2005: 21f.). Thus, in an ideal case, learners would be interviewed individually regarding their choice of verbs and asked to comment on their own productions, i.e., eliciting explanations for their language use. However, as the present study is a corpus-based study, this option is unfeasible.

The main purpose of this chapter is to report the frequency results attained from the data analysis – both token and type frequency. That is, we answer the question of how frequently multi-word verb types occur in the learner corpus investigated and how productively the learners employed them. Nevertheless, some qualitative comments will be made at relevant points since frequency information alone, albeit necessary, is not enough to explore potential reasons and implications of the findings.

In the chapter only selective examples are displayed to illustrate a given point. An unacceptable or doubtfully acceptable form or usage is preceded by an asterisk (*).

The remainder of the current chapter is organized as follows: Section 5.2 focuses on phrasal and phrasal-prepositional verbs. Raw and normalized frequencies of phrasal and phrasal-prepositional verbs found in the learner corpus are presented first and are then compared to that of

the reference corpus. This comparison is followed by information on type frequencies, focusing on the most common phrasal and phrasal-prepositional verbs in the data. Section 5.2.4 addresses the productivity of verbs and particles used in the formation of phrasal and phrasal-prepositional verbs. The last section on these two verb categories investigates whether it is feasible to account for the findings attained by the extralinguistic factors available on the ICLE-CD. Section 5.3 focuses on prepositional verbs. First, the overall type and token frequencies are provided, followed by a discussion of the most frequent prepositional verb types. Thereafter, the productivity of verbs and prepositions used in forming prepositional verbs in the data are presented. Section 5.4 focuses on verb-noun collocations. Similar to the preceding section, overall quantitative results are provided first, followed by a discussion on the most commonly used verb-noun collocations in the data. Next, productivity information on the verbs (based on the lexical variety of noun collocates) is presented. The chapter ends with some concluding remarks in Section 5.5.

5.2 Phrasal and phrasal-prepositional verbs in TICLE

The methodological aspects of how phrasal verbs were extracted from the data were already discussed in chapter four (see Section 4.5.1). As a reminder, phrasal and phrasal-prepositional verbs are quantified independently of inherent semantic differences because there are cases in which the meaning of a multi-word verb deviates from one of the various dictionary meanings and where a clear-cut semantic differentiation

is not possible. Moreover, in some cases learners use phrasal and phrasal-prepositional verbs in wrong contexts and may assign them inappropriate meanings in a given context.

In the following, the results of the quantitative analysis are presented.

5.2.1 Overall results regarding phrasal and phrasal-prepositional verb usage

As the aim of this part of the study was to investigate the overall use of phrasal and phrasal-prepositional verbs by Turkish learners of English, the whole data was investigated in an exhaustive manner, following the methodology described in chapter four; all phrasal and phrasal-prepositional verb tokens were extracted from the learner corpus (TICLE) in order to guarantee an in-depth analysis. The results were then compared with the findings attained for native speakers by Waibel (2007).

The initial hypothesis was that Turkish learners would make less use of phrasal and phrasal-prepositional verbs in their writing than their native speaker counterparts, given their manifold semantic, syntactic and stylistic inherent difficulties, and the different verb systems of the learners' mother tongue and that of the target language. The results of the quantification of all phrasal and phrasal-prepositional verb tokens in the data clearly confirm this hypothesis. In total, 1,378 instances of phrasal and phrasal-prepositional verbs were reported for LOCNESS (1,371 phrasal and 7 phrasal-prepositional verbs), corresponding to a relative fre-

quency of 5,197 per one million words (Waibel 2007: 83, see also Appendix 3, pp. 188-196). In contrast, only 458 phrasal and 64 phrasal-prepositional verbs were attested in the learner corpus TICLE, together corresponding to a relative frequency of 2,596 per million words.¹⁰² That is, learners used approximately only half as many phrasal and phrasal-prepositional verbs compared to native speakers. It can thus be concluded that, overall, Turkish learners underused phrasal and phrasal-prepositional verbs in comparison to their native speaker counterparts to a great extent.

Regarding the phrasal and phrasal-prepositional verb types in both corpora, while Turkish learners used 170 different types, native students used 422 different ones. The type/token ratios of phrasal and phrasal-prepositional verbs for both corpora are, however, about the same: for LOCNESS the type/token ratio is 30.6, for TICLE it is 32.5 percent. According to the chi-square test, the difference between the two corpora

¹⁰² Being rather surprised at the very low overall frequency of phrasal-prepositional verbs reported for native speakers by Waibel (2007) – namely two phrasal-prepositional verb types were used seven times in total (*come down to* (4) and *do away with* (3)), I cursorily investigated the native speaker corpus and realized that some of the verbs considered phrasal-prepositional verbs here (e.g. *pass on to* or *stand up for*) were regarded as phrasal verbs by Waibel. This does not cause any change in terms of quantitative results though. Indeed, as will be mentioned later on, some of the phrasal-prepositional verbs listed for TICLE were actually erroneously used phrasal or prepositional verbs, i.e., likely target verbs by learners were phrasal or prepositional verbs (e.g., **bring about with* for *bring about* or *come across with* for *come across*). As mentioned in the previous chapter, all multi-word verbs attested in TICLE were considered as they were found in the data – not in terms of their likely targets.

is statistically not significant.¹⁰³ This result translates into the fact that both groups make use of nearly the same degree of repetition of phrasal and phrasal-prepositional verbs in their written productions. As learners are known to show a tendency to repeat certain words and expressions that are familiar to them (see Section 2.2), this finding was expected for the learner group but having nearly the same amount of repetition in the use of these multi-word verbs by the native speakers was surprising.

On the whole, however, the frequencies of the individual phrasal and phrasal-prepositional verbs in TICLE were often too low for meaningful conclusions to be drawn. Moreover, 79 out of 135 phrasal verb types were attested only once in the learner data, viz. 17 percent of the total phrasal verb occurrences in the learner data are hapax legomena. The same holds for phrasal-prepositional verb use by learners: 25 out of 35 phrasal-prepositional verb types were attested only once in the learner data. That is, 39 percent of the total phrasal-prepositional verb occurrences in the learner data are hapax legomena. Taken together, nearly 20 percent of all phrasal and phrasal-prepositional verbs attested in the learner data were used only once. Furthermore, as will be discussed at more length in the next chapter, 17 out of 64 phrasal-prepositional verb tokens were used erroneously at the surface level, i.e. phrasal verb use with additional prepositions (e.g. **come across with*, **bring about with*), wrong choice of preposition (e.g. **give up from* for *give up on*)

¹⁰³ χ^2 : LOCNESS: TICLE is 0.347. The *p*-value is 0.555. This result is not significant at $p < 0.05$.

or reversed order of the adverbial particle and preposition (**come with up*).

With the methodology outlined in Chapter 4, not only the quantification of all phrasal and phrasal-prepositional verbs was possible but also of main verbs in general, which in turn allowed the determination of the proportion of multi-word verbs in relation to the overall amount of verbs (see Figure 2 below).¹⁰⁴ The proportion of phrasal and phrasal-prepositional verbs with respect to the overall number of verbs in terms of percentages has been found to be 2.5 percent in TICLE while this proportion has been reported to be 4.7 percent in LOCNESS (Waibel 2007: 84).¹⁰⁵

¹⁰⁴ Modal auxiliaries and *be*, *do*, and *have* were not quantified here. The decision in favor of excluding modal auxiliaries was by and large made on the grounds that modals cannot be counted as independent verbs since they qualify main verbs and, moreover, they do not combine with particles. In addition, *be*, *do* and *have* were excluded from the quantification since the figures obtained for phrasal and phrasal-prepositional verbs in TICLE are to be compared to those obtained by Waibel for LOCNESS. Occurrences with an adjectival character (e.g., *be lost*) and *going to*-future occurrences were also omitted from the verb counts, following Waibel (2007).

¹⁰⁵ Phrasal and phrasal-prepositional verbs formed with the verbs *be*, *do*, and *have* are included in these figures although the main and auxiliary uses of these verbs were not quantified (see the preceding footnote).

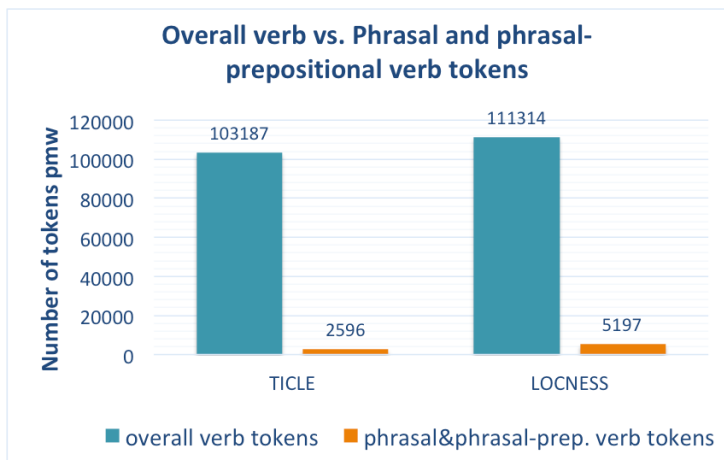


Figure 2. Total number of verb tokens compared to the total number of phrasal and phrasal-prepositional verb tokens in TICLE and LOCNESS

The significance of this identified difference between the learners and native speaker students in their usage of these two multi-word verb categories in relation to the overall number of verbs has been further corroborated by the chi-square test.¹⁰⁶

5.2.2 Frequency of individual phrasal verbs

By combining the 28 particles (see Section 3.2) with all possible verbs extracted from the learner corpus, all phrasal verbs were retrieved and then ordered according to their frequency in the learner corpus. In what follows the overall quantitative comparison between the learners and the native speakers is presented first, followed by a discussion of those

¹⁰⁶ χ^2 : LOCNESS: TICLE $p=0$

phrasal verbs that are shared and not shared by the two groups. In Table 5, the most frequent 25 phrasal verb types found in TICLE are listed along with the frequency of the same phrasal verbs in LOCNESS.¹⁰⁷ Note that although all forms of the verbs were investigated during the search for these phrasal verbs in the data – for example for the verb ‘go’ the forms ‘go’, ‘goes’, ‘went’, ‘gone’ and ‘going’ were all included in the search¹⁰⁸ – only the base forms are given in the tables. All the figures mentioned in the table are normalized (n/million words) and rounded for ease of reading. In order to provide a better idea to the reader, raw numbers are given in parentheses, since, as will be seen, the frequencies of the individual phrasal verbs in TICLE are often very low.

A detailed interpretation of the qualitative use of phrasal verbs will be provided in the following chapter. Some striking observations that can be made from Table 5, however, may not seem out of place at this point.

First of all, the list of the most frequent phrasal verbs of the learner corpus is headed by the phrasal verb *go on*, which is also at the top of the list in the reference corpus (LOCNESS). This finding is in line with Biber et al.’s (1999) finding: In LGSWE, *go on* was found to be the most common phrasal verb overall (Biber et al. 1999: 410). It occurs in all registers investigated – conversation, fiction, news and academic prose – serving a number of different functions and based on the overall

¹⁰⁷ The list of all phrasal verbs according to their frequency can be found in Appendix 1.

¹⁰⁸ In the cases when a verb form is misspelled, such as *graving* for *growing* or *growe* for *grow* it was included as well in the search list.

frequencies compared in the four registers, *go on* is very common in conversation, fiction, news, but less common in academic prose (Biber et al. 1999: 410f). Although *go on* serves a number of different functions in native speaker data (see Biber et al. 1999: 411), in TICLE it is mostly used in the meaning of ‘continue’ (47 clear cases where the intending meaning by the learner is ‘to continue’). This finding is interesting as this meaning of *go on* (‘to continue’) is given as the sixth meaning in the OALD. However, *continue* is a commonly provided one-word alternative to *go on* in the phrasal verb lists prepared for ESL learners. Therefore, the figures for *go on* with the frequencies for *continue* was contrasted and the results showed that both groups behaved in the same way and opted for one-word verb option significantly more often than the phrasal verb option.¹⁰⁹ However, native speakers’ preference for one-word verb was higher compared to the learners. Expressed in percentages, whereas the one-word verb option accounts for 78 percent of native speakers’ use, it accounts only for 61 percent in the case of learners.

¹⁰⁹ Only the instances of *go on* which can be substituted by *continue* are considered and the result is LOCNESS:TICLE $\chi^2=10.14$, p-value= 0.001453. p <0.002.

5.2 Phrasal and phrasal-prepositional verbs in TICLE

Table 5. The 25 most frequent phrasal verbs in TICLE compared with LOCNESS

Rank in TICLE	Phrasal Verb	Frequency in TICLE pmw (raw)	Rank in LOCNESS	Frequency in LOCNESS pmw (raw)
1	<i>go on</i>	284 (57)	1	201 (53)
2	<i>sum up</i>	199 (40)	395	4 (1)
3	<i>grow up</i>	174 (35)	8	98 (26)
4	<i>go out</i>	144 (29)	21	42 (11)
5	<i>bring up</i>	114 (23)	5	110 (29)
6	<i>carry out</i>	104 (21)	2	182 (48)
7	<i>give up</i>	85 (17)	9	95 (25)
8	<i>come together</i>	35 (7)	59	19 (5)
9	<i>go back</i>	35 (7)	14	61 (16)
10	<i>find out</i>	30 (6)	11	72 (19)
11	<i>get on</i>	30 (6)	73	15 (4)
12	<i>point out</i>	30 (6)	3	155 (41)
13	<i>take away</i>	30 (6)	4	117 (31)
14	<i>wake up</i>	30 (6)	69	19 (5)
15	<i>keep on</i>	25 (5)	-	0
16	<i>turn out</i>	25 (5)	23	42 (11)
17	<i>build up</i>	20 (4)	33	30 (8)
18	<i>come back</i>	20 (4)	42	27 (7)
19	<i>come over</i>	20 (4)	-	0
20	<i>lock up</i>	20 (4)	-	0
21	<i>bring about</i>	15 (3)	10	87 (23)
22	<i>break out</i>	15 (3)	132	8 (2)
23	<i>break up</i>	15 (3)	213	4 (1)
24	<i>bring back</i>	15 (3)	29	34 (9)
25	<i>come out</i>	15 (3)	30	34 (9)

Table 6. *continue* versus phrasal verbs *go on*, *carry on* and *keep on*

	LOCNESS	TICLE
	n/million words (raw)	n/million words (raw)
<i>go on</i>	148 (39)	224 (47)
<i>carry on</i>	42 (11)	5 (1)
<i>keep on</i>	0	25 (5)
<i>continue</i>	515 (136)	358 (72)

Two further phrasal verbs worth mentioning in this context are *carry on* and *keep on*, meanings of which are roughly the same as *go on* and which can be paraphrased by *continue* and found to be provided as the formal alternative to these two phrasal verbs for English language learners.¹¹⁰ Thus, they are also included in Table 6.

As can be seen, even adding up the figures for these three phrasal verbs did not change the fact that the one-word verb option was opted for by both learners and native students in most instances.

A cursory glance at the learners' essays showed that some learners used both types of verbs (namely, *go on* and *continue*) within the same essay interchangeably (e.g., TRCU1010 and TRCU1086).

Despite the fairly high frequency of *go on*, the learners in quite a few cases did not manage to use this phrasal verb correctly; there are many

¹¹⁰ See page 8 on the following link:

<http://temple.edu/writingctr/english-language-learners/documents/Prepositionsvs.Prepositionalphraseshandout.pdf>

(last access on 7 August 2017)

instances where the phrasal verb is directly followed by NPs, such as *go on one's life/their education/a marriage*, etc. Some of these deviant usages show that the aimed verb was indeed a phrasal-prepositional verb – *go on with* – and the necessary preposition was missing (see sentence 52; the use of *go on* in TICLE will be discussed at more length in the following chapter).

(52) *Last of all, teenegars [sic] and married women have a chance to go on their ordinary lives.* <TRCU1100>

Bring about is another example of the parallel use of a one-word verb and a phrasal verb. However, in contrast to *go on*, *bring about* is underused by learners in comparison to the native students. With $p=0.0002$, this difference is highly significant.¹¹¹ As stated by Waibel (2007: 89), “[u]nderuse frequently goes together with avoidance strategies such as the increased use of another lexical item.” Thus, a possible one-word verb alternative to *bring about*, namely *cause*, was quantified in order to find out whether learners opted for it – taking into account that both verbs cannot be used interchangeably in all contexts.¹¹² The figures found are in Table 7.

¹¹¹ $\chi^2_{(1)} = 13.3704$, $df = 1$, $p = 0.0002556$

¹¹² This selection is based on Oxford Advanced Learner’s Dictionary of Current English, 6th ed.; *cause* is given as a synonym to *bring about*.

Table 7. *bring about* versus *cause*

	LOCNESS	TICLE
	n/million words (raw)	n/million words (raw)
<i>bring about</i>	87 (23)	15 (3)
<i>cause</i>	492 (130)	905 (182)

In statistical terms, based on chi-square test, Turkish learners preferred *cause* over *bring about*. With p-value < 2.2e-16, this distribution is highly significant. There is one other use of *bring about* in TICLE which is not included in the quantification here but under phrasal-prepositional verbs because it is erroneously followed by an additional preposition (see sentence 53). This additional use of a preposition can be directly attributable to the influence of L1 translation; one equivalent verb to *bring about* in Turkish (*beraberinde getirmek*) carries the notion of ‘accompanied by’ which literally equates to the preposition *with* in English:

(53) *The effects of the radiation **brings about with** some signs on the bodies of people and also it affects the next generations.*
<TRCU1133>

The three learners (TRCU1133, TRCU1167, TRME3019) who produced *bring about* in their essays – used only once by each learner – also produced *cause* – again only once. The learner TRKE2011, however, did not show any use of *cause*.

Although *cause* is opted for by the learners, quite a few of its uses show wrong complementation: *to-infinitive* directly follows the verb. Here are a few examples of the erroneous uses found in the learner data:

(54) *Decreasing the snakes **cause** to increase the number of mice. [...] In conclusion, we need to have scientific developments of course. but experts shouldn't **cause** to extinct the animals.* <TRKE2054>

(55) *First they did it by valuable things such as gold, but later payment **had caused** to invent money. [...] Power of the money **caused** to be exploited and to be suffered privation by making use of people by each other.* <TRCU1054>

(56) *As a result, this situation **causes** to be many jobless graduates not well educated.* <TRCU1070>

(57) *It **causes** to lose their good emotions such as love, mercy, affection, etc. towards each other [...]* <TRCU1121>

The second most frequently used phrasal verb by the learners is *sum up*. As it is commonly used as a text-structuring device by English learners, i.e. to introduce a summary, its frequent use was indeed expected on the basis of the literature (see Gilquin et al. 2007: IW28). While this phrasal verb was found to be used only once in the native speaker data, in total there are 40 instances of *sum up* in TICLE, corresponding to a relative frequency of 199 per million words – one instance of *summing up*, one instance of **in sum up* and 38 instances of *to sum up*. Overuse of this phrasal verb as a text connector has already been noted for other learner groups; for instance, for French-speaking learners of English (Gilquin 2015: 79), Korean-speaking learners of English (Sung 2017: 14), and

German- and Italian-speaking learners of English (Waibel 2007). This overuse may point to different preferences among the native and non-native speakers. A close look at the data reveals that while Turkish-speaking learners had a clear, statistically significant preference for formulations on the basis of *sum*¹¹³, native students opted for formulations with *conclude* and *conclusion* instead (see Waibel 2007: 90) – although they still used them less often compared to learners (see Table 8, adapted from Waibel (2007: 90), for the figures of *sum up*, complemented by the figures for possible alternatives).¹¹⁴ As for the combinations with *to conclude* and *conclusion* – *to conclude/in conclusion/to come to a conclusion/as a conclusion* – in total 323 (raw 65) occurrences were found in TICLE, which is significantly high compared to the native speakers' use of these phrases, which is 76 times per million words (see Waibel 2007: 90). The frequent use of text structuring devices by the learners may be due to acquired behavior from teaching, i.e. the way essay organization is taught to the learners. As rightly pointed out by Field and Yip (1992: 24), “[t]eacher and textbook influences may take many forms.” The phrases found in the essays based on the verbs *sum* and *conclude* are among the commonly taught ‘model’ formulations to signal conclusion in writing. The statement of Waibel (2007) for learners of English with German and Italian backgrounds in

¹¹³ For *sum up* *LOCNESS:TICLE* $\chi^2 = 37.0976$, $df = 1$, $p\text{-value} = 1.124e-09$

¹¹⁴ There are, of course, other structuring devices, i.e. concluding text connectors, used in the learner corpus, e.g. *at last*, *lastly*, *in short*, *finally*, *briefly*, etc. Here only a few are chosen to provide a rough picture of possible alternatives.

this context can also be affirmed for Turkish learners: The frequent use of these text-structuring devices may indicate that “learners are intent on making the macro-structure of their essays obvious” (p. 91), which may derive from their limited ability to write well.

Table 8. *sum up*, *summarize*, *in sum/summary* and some alternatives with *conclude*

	LOCNESS	TICLE
	pmw (raw)	pmw (raw)
<i>to sum up/summing up</i>	0	204 (39)
<i>to summarise/summarizing, as a summary*, in sum/in summary</i>	8 (2)	40 (7)
<i>sum</i> total	8 (2)	244 (46)
<i>to conclude/concluding, in conclusion, to come to a conclusion*, as a conclusion*</i>	76 (20)	323 (65)

A striking finding is that all usages of *sum up* and other formulations related to *sum* are used sentence-initially, some of which are used paragraph-initially. With the exception of one case, the same finding holds true for the phrases with *conclude* and *conclusion*.¹¹⁵ These results correlate closely with Field and Yip’s (1992) results; investigating the use of conjunctives in the English essay organization of native speakers and Cantonese learners of English, the researchers found that learners used a significantly higher number of organizational cohesive devices – additive, adversative, causal and temporal – than the native speakers and

¹¹⁵ What I wanna say **as a conclusion** is, both of the sexes are unique, [...] <TRCU1153>

they also tended to choose the initial sentence and paragraph position rather than placing the structural devices within the sentence. According to Field and Yip (1992: 26), the tendency of learners to use organizational devices in initial sentence or paragraph position may result from “a lack of complex sentence writing abilities among L2 writers or they may feel less competent to insert within a sentence.”¹¹⁶ Another plausible explanation for this tendency, already noted by Waibel (2007: 91) for German and Italian learners of English, is that by putting these organizational devices at the beginning of a sentence learners may be willing to make the structure of their text explicit for their reader so that s/he can better follow the line of argumentation. This can be the case especially if the texts are written under time pressure, as pointed out by Field and Yip (1992: 24) as follows:

[...] the high frequency of devices in L2 and even in L1 scripts may be due to the limited time provided for completion of the task. Content had to be devised quickly and writers may have relied on organizational devices to shape the essay rather than a strong development of their thought.

As the essays in TICLE were not written under time constraints, the former explanation is more likely for the learner group at hand.

¹¹⁶ Field and Yip (1992: 25f) note that “[i]f the effect of the NIP position [not in initial position] is to reduce the prominence of organizational devices and to convey a fluent and natural sounding speech tone, then a heavy use of ISP [initial sentence position] / IPP [initial paragraph position] where it is possible to place the device within sentence, can make the prose sound ponderous and stilted. [...] Being less influenced by native speaker intonational positions, they [the learners] may simply not see a need to vary position in this way.”

Before moving on to the next frequent phrasal verb, one other thing to note with regard to the concluding structuring devices is that there are some erroneous uses in the learner corpus: **in a conclusion* (1), **as conclusion* (1), **in sum up* (1); **to sum* (1) – the latter was not included in the counts of phrasal verbs.

The third most frequent phrasal verb in TICLE is *grow up* with 174 occurrences (raw 35), which is also one of the most frequent phrasal verbs in LOCNESS (rank 8 with 98 occurrences). *Bring up* is another highly used phrasal verb, both in the learner and native student corpora. In the OALD, *raise* (a child) is given as a synonym to *bring up*. Another main meaning of this phrasal verb is to ‘introduce or mention for attention, discussion, action, or consideration; start to talk about a particular subject’: *bring up* a subject. A semantic analysis of the data was performed in order to determine whether both of these major meanings were represented in the learner corpus, which is the case for the native speaker corpus: ‘raise a child’ = 34.5 percent, ‘start to talk about a particular subject’ = 65.5 percent (Waibel 2007: 91). Learners, on the other hand, except in two questionable cases (see sentences 58 and 59), used *bring up* in the sense of ‘raise a child’. This can be seen as evidence in support of the claims of restricted nature of learners’ phrasal verb knowledge (see Garnier and Schmitt 2016; also Waibel 2007: 93).

(58) *According to a traditional thought, men work out of the home and women work in their home. Their place is determined by society. The reason of this situation is the style of people’s **bringing up**.* <TRKE2034>

(59) *Second people are not **brought up** self-confidently during the university years.* <TRME3009>

Three essay topics – namely ‘abortion’, ‘sex equality’ and ‘divorce’ – were expected to have triggered the use of phrasal verbs like *grow up* and *bring up*. This possibility was, however, confirmed only for two essays for both phrasal verbs: ‘abortion’ and ‘sex equality’. *Grow up* was found to occur in essays with various topics – 11 different essay titles. Expressed in percentages, 31.43 percent of all instances of *grow up* (11 occurrences out of 35) and 21.74 percent of all instances of *bring up* (5 out of 23) occur in the essay entitled ‘abortion’. 25.71 percent of all occurrences of *grow up* (9 usages out of 35) and 47.83 percent of all occurrences of *bring up* (11 out of 23) were found in the essay entitled “sex equality”. When combined, 57.14 percent of all instances of *grow up* and 69.57 percent of all instances of *bring up* are concentrated on only two different types of essays in TICLE. Thus, it can be stated that the use of these two phrasal verbs is clearly topic-dependent.

One other thing noticed during the close semantic reading of the instances of *grow up* is the high number of deviant uses: In quite a few cases it is used as a transitive verb or in a wrong context, i.e. it was confused with *bring up*. Sentence (60) illustrates the confusion between the two verbs: *grow up* was erroneously used as a transitive verb by the learner:

(60) *Giving a birth to a child is not difficult; however, **growing** him or her **up** as a conscious, educated, healthy and beneficial individual is difficult. A family should have children as many*

*as they can be in charge of and they can afford to **grow up** perfectly.* <TRCU1094>

As will be elaborated in the next chapter, teaching and L1 influence are two possible factors causing confusion among the learner group under scrutiny in the use of these two verbs.

The one-word synonym of the phrasal verb *bring up* – i.e. *raise* – was attested ten times in the learner data in the meaning of ‘raising a child’. Three of these uses were erroneous due to an additional particle: **raise up (child(ren)/baby/honest people)*.

Coming back to the issue of topic-sensitivity, another clear case of it is *go out* – the fourth most common phrasal verb in the learner corpus. Although *go out* is also used frequently in the native corpus, the frequency difference between the two corpora is statistically significant (X-squared = 8.1, df = 1, p-value = 0.004427). In TICLE, more than half of its occurrences are concentrated on one essay topic: ‘sex equality.’ Expressed in percentages, 55.17 percent of all instances of *go out* were used in essays written on this specific topic.

As the sixth most frequent phrasal verb in the list comes *carry out*, which is the second most frequent phrasal verb in the native speaker corpus. *Carry out* and *point out* (12th most frequent verb in TICLE) were reported to be common in academic prose by Biber et al. (1999: 410). Although these two verbs are among the most frequent phrasal

verbs in the learner corpus, they are not used as frequently by the learners as by the native students (see Table 9). One other thing noticed is that all occurrences of *point out* in TICLE follow the VPO pattern.

Table 9. Frequencies of *carry out* and *point out*

	LOCNESS pmw (raw)	TICLE pmw (raw)
<i>carry out</i>	182 (48)	104 (21)
<i>point out</i>	155 (41)	30 (6)

A probable explanation for the differences seen between the learners and native speakers regarding the overall frequency of these two phrasal verbs is provided by Waibel (2007: 90) as follows:

British and American students are more aware of the fact that essays written in an academic context require a specific vocabulary; they may therefore use the same phrasal verbs as scientists do. At the same time, learners may not be aware of the stylistic implications of specific phrasal verbs, i.e. they do not differentiate between those that are very colloquial – and may thus not be ideal for essay writing – and others that are appropriate also in academic writing.¹¹⁷

¹¹⁷ Actually, some language usage observed in the investigated learner corpus indicates clearly that learners are not aware of register restrictions in academic prose; some speech-like features can be discerned in the data such as the use of contractions (*don't*), phrases involving personal pronouns and private verbs (*I think*), use of shortenings (*u*, *wanna*, *cause*), colloquial word choices, etc. (see Biber et al. 1999: 1041-1052). The following sentences drawn from the data may give a flavor of the colloquial style of the learner language under investigation:

(a) [...] and the woman has to **okay** the rules and restrictions designated by her familiar husband **I think** it is acceptable and [...] <TRKE2052>

As for the usage of *carry out* by the learners, in some contexts its use is questionable or ‘ambiguous’ whereas in others it is clearly deviant from native speaker norms of conventionality. Here are some illustrative examples:

(61) *I am not saying all of a sudden they become cured of such disease but perhaps less pain and problems, so they can **carry out** normal life.* <TRCU1176>

(62) *Again, another biggest problem of our universities is not achieving the cooperation between school and industry. This results in the way not **carrying out** the usage of their knowledge in real life. As it's known before, the education programmes completely based on theory, don't give permission for this. Many graduates have difficulties in **carrying out** the knowledge; they learnt it before as theoretical.* <TRCU1069>

(63) *They must not be allowed to be in the public like men because they could not **carry out** an occupation as men have done.* <TRKE2019>

(64) *One of its negative dimension is that it is very difficult and impossible to **carry out** equality between sexes in a nation like us which has a patriarchal national structure.* <TRKE2052>

Sentence (61) reveals that the deviation can be rectified by changing only the particle – as far as it can be understood from the context, the intended phrasal verb by the learner was *carry on*. The other uses of

(b) **I mean** this type of cheating or cheating attempt (**whatever**) refreshes the information that hardworking students have. <TRKE2057>

carry out (62-64) display unidiomatic usage; they are semantically inappropriate in the context they were used.

Although *give up* is reported by Biber et al. (1999: 410) to be common in conversation, fiction and news but not in academic prose, it is one of the most frequently used phrasal verbs both in the learner corpus and native control corpus. With 99 occurrences in per million words (pmw), it is the 7th most frequent phrasal verb in the learner corpus. There is a similarity in the degree of usage with the native speakers: with 95 occurrences per million words, its rank is ninth place in LOCNESS. Occurrences of *give up* are evenly distributed across different essays in the learner corpus: it is used in eleven different topics, i.e. no dependency on specific topics is discernible. In two occurrences of this phrasal verb, mother tongue influence can be seen very clearly: As the use of ablative suffix with the verb's object is necessary with the corresponding verb in Turkish, the preposition *from* – the equivalence of the ablative suffix in Turkish – follows the phrasal verb in the learner usage:

(65) *Although it has some advantages and disadvantages people are using and will be continue to use it because the advantages of it are superior to disadvantages so people can't **give up from** this.* <TRCU1016>

(66) *People are so addicted to money in recent days that they can do anything for it, they can **give up from** their relatives, from their friends, they can tell lie, they can steal ... etc. [...] Even it is good or bad, I think people will never **give up from** money, they'll always work for it.* <TRCU1044>

Another fairly colloquial phrasal verb found in the most frequent phrasal verb list of the learner corpus is *find out* (rank 10 with 30 occurrences). Its rank, with 72 occurrences, is eleven in the native speaker corpus. *Find out* is reported by Biber et al. (1999: 410) to be common mostly in conversation, less so in fiction and news; its use is also seen in academic writing yet compared to its frequency in other registers, it is not very common in academic writing. However, it should be stated that there is a statistically significant difference between the learner and the native speaker corpus despite their similar rankings ($X^2 = 6.76$, $df = 1$, $p\text{-value} = 0.009$). As there is an overall low frequency for phrasal verbs in TICLE, *find out* is in the list with raw six occurrences. In addition to its low frequency, half of its usages are erroneous (sentences 67-69). In sentence (67) and (68) the deviation could be rectified by omitting only the particle. However, in (69) *find out* was produced inappropriately instead of *discover*. Given the fact that *discover* is one of the one-word synonyms provided on ESL lists for the phrasal verb *find out*, this deviation can be foreseen. The learners under investigation either used *discover* (raw 18 occurrences) or *find out*; however, they were not used by the same learner interchangeably.

(67) *Are men and women equal? I think this one of the most difficult questions to which human being have not **found out** a satisfactory answer yet.* <TRCU1110>

(68) *So, when they do this experiments, they have to use microscopes; otherwise they could not **find out** that virus.* <TRCU1021>

(69) *He **found out** conditioned learning method.*
<TRKE2032>

Two uses of *built up*, which has four raw occurrences overall, were considered colloquial by the NS judges:

(70) As a first step, physical features **built up** differences between women and men. <TRKE2063>

(71) *To prison him to **build up** many handicaps in front of him in order not to unite with his readers and then after his death to declare him as a hero?* <TRCU1126>

Although used by the NSs, *build up* is not among the most commonly used phrasal verbs with eight raw occurrences corresponding to thirty per million words. The same statement also holds true for the other eight commonly used phrasal verbs by the learners: *come together*, *get on*, *wake up*, *come back*, *break out*, *break up*, *bring back*, and *come out* (for the individual frequencies of these verbs in LOCNESS see Appendix 1).

Turn out, which was reported by Biber et al. (1999: 410) to be most frequently used in fiction but also used in news and academic prose, was attested both in the learner and native speaker corpus: the 23rd most frequent verb in LOCNESS, 16th in TICLE. The other most commonly used phrasal verbs found in both corpora are *go back* and *take away*. The latter was reported to be one of the common activity phrasal verbs in LGSWE Corpus (Biber et al. 1999: 409). As for *go back*, its one-word equivalent *return* was checked in the learner data: Out of seven instances of the verb *return*, five were used in the meaning of *go back*.

The learners opted either for the multi-word verb or its one-word alternative, i.e. no learner used both verbs at the same time in their essay. One other verb relevant in this context is *come back*, which ranks at 18th position in TICLE – in three out of four attested instances, it was used in the meaning of ‘return’. Again, the learners who used this phrasal verb did not use the one-word verb equivalent.

There are three phrasal verbs used frequently by the learners but not attested in LOCNESS: these verbs are *come over*, *keep on* and *lock up*. *Come over* is reported to be quite frequently used in conversation and fiction by Biber et al. (1999: 410) and was attested four times in the learner data – used by three learners. However, as far as the context enables one to guess, all four usages actually targeted the single verb *overcome* – in three instances *come over* collocates with *problem(s)* (see sentences (166) to (169) in Section 6.2). As also noted by one of the NSs consulted, *come over* can only be used in the sense of physical action, as in *He is coming over to my place later*.

As for the unattested verb in the NS corpus *lock up*, it is fairly plausible to speculate that its use is topic-dependent.

Chapter 5: Multi-word verbs in learner writing – a quantitative approach

Table 10. The 25 most frequent phrasal verbs in LOCNESS along with their frequencies in TICLE

	Phrasal Verb	Frequency in LOCNESS pmw (raw)	Frequency in TICLE pmw (raw)
1	<i>go on</i>	201 (53)	284 (57)
2	<i>carry out</i>	182 (48)	104 (21)
3	<i>point out</i>	155 (41)	30 (6)
4	<i>take away</i>	117 (31)	30 (6)
5	<i>bring up</i>	110 (29)	114 (23)
6	<i>take on</i>	102 (27)	10 (2)
7	<i>end up</i>	98 (26)	15 (3)
8	<i>grow up</i>	98 (26)	174 (35)
9	<i>give up</i>	95 (25)	86 (17)
10	<i>bring about</i>	87 (23)	15 (3)
11	<i>find out</i>	72 (19)	30 (6)
12	<i>make up</i>	68 (18)	15 (3)
13	<i>set up</i>	64 (17)	5 (1)
14	<i>go back</i>	61 (16)	35 (7)
15	<i>break down</i>	53 (14)	0
16	<i>get away</i>	53 (14)	0
17	<i>cut off</i>	45 (12)	0
18	<i>be out</i>	45 (12)	5 (1)
19	<i>bring in</i>	42 (11)	0
20	<i>carry on</i>	42 (11)	5 (1)
21	<i>go out</i>	42 (11)	144 (29)
22	<i>run up</i>	42 (11)	0
23	<i>turn out</i>	42 (11)	25 (5)
24	<i>fit in</i>	38 (10)	5 (1)
25	<i>go on</i>	201 (53)	284 (57)

In Table 10, the 25 most frequent phrasal verbs found in LOCNESS are given along with their frequency in TICLE¹¹⁸. If the frequencies of these verbs are compared in the two corpora, it is seen that six of the frequent phrasal verbs used by the native speakers are not attested in the learner data: *break down*, *get away*, *cut off*, *bring in*, *run up* and *get out*. Four are attested but only once: *set up*, *be out*, *carry on*, and *fit in*. There are, however, four phrasal verbs in Table 10 that the learners use with higher frequency: *go on*, *bring up*, *grow up* and *go out*. Fluctuations of this kind probably depend more on the subject matter of the text than the productivity of the combinations. According to *Macmillan Phrasal Verb Plus* (Rundell 2005), most of the phrasal verbs used very frequently by the native speakers in comparison to learners belong to the group of most common and basic phrasal verbs – namely, *take away*, *take on*, *end up*, *make up*, *set up*, *break down*, *get away*, *cut off*, *bring in*, *carry on* and *get out*. All of these verbs are highly polysemous in nature and one may feel tempted to ascribe the underuse of such common and ‘basic’ phrasal verbs by learners to this feature. Though reasonable, this possibility remains speculative to date, lacking data to substantiate it. Searching for other possible explanations for the underuse observed in the learner data, I looked at the phrasal verb lists provided by Kartal (2018) for the four commonly used upper-intermediate level ELT textbooks in Turkey.¹¹⁹ The results have revealed that whereas

¹¹⁸ The frequency results for LOCNESS are taken from Waibel (2007: 87).

¹¹⁹ These coursebooks are New Headway, New English File, Face2Face and English Unlimited.

some of these common verbs are presented in some of the coursebooks (e.g. *end up* and *set up* appear in three of the books; *make up*, *take on*, and *carry on* in two; *get out* and *break down* in only one), some are not included at all (e.g., *cut off*, *bring in*, and *take away*) – the latter finding may be explained by the possibility of their being introduced in lower proficiency level textbooks at an earlier stage.

To round off this section, the analysis of the 25 most frequent phrasal verb types in the two corpora has revealed remarkable differences in the overall frequencies, along with some similarities of usage. The learner essays show not only a clear underuse of phrasal verbs, but they also contain a more restricted range of phrasal verbs, with the ten most frequently used phrasal verbs accounting for slightly more than half of the total, i.e. for 51 percent of the total items (the so-called Zipfian distribution). In fact, the top eight items constitute half of the total items. The ten most frequently used phrasal verbs in the native speaker corpus, however, constitutes only 24 percent of the total phrasal verb use.

The learners also make restricted use of some phrasal verbs, i.e. they do not apply them in a wide range of meaning senses as their native speaker counterparts (e.g., *bring up*, *go on*). Moreover, the uses of phrasal verbs by the learners occasionally deviate from native speaker norms. Some of these deviations can be accounted for by L1 interference but some cannot. There are only very few formal, ‘academic’, phrasal verbs attested in the learner data: *sum up*, *point out*, and *carry out*, occurring side by side with less formal phrasal verbs, such as *find out* and *turn out*. On the whole, however, the frequencies of most phrasal verb types

in the learner corpus are too low for meaningful conclusions to be drawn. One safe conclusion is that an important amount of the attested phrasal verbs show topic-dependency and they are not expected in a formal writing, such as *go out*, *wake up*, *wash up*, *bring up*, *grow up*, *lock up*, etc.

5.2.3 Frequency of individual phrasal-prepositional verbs

This section presents the overall quantitative results obtained for phrasal-prepositional verbs attested in TICLE. As mentioned in Section 5.2.1, TICLE yielded only a total of 35 types of phrasal-prepositional verbs, used a total number of 64 times. In Table 11, the most frequent eight phrasal-prepositional verb types occurring in TICLE are listed along with the frequency of the same verbs in LOCNESS. As can be seen, the overall frequency of these verbs is quite low, especially in the native speaker corpus. The other verbs attested in TICLE and not listed in the Table 11 were used only once by the learners (25 verb types; see Appendix 2 for these verbs).

Table 11. The most frequent phrasal-prepositional verbs in TICLE along with their frequencies in LOCNESS

	Phrasal-prepositional Verb	Frequency in TICLE pmw (raw)	Frequency in LOCNESS pmw (raw)
1	<i>*come across with</i>	45 (9)	0
2	<i>get on with</i>	30 (6)	8 (2)
3	<i>get along with</i>	25 (5)	0
4	<i>come up with</i>	20 (4)	20 (5)
5	<i>keep away from</i>	20 (4)	0
6	<i>*give up from</i>	15 (3)	0
7	<i>get back to</i>	10 (2)	0
8	<i>put up with</i>	10 (2)	4 (1)

The most frequently used phrasal-prepositional verb by the learners, namely *come across with*, was used erroneously in all cases. There are two entries for *come across* in Cowie and Mackin (1975). The first one is as an intransitive prepositional verb with the meaning of ‘find, meet (usually by accident)’ and the second is as an intransitive phrasal verb with the meaning sense of ‘be communicated, understood, heard’. There is one other entry for its intransitive phrasal verb use, which is informal, with the meaning of ‘give, hand over [money, information, keys]’, optionally followed by the preposition *with* turning it into a phrasal-prepositional verb, as in the following sentence (72) given as an example by Cowie and Mackin (1975):

(72) *The robber gave the old man’s arm a twist, and he at once **came across with** all the money he had in his pocket.*

In none of the nine attested uses of *come across with* in TICLE, the learners aimed for the meaning in sentence (72). According to NS judgments, the learners aimed for the prepositional verb *come across* in eight out of the nine cases and in one case the verb choice was inappropriate; it did not fit the context in which it was used (see sentence 73).

(73) *So, knowledge should not only be given theoretically, It should not be forgotten that no matter has [sic] much you read, study or memorize, you can not be successful when you **come across with** the application of what you do if you have never tried it.* <TRME3023>

One of the NSs approached suggested that the meaning the learner tried to express in sentence (73) could be achieved through ‘when you are faced with’ and ‘when you are tasked with’.

The second most frequent phrasal-prepositional verb in the TICLE list is *get on with*, used six times by five learners. This verb has two meanings: The first one is ‘have a harmonious relationship (with)’ – in this usage the multi-word verb can be used without the preposition; the second meaning is ‘continue to do [work, job, the gardening]’ (Cowie and Mackin 1975). The learners, however, used *get on with* only in the former meaning in five of the attested six cases. In contrast to learners, native speakers used *get on with* in the latter meaning, attested twice in LOCNESS. In one of its uses by the learners, *get on with* does not fit the context at all (see sentence 74). Possible alternative verbs provided by native speakers are *match with*, *fit with*. However, as one native

speaker commented, in order to convey the learner's intended meaning it would be better to rephrase the sentence to a greater extent.

(74) *Also the knowledge people took in universities do [sic] not **get on with** reality so we can name these [sic] knowledge as unnecessary.* <TRCU1128>

The next most frequently used verb on the list has the same meaning as *get on with*: *get along with*. It is used five times by three learners and there was no hit for it in LOCNESS. The use of these two synonymous phrasal-prepositional verbs in TICLE shows no dependency on specific essay topics; they are distributed across five different topics.

Come up with features as the fourth most common phrasal-prepositional verb in TICLE – used four times by four learners, five times by the native speakers, corresponding to a relative frequency of 20 per million words in both corpora. In the meaning of 'to produce, find', it has been considered an informal verb (Cowie & Mackin 1975). In three cases, this is the meaning intended by the learners. In the fourth one, this meaning does not fit the context (see sentence 75). Intended meanings with this verb in this context could be 'come up against', 'deal with' or 'endure'.

(75) *Men can do everything, but women can't. [...] Only the educated society, educated women and man prevent this situation. [...] But, the most important duty is the women's duty, they should know their place in the life, and they should **come up with** each matter.* <TRKE2034>

The same learner produced this multi-word verb for the very same meaning erroneously in an earlier part of her essay; she used the particle and preposition in a reversed order:

(76) *While they [women] are walking around the street, or while driving a car they are felt it, with some rude words and other similar bad behaviours. In order to prevent from it women try to be strong but it is really very hard to **come with up**.*
<TRKE2034>

Keep away from is the fifth verb on the list, used four times by three learners and not at all by the NSs.

The next verb in the list is another case of an inappropriate use – including a wrong choice of preposition: **give up from*. It was produced three times by two learners. This multi-word verb with the correct choice of preposition (*give up on*) was also attested in another learner's essay, used only once. The substitution of the preposition in this case can be accounted for by L1 interference. The Turkish equivalents of this phrasal-prepositional verb *birinden ümidi(ni)/umudu(nu) kesmek* ('give up on someone; despair of someone') and *birşeyden vazgeçmek/umudu(nu) kesmek* ('give up on/quit something') necessitate the use of the ablative suffix, which is comparable to the preposition *from* in English. *Give up on* was attested only once in LOCNESS.

The seventh verb on the list with only two raw tokens is *get back to*, which was not attested in the NS data. This verb has two different meanings, one of which has been reported to be informal: 'to phone, write,

or speak to someone at a later time’ (Rundell 2005: 173). Yet the two learners used *get back to* in the meaning of ‘return to’. Having a closer look at these two learners’ essays to see whether they also used the single verb *return* in their essays yielded negative results. That is, the learners did not use the single verb and their multi-word verb equivalent interchangeably.

The last verb in Table 11 is *put up with* – attested twice in TICLE and only once in LOCNESS. This verb is commonly given in ESL phrasal verb lists with its one-word ‘equivalent’: *tolerate*. The two learners who used *put up with* did not use its one-word equivalent in their essays. *Tolerate* was utilised in five essays, with a total number of six times. LOCNESS also yielded the same overall frequency for this multi-word verb.

As it should be clear from the foregoing account, phrasal-prepositional verbs were very rare both in the learner and NS data investigated. Given their general rarity compared to phrasal verbs and the fact that they are mostly considered informal, their overall low frequency was thus not entirely unexpected. Despite their low number, quite a few of the attested usages in the learner data were used in inappropriate contexts. Other uses showed deviations from native speaker norm, which can be directly attributable to the influence of L1 translation equivalents.

5.2.4 Productivity of verbs and particles in forming phrasal and phrasal-prepositional verbs

One other aspect to consider in the context of phrasal and phrasal-prepositional verbs is the productivity of their constituents. Following Biber et al. (1999), phrasal and phrasal-prepositional verbs were divided into their parts to investigate their productivity since information gleaned from such an analysis can be very revealing with regard to the differences and similarities between groups as to which verbs are selected most frequently as the basis for forming phrasal and phrasal-prepositional verbs and which particles are most productive. Productivity here refers to the potential of lexical verbs to combine with different particles to form phrasal and phrasal-prepositional verbs. Similarly, in the case of particles, it refers to the potential of a given particle to combine with different verb types. As Waibel (2007: 96) put it, “a small number of verbs may, in combination with different particles, generate a large number of different phrasal verbs; the same may apply to a small number of particles”.

The analysis has revealed that the learners used 83 verb types, which were combined with 23 different particles to form phrasal and phrasal-prepositional verbs. Native speakers used, as reported by Waibel (2007: 96), 222 different verb types, combined with 24 particles. Despite the similarity in the overall number of particle types, native speakers made use of almost three times more verb types as the verbal basis in forming phrasal and phrasal-prepositional verbs compared to the learner group. Considering the overall lower number of phrasal verbs in the learner

corpus, the remarkable difference in the overall frequency of verb types in these constructions between the two groups was thus not entirely unexpected.

57 out of 83 verb types in the learner corpus were used only once as the verbal basis in the formation of phrasal and phrasal-prepositional verbs, 11 only twice. These verbs in the learner corpus have been considered non-productive in forming verb-particle combinations. Only the verbs that combine with at least three different particles qualified as a productive verb in the present study. Therefore, the list of productive verbs in the formation of phrasal and phrasal-prepositional verbs in TICLE contains only 15 verbs whereas the number is as high as 36 in the productive verb list of LOCNESS. The entire list of all productive verbs forming these two types of multi-word verbs in the two corpora can be found in Appendix 3.

Despite differences in frequency, there is considerable similarity in the verbal items used by the learners and native speaker students, especially with regard to the ten most productive verbs: The ten most productive verbs used in the formation of phrasal and phrasal-prepositional verbs in TICLE are *go*, *come*, *take*, *get*, *bring*, *give*, *put*, *fall*, *stand*, and *be*, in decreasing order (see Table 12). Seven of these verbs are also among the ten most productive verbs in the formation of phrasal and phrasal-prepositional verbs in LOCNESS – namely, *go*, *come*, *get*, *bring*, *put*, *take* and *turn*. In both corpora *go* is the most productive verb, followed by *come*. There are two verbs functioning as a productive verb in verb-

particle combinations in TICLE but not in LOCNESS, namely, *fall* and *leave* – they were used in combination with four and three different particles by the learners, respectively. However, it should be noted that despite this similarity between the two groups with regard to the most productive verb types, the potential of the individual verbs for forming phrasal and phrasal-prepositional verbs in the two corpora vary to a great extent; verbs were combined with a higher range of particles by the native speakers. For instance, whereas *turn* was combined only with three different particles by the learners to form phrasal verbs, it was used in combination with nine different particles by native speakers. The exact same goes for the verb *be* (see Appendix 3). The graph below (Figure 3) illustrates this divergence in the two corpora for the most productive ten verbs in TICLE. That is, the productive verb list from TICLE functions as the basis in the graph, followed by figures found for these verbs in LOCNESS.

As mentioned in Section 3.2, the verbs that are commonly found in verb-particle combinations are mostly monosyllabic and of Germanic origin. All of the productive verbs attested in these combinations in TICLE are monosyllabic verbs, the most productive ten monosyllabic verbs of Germanic origin. Table 12 below shows the entire list of productive verbs found in the learner corpus along with the adverbial particles they co-occur to generate phrasal and phrasal-prepositional verbs.

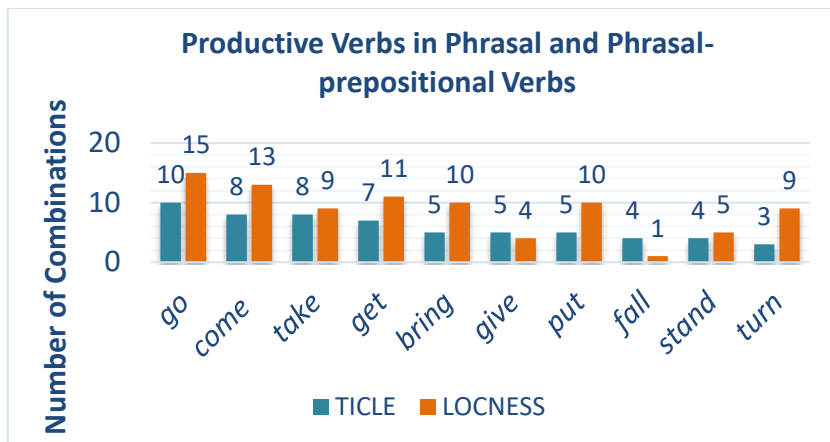


Figure 3: Number of different particle combinations with ten productive verbs in TICLE, along with the corresponding figures from LOCNESS

Table 12. The productive verbs in TICLE and the particles they combine with

1	go	+	<i>away, back, by, down, forward, on, out, round, up, together</i>
2	come	+	<i>across, along, back, out, over, together, up, with</i>
3	take	+	<i>away, back, in, off, on, out, over, up</i>
4	get	+	<i>along, away, back, down, on, up, together</i>
5	bring	+	<i>about, back, on, together, up</i>
6	give	+	<i>back, off, out, over, up</i>
7	put	+	<i>down, forward, on, out, up</i>
8	fall	+	<i>apart, away, down, ou</i>
9	stand	+	<i>by, on, together, up</i>
10	be	+	<i>apart, off, out</i>
11	keep	+	<i>away, on, together</i>
12	leave	+	<i>aside, behind, out</i>
13	pass	+	<i>by, on, over</i>
14	run	+	<i>across, around, aw</i>
15	turn	+	<i>back, on, out</i>

The above findings attained from the analysis of the productivity of verbs in TICLE tally well with Biber et al.'s findings (1999: 413): According to their results, the high-frequency verbs *take*, *get*, *put*, *come*, *go*, *turn*, *bring* and the verb *set* are particularly productive in forming phrasal verbs.¹²⁰ With the exception of the verb *set*, which is used twice and with two different particles in TICLE (*set forth* and *set up*), these high frequency verbs are also productive in the learner corpus in forming phrasal verb constructions. We see three other productive verbs in the learner corpus, which are not given in Biber et al.'s list: *give*, *fall* and *stand*. However, as mentioned earlier, verbs are considered productive in this study if they combine with at least three different particles. Thus, although *give*, *fall*, and *stand* are considered productive verbs, they are not used in as many different combinations as the other verbs in the list, *give* occurring with five different particles, *fall* and *stand* with only four.

Having the potential to combine with a high number of different particles does not necessarily mean that these verbs are also the most frequently used ones. Reconsidering the data for learners in Table 12, this statement is corroborated: *Sum up*, for instance, features as the second most frequent phrasal verb used by the learners although the verb *sum* is not very productive itself in the formation of phrasal verbs. The same goes for the verbs *grow*, *carry*, *point* – they function as the basis for the

¹²⁰ It is important to note that Biber et al. (1999) did not include phrasal verbs on the basis of main verb use of *be*, *do*, and *have* whereas *be* was included in this study.

most frequently used phrasal verbs in both the learner and the reference corpora.

It is informative to see that there are no major differences between the learners and native speaker students in the overall choice of verbal basis in the formation of phrasal and phrasal-prepositional verbs. However, native speaker students not only use a wider range of verbs in the formation of these multi-word verbs, they also use each of these verbs with a higher number of particles, i.e. more productively.

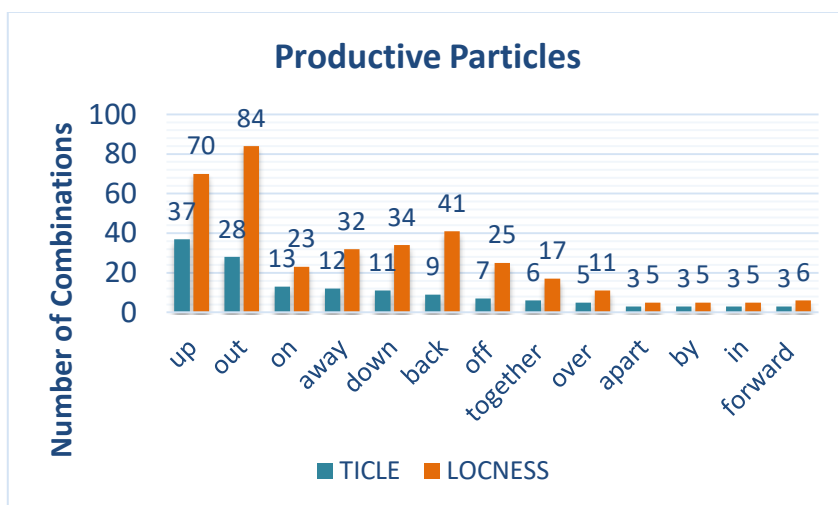


Figure 4. Number of verb types combined with the productive particles in TICLE, along with the corresponding figures from LOCNESS

As for the productivity of particles, as stated earlier in this section, both groups used about the same number of different particles – learners used 23 and native students 24. Two particles occur only in the native

speaker corpus (*ahead, under*) and one only in TICLE (*round*). Seven out of the 23 particles in the learner corpus were combined only with one verb, whereas only two particles in the native corpus show such low productivity: *under* and *with*. While the former was not used by the learners at all, the latter appeared only once and erroneously in a phrasal-prepositional construction: **come with up*. Figure 4 demonstrates the productive particles in the learner data (13 in total); the figures illustrating the total number of different verbs combined with each particle for the learner data are complemented by the corresponding figures from LOCNESS. Out of the 13 productive particles listed in the graph, ten also feature among the 12 most productive particles in the native speaker corpus. Instead of the particles *apart* and *by*, we see *in* and *around* in the list of the top 12 productive particles of LOCNESS. The remaining ten particles attested in the learner corpus lacks productivity; they are combined with either one or two verbs. However, the details on them and the other particles seen in the native speaker corpus but not listed in Figure 4 can be found in Appendix 4a. Overall, it can be stated that the productivity of particles in TICLE, in accordance with the general low frequency of phrasal and phrasal-prepositional verbs reported earlier in this section, is significantly lower compared to the native speaker group. The number of verb types that combine with an individual particle varies to a significant degree between the two groups. Particles are combined with a far smaller number of verbs by the learners than the native speakers, as can be seen in Figure 4 and in the Appendix 4a. For instance, even the most productive particles *up*

and *out* are used in less than half as many varied combinations by the learners compared to their native speaker counterparts.

An interesting finding that should be highlighted is the striking difference between particles with regard to their productivity in forming phrasal and phrasal-prepositional verbs. There are particles in the investigated corpora that are used in combinations with as many different verbs as 84, along with particles that occur in combination with only one single verb. The most frequent particle used by the learners is *up* with 188 instances – it was combined with 37 different verbs. This finding was not unexpected since the productivity of the particle *up* has long been noted in the literature (e.g., Meyer 1975; Bolinger 1971; Biber et al. 1999), as mentioned earlier in Section 3.2. *Up* is followed by *out* in the learner corpus with 103 instances, used in combination with 28 different verbs. These findings accord precisely with Biber et al.'s findings: They have shown that *up* is the most productive particle in the formation of phrasal verbs, followed immediately by *out* (1999: 412f). The analysis of the productivity of these two particles in the native speaker data, however, has revealed a slight deviation from Biber et al.'s findings: In LOCNESS the most productive particle is *out*, followed by *up* (Waibel 2007: 100). The importance of these two particles in forming phrasal and phrasal-prepositional verbs is verified, especially for the learners at hand, when the figures obtained for them are added and related to the total number of phrasal verb types: *Up* and *out* constitute 71 percent of phrasal verb types in TICLE. To put it differ-

ently, 60 out of the 83 verb types used in phrasal and phrasal-prepositional verb constructions in the learner corpus have been found to combine with either one or the other of these two particles. With regard to the overall frequency, more than half of the phrasal and phrasal prepositional verbs in the learner data were formed with these two particles: 296 out of 522.

On is the next productive particle in the learner data, which has also been listed by Biber et al. (1999: 413) as one of the particularly productive particles. It ranks as the third most productive particle in TICLE, combining with 13 different verbs. It is the eighth most productive particle in LOCNESS, occurring with 23 verb types. Its overall frequency in both corpora is close to each other, unlike in the case of the first two particles in the list; 443 (pmw) occurrences of *on* in these constructions in TICLE, 508 (pmw) occurrences in LOCNESS.

The fourth productive particle in TICLE – *away* – has a similar productivity degree as *on*. It was attested in 25 multi-word verb constructions, in combination with 12 different verbs. This particle features quite prominently in LOCNESS; it ranks as the fifth most productive particle, occurring with 32 verb types in the formation of phrasal verbs. Thus, it has nearly three times more range of use in LOCNESS than in TICLE.

The fifth most productive particle is *down* in TICLE – one of the particles reported to be productive in the formation of phrasal verbs by Biber et al. (1999: 413). In TICLE, it was attested with 11 different verbs for

a total number of 17 times. This particle is another very prolific particle in native speaker essays; although its rank is the same as in TICLE, it was used with 34 different verb types in LOCNESS – its range is three times wider than found in learner data.

Further productive adverbial particles listed by Biber et al. are *in* and *off*. Whereas *in* is among the ten most productive particles in the native speaker corpus, occurring in combination with 26 different verbs, it is not very productive in the learner corpus, as seen in Figure 4 above: It is combined only with three different verbs – each used only once. The productivity of the particle *off* is almost as high as the particle *in* in the native speaker corpus. Although its rank is the same in the both corpora, its productivity varies to a large extent between the two groups: whereas it was combined with 25 different verbs to form phrasal verbs by the native speaker students, learners used it in combination with seven verbs – each verb and particle combination was used only once.

The remaining particles used in both corpora such as *back*, *over* and *together*, which are especially productive in the native speaker corpus compared to the learner corpus, have not been listed in LGSWE by Biber et al. (1999). This is most probably because Biber et al. (1999) listed only the most frequent phrasal verbs and particles attested in the LGSWE. Meyer (1975: 5), however, mentions these remaining particles as productive particles in the list he provided. According to Meyer, the number of the most productive adverbial particles (what he calls “second elements”) used to form phrasal verbs – with his term “two-

word verbs” – are seventeen: *about, across, along, around, aside, away, back, by, down, in, off, on, out, over, through, under, up*. Out of these particles, *back* and *away* are particularly prolific and overly frequent in the control corpus compared to the learner corpus (see Appendix 4a).

As for the frequency of particles, a common-sense assumption would be that the most productive particles in the data should also be the ones used most frequently. This assumption was justified to a large extent for both corpora although there are some slight differences in the orders of productivity and frequency of individual particles. *Up, out, and on* are the three most frequent particles in TICLE as well as in LOCNESS – the only difference is the reverse order of the first two particles in the two corpora. *Up* and *out*, two particles which have already been noted as the two most frequent adverbial particles in American and in British English (see Waibel 2007: 102), are not only the most productive particles in forming phrasal and phrasal-prepositional verbs in the corpora investigated, but they are also the most frequently used ones by both learners and native speaker students. Table 13 shows the frequencies of the most frequent six particles in the two corpora – given in both absolute and relative frequencies per million words (pmw). The order in the table is based on the frequency of the particles in TICLE. The frequencies of the remaining particles can be found in Appendix 4b. As can be seen in the Table 13, the most frequently used particles are the same in the two corpora but despite this similarity, there is a remarkable difference between the figures for learners and those for native students,

which is, as already pointed out above, result from the general low frequency of phrasal and phrasal-prepositional verbs in TICLE.

Table 13. The frequencies of the six most frequent particles pmw (raw)

	Particle	Frequency in TICLE pmw (raw)	Frequency in LOCNESS pmw (raw)
1	<i>up</i>	935 (188)	1163 (308)
2	<i>out</i>	512 (103)	1201 (318)
3	<i>on</i>	443 (89)	508 (135)
4	<i>away</i>	124 (25)	409 (108)
5	<i>back</i>	124 (25)	405 (107)
6	<i>down</i>	85 (17)	333 (88)

The significance of these six particles in forming phrasal verbs in LOCNESS was noted by Waibel (2007). She found that 77.3 percent of all phrasal verbs were made up of a verb plus one of these six particles (p. 102). Following Waibel, the frequencies of these six particles were added and the sum figure attained was then related to the overall amount of phrasal and phrasal-prepositional verbs in TICLE. The results have revealed that Turkish learners relied more on these six particles than native speaker students. While in LOCNESS almost a quarter of all phrasal and phrasal-prepositional verbs consist of a verb plus an adverbial particle other than *out*, *up*, *on*, *back*, *away*, and *down*, this percentage is even lower in TICLE – only 14.3 percent of all phrasal and phrasal-prepositional verbs extracted from the learner corpus are not made up of a verb plus either *out*, *up*, *on*, *back*, *away*, or *down*. The particles *up*, *out* and *on* are particularly significant in the learner corpus

– almost 73 percent of all phrasal and phrasal-prepositional verbs consist of one of these three particles. So, it can be concluded that there is a high degree of reliance on a few productive particles in forming phrasal and phrasal-prepositional verbs by the learners.

To round off this section, native speaker students and the learners exhibit considerable similarities of usage with regard to the most productive and frequent verb and particle types. However, there is a remarkable difference between the two groups with regard to the overall frequency of verbs and particles. Learners differ significantly from their native speaker counterparts in relying on a more limited range of verbs in the formation of phrasal and phrasal-prepositional verbs; native speakers use three times more verb types than the learners. The learners also show a higher degree of reliance on a limited number of productive particles in constructing these two multi-word verb categories.

5.2.5 Extralinguistic factors correlating with the quantitative use of phrasal and phrasal-prepositional verbs

In TICLE, overall 458 phrasal and 64 phrasal-prepositional verbs were used in a total of 278 essays. Assuming that all learners in the data behaved in the same way, i.e. we are dealing with an ‘average learner’ (cf. Waibel 2007: 103), we would expect each learner to use 1.8 phrasal or phrasal-prepositional verbs in their essay. In reality, however, the analysis revealed that 69 learners did not use any phrasal or phrasal-prepositional verb at all. In other words, 25 percent of the learners show a

deviation from the average amount of the expected phrasal verb use. This variation in the quantitative use of phrasal and phrasal-prepositional verbs between the learners raises the question of whether any of the independent variables and/or their correlation could explain the results obtained, and if so, to what extent.

As mentioned in Section 4.2, for the level assignment in the ICLE project, external criteria, i.e. non-linguistic factors, were opted for: Learners were considered to be advanced because of their institutional status. However, despite the similar (linguistic) backgrounds, the learners in the data unavoidably demonstrate different levels of attainment in their target language. There are potentially influential variables accounting for the differences in the proficiency level attained by each learner – individual differences such as motivation, learning styles and age of first exposure to the TL as well as extralinguistic factors, such as quality and quantity of input, length of exposure to the TL.

In addition to the individual and extralinguistic factors determining the ultimate success attained in learning an L2, there are potentially influential task-related factors that may account for the differences in the quality of the essays to some extent, such as access to reference tools and time-constraints imposed on the learners during the writing task, i.e. whether essays are timed or not. In the current study interaction between variables in various ways is expected. Both the effects of possibly influential individual variables recorded on the ICLE-CD and the cir-

cumstances of production are to be investigated (see section 4.2). Although it is a fairly homogeneous learner group, there are some extralinguistic factors that may account for the variation found in the data with regard to the number of phrasal and phrasal-prepositional verbs used. One such factor is L2 exposure – length of stay in an English-speaking country: it is commonly assumed that there is a positive correlation between a stay in the country where the target language is spoken and the learner’s foreign language proficiency (cf. Masangya and Lozada 2009), regardless of the purpose behind the stay (see Waibel 2007: 104). However, mixed results have been forthcoming for this common-sense assumption, as noted in Chapter 2. Investigating the effect of the extent of L2 exposure on the choice of multi-word verbs rather than their one-word equivalents by advanced learners of English whose L1 is outside the Germanic group of languages, Siyanova and Schmitt (2007) did not find a discernible effect. According to their results, “even over a year’s worth of native-environment exposure was insufficient to increase the likelihood of using the multi-word verbs” (Siyanova & Schmitt 2007: 132). Waibel (2007) was also interested in the question of whether the amount of time spent in an English-speaking country has an influence on the frequency of phrasal verb use by advanced learners of English but her subjects had German and Italian as their L1. In contrast to the findings obtained by Siyanova and Schmitt (2007), Waibel’s results showed a positive correlation for both learner groups between the amount of time spent abroad and the quantity of phrasal verbs used, i.e., about 20 percent of the learners used more phrasal verbs after having

spent time in an English-speaking country. These contrary results could have to do with the typological differences in the L1 of the subjects investigated in the two studies. As Siyanova and Schmitt (2007: 132) have speculated, learners with non-Germanic L1s may need a longer time to become comfortable with “alien” multi-word verbs. It should also be mentioned that despite the positive correlation found in Waibel’s study, the researcher noted that this correlation was only related to the quantity of the phrasal verbs – not the quality; there were many erroneous uses of phrasal verbs. Therefore, the duration of L2 exposure, Waibel concluded, may be a determining factor for an increase in the phrasal verb use for a fifth of all essays but this correlation does not reveal anything about the quality of phrasal verb use (2007: 109). These arguments and the idiomatic nature of multi-word verbs taken together suggest that L2 exposure is considered to be an influential factor in the use of phrasal and phrasal-prepositional verbs for the learner group at hand with a non-Germanic L1, but unfortunately, as mentioned in Chapter 4, there are only four learners in the data who spent time in an English speaking country, making the investigation of possible impact of this factor unfeasible.

One related variable to the L2 exposure that may contribute to the learners’ proficiency in general is the length of learning the target language. The common-sense assumption related to this variable is the longer, the better although this is usually not that straightforward given the fact that the quality and extent of instruction available to the learner plays a determining role in the proficiency level attained in foreign languages (see

Section 1.4). Dealing with the very same lexical aspect of learner language but of German and Italian learners of English instead of Turkish, Waibel (2007) combined the years of teaching at school and the years at university, which are recorded separately on the ICLE-CD, in order to investigate whether the learners who had learnt English for a longer period used more phrasal verbs in their essays. Her findings for the two groups differed: Whereas no correlation between the duration of learning the target language and the frequency of phrasal verbs emerged for the Italian learners, learners in the German corpus who had learnt English for a longer period of time used significantly more phrasal verbs than the ones who had a shorter period of learning English but the difference was slight (Waibel 2007: 114f). Waibel's contradictory findings for these two learner groups made this variable more interesting to investigate further for the learner group at hand.

The variable of time pressure, recorded on the ICLE-CD and shown to be an influential variable in lexical studies (see e.g., Nessalhauf 2005: 230; Waibel 2007: 113), is unfortunately not relevant for the data at hand since all essays in the learner corpus were written untimed (see Section 4.2). That means the influence of time constraints imposed on the learners on the use of phrasal and phrasal-prepositional verbs cannot be dealt with in the present study. However, one other influential variable in lexical studies is whether learners are given vocabulary support, i.e. whether use of reference tools such as mono- or bilingual dictionaries are allowed during the writing task. This variable is very crucial for the current study since more than half of the students (145 students)

were allowed to have access to reference tools during the writing task. Stated more precisely, 133 essays (47.8%) were written without the help of a reference tool and 145 essays (52.1%) were written using reference tools. This variable was found to be positively correlated with the quantitative use of phrasal verbs for German learners (Waibel 2007: 113f). The same effect was found in the use of collocations for the same group of learners by Nesselhauf (2005: 231), i.e., slightly more collocations were used by German learners working with a dictionary. A similar effect is expected in the present study.

Lastly, one other possibly influential variable relevant in the present study is text-length. As writing in a second language is “more constrained, more difficult, and less effective” than writing in a first language (Silva 1993: 668) – mostly because the foreign language learner needs to devote his or her cognitive resources not only to the content but also to issues of language – the ability of writing longer texts may be reasonably considered as an indicator of higher competence in a foreign language. And higher proficiency in a foreign language has been reported in various studies to be positively correlated with familiarity and ease in using multi-word verbs (see Chapter 2). Furthermore, it is usually assumed that the longer a text is, the more likely it is that a certain feature or item occurs in a text. Based on these assumptions, the variable of text length is believed to exert influence on the quantitative use of phrasal and phrasal-prepositional verbs in the present study. Investigating this very research question in the context of phrasal verbs,

Waibel (2007), in contrast to expectations, found no significant correlation between text length and the number of phrasal verbs used in the learners' texts. That means, according to her results, that it is possible for a short essay to include more phrasal verbs than a longer essay (Waibel 2007: 110).

Rather than looking individually at all of the potentially influential independent variables, a multivariate analysis (namely, binary logistic and multiple linear regression analyse) was conducted in the present study since some overlap and interaction between independent variables could be expected. As Gries (2015: 175) rightly claimed, “[n]othing in linguistics is truly monocausal.”

Out of the 15 learner-dependent variables recorded on the ICLE-CD, only eight were included in the analysis. The excluded variables are *native language*, *other foreign language 2*, *other foreign language 3*, *language at home 1*, *language at home 2*, *language at home 3* and *length of stay in an English-speaking country*, i.e. months abroad. The decision in favor of their exclusion was taken mainly on the grounds that these excluded variables bear little significance if viewed at the level of all learners under scrutiny. For instance, whereas only one learner had knowledge of a 3rd foreign language, Turkish is the NL of 98 percent of the learners – i.e., 276 out of 280 learners. The remaining four learners did not specify which language their NL is. As mentioned earlier, other languages spoken at home were not specified, either (see Section 4.2. for the details of the other excluded variables). Therefore,

their inclusion in the analysis would not be of any help in the interpretation of the results. Besides, although it is considered to be an influential factor, the variable *length of stay in an English-speaking country* was also excluded from the analysis because there were only four learners who have been in an English-speaking country. Its inclusion would distort the analysis.

The independent variables included in the analyses are:

- *gender* (0=male / 1=female)
- *use of reference tools* (0=no / 1=yes)
- *other foreign language 1* (2=other / 3=German / 4=French)
- *essay topic* (1-18)
- *age* (19-38 years)
- *text length* (500 – 1390 tokens)
- *years of English at school* (4-13 years)
- *years of English at university* (3-6 years)

The first four of the variables are nominal; the remaining four are metric in nature.

Firstly, a binary logistic regression analysis was run to determine whether the use of phrasal and phrasal-prepositional verbs or lack thereof could be predicted from the individual characteristics of the learners. Binary logistic regression was chosen as the most appropriate statistical test as the dependent variable, use of phrasal/phrasal-prepositional verbs, was categorical and binary. To this end, the independent

variables which were considered to be potentially influential were included in the analysis in order to ascertain their effect on the likelihood of phrasal and phrasal-prepositional verb use (yes or no). Although the binary logistic regression model was statistically significant $\chi^2(1)=9,379, p <0.002$, the model explains only 4.9 percent (Nagelkerke R^2) of the variance in the phrasal and phrasal-prepositional verb use. It correctly classified 74.8 percent of the cases. Out of the eight selected predictor variables, only one variable, namely *text length*, reached a level of significance ($p =.003$; see Table 14). That is, the variable *text length* was found to be the only significant variable in determining whether a learner would make use of a phrasal or phrasal-prepositional verb in their essays. In other words, an increase in text length is associated with an increased likelihood of phrasal or phrasal-prepositional verb use in the present study.

Table 14. The impact of the independent variable text length on the use of phrasal and phrasal-prepositional verb use

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1	Text length	,003	,001	8,722	1	,003	1,003
	Constant	-,971	,696	1,948	1	,163	,379

a. Variable(s) entered on step 1: text length.

Since the remaining seven variables did not contribute to the model (i.e., they did not attain a level of significance), they were not selected by the model. Regarding the only significant predictor variable, i.e. *text*

length, it is worth noting that its factor is 1.003. That is, when all other independent variables were kept constant, for each extra word the probability of a learner's use of a phrasal or phrasal-prepositional verb would increase by a factor of 1.003. To put it differently, for every one hundred more words in a text, we expect to see 1.349 (1.003^{100}) more phrasal or phrasal-prepositional verb.

The next question was to find out which learner variables would be selected when the question was how many phrasal and/or phrasal-prepositional verbs a learner would make use of in his/her text. That is, the aim was to see the impact of the independent variables on the frequency of phrasal and phrasal-prepositional verb use. To this end, a multiple linear regression analysis was performed, using the frequency of phrasal and phrasal-prepositional verb use as the criterion variable. As one can only insert binary or metric independent variables in this type of regression analysis, only the variables *text length*, *gender*, *age*, *use of reference tools*, *total years of English* (combination of the variables *years of English at school* and *years of English at the university*) were included in the analysis – *essay topic* had to be excluded as it is a non-binary categorical variable. The results of this multiple linear regression analysis indicated that the model could explain 8.7 percent of the variability (adjusted R^2 of .087) seen in the data with regard to the overall frequency of phrasal and phrasal-prepositional verb use by each learner ($F(1, 276) = 27.489, p = .000$). In accordance with the results of the binary logistic regression, among the predictor variables entered into

5.2 Phrasal and phrasal-prepositional verbs in TICLE

the model, the only significant predictor variable of the overall frequency of phrasal and phrasal-prepositional verb use was found to be *text-length*, as can be seen in Table 15. The probability of these verbs in an essay increases by .004 for each additional word.

Table 15. Results of linear regression analysis (stepwise) for the overall phrasal and phrasal-prepositional verb frequency for each learner.

predictor variable	B	S.E.	standardized coefficient Beta	T	Sig.
total years of English study	,051	,050	,063	1,027	,305
gender	,269	,284	,055	,949	,343
age	-,094	,069	-,081	-1,354	,177
use of reference tools	-,315	,225	-,082	-1,402	,162
text length	,004	,001	,302	5,216	,000

After the relative frequency of phrasal and phrasal-prepositional verb use per essay was calculated ($M = .0025$, $SD = .0025$) and the variable *text length* was taken out, another multiple linear regression analysis was performed with the independent variables *use of reference tools*, *age*, *gender* and *total years of English*. The results revealed no change in the initial results. None of the predictor variables entered into the model reached a level of significance. That is, none of these variables could explain the variability in the data with regard to the relative frequency of phrasal and phrasal-prepositional verb use by each learner

Chapter 5: Multi-word verbs in learner writing – a quantitative approach

($F(4,273) = 1.047, p = .383$, with an R^2 of .15). The full results of this regression model are displayed in Table 16.

Table 16. Results of linear regression analysis (enter method) for the relative phrasal and phrasal-prepositional verb frequency for each learner.

predictor variable	B	S.E.	standardized coefficient Beta	T	Sig.
total years of English study	5,999E-5	,000	,056	,879	,380
gender	,000	,000	,056	,927	,355
age	-,000	,000	-,090	-1,429	,154
use of reference tools	-,000	,000	-,073	-1,195	,233

The only significant factor – *text length* – deserves some further consideration here. The essay tokens in TICLE range from 500 to 1390, with most essays ranging between 500 and 900 words. Overall, however, the essays cover a wide range of essay lengths, as can be seen in the histogram below (Figure 5). As mentioned earlier in this section with regard to this variable, the common-sense assumption is that the longer a text, the more likely it is that a certain feature occurs more frequently. Moreover, text length is believed to show a parallel increase with increasing proficiency in an L2, which is characterized by an ease and greater familiarity with the use of multi-word verbs. Based on these

assumptions, it was reasoned that the longer the text, the more occurrences of phrasal and/or phrasal-prepositional verbs should be expected.

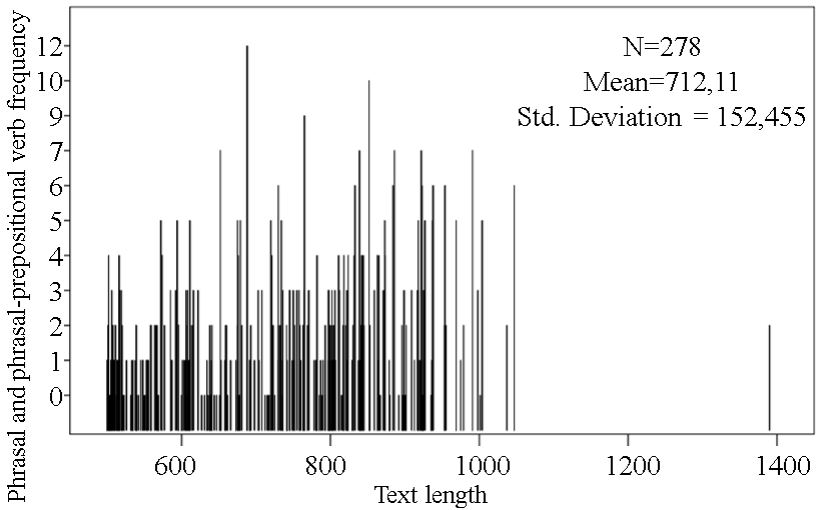


Figure 5. Text length and frequency of phrasal and phrasal-prepositional use in TICLE

The trend line and the corresponding confidence intervals in Figure 6 below indicate that such a prediction is valid for the current dataset yet it should be noted that there is only a very weak correlation: R^2 is 0.089 – although it is statistically significant ($p=0.000$). This means that *text length* is a relevant factor for the quantitative use of phrasal and phrasal-

prepositional verbs. This finding is in line with our predictions and refutes earlier findings reported by Waibel (2007).

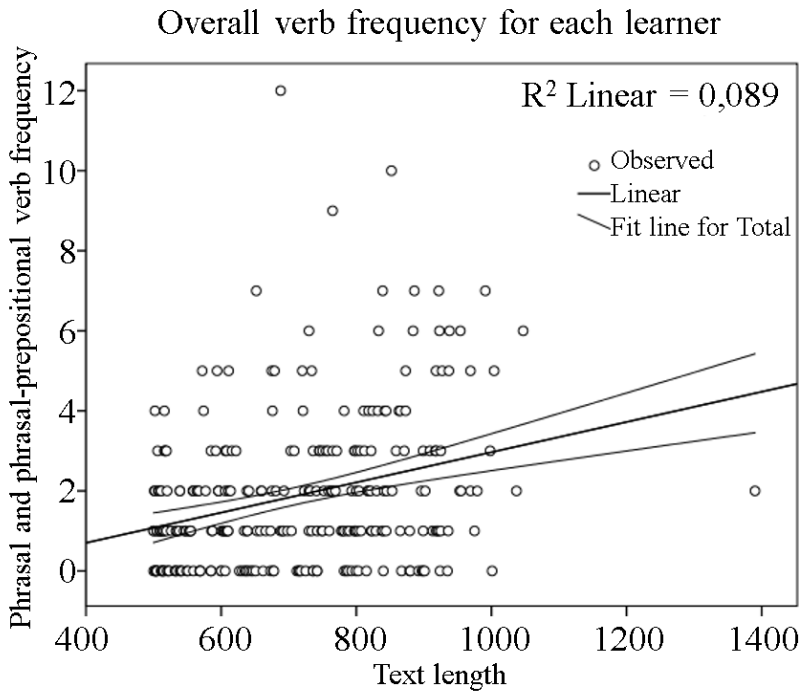


Figure 6. Correlation between text length and frequency of phrasal and phrasal-prepositional verb use in TICLE

Before we move on to the variable *essay topic*, variables that were not selected in the regression models deserve a brief discussion. Possibly the most unexpected variable in this respect is *the use of reference tools*. As mentioned earlier, this variable has been reported to have positive

correlation with the frequency of multi-word verbs in studies conducted by, for instance, Waibel (2007) and Nesselhauf (2009). However, the results of regression analyses along with the results of Mann Whitney U test, which was conducted as a preliminary test, revealed no impact of reference tools pertaining to the frequency of phrasal and phrasal-prepositional verb use for the dataset at hand ($U = 9,568.000$, $p = .909$). That is, it is not possible to state that those who used reference tools ($M = 1.7586$, $SD = 1.60414$) outperformed those who did not use reference tools with regard to the overall number of phrasal and phrasal-prepositional verbs ($M = 2.0075$, $SD = 2.19675$). The reason for this result is not entirely clear. One can only speculate that perhaps the insufficient or inappropriate information provided in the bilingual dictionaries or the ways learners make use of dictionaries account for this result.

The variable of *total years of English study* – the combination of the variables *years of English learning spent at school* and *years of English learning spent at the university* – is another variable expected to exert influence on the use of multi-word verbs. As mentioned in Section 4.2, the former variable is the variable that differs most between the participants; it ranges between 4 and 13 years. The years spent at the university studying English range between 3 and 6 years. The mean of *total years of English study* is 12.4 ($SD = 2.350$). One may feel tempted to relate the increasing years of studying a foreign language with increased proficiency, and in turn with confidence and ease in using phrasal and phrasal-prepositional verbs – for the acquisition of which a great deal of exposure is believed to be necessary (see Section 2.2). However, a

closer look at the data reveals that there are learners who studied English for 17 or 18 years but did not use a single phrasal or phrasal-prepositional verb in their essays (e.g., TRKE2009, TRKE2033, TRME3011), along with the learners who studied English for a shorter time but made use of quite a few such verbs in their essays (e.g., the learners TRKE2068 and TRCU1003, who studied English for 11 years, used 7 phrasal/phrasal-prepositional verbs). Therefore, without knowing anything about the learners' language learning skills, their motivation, their attitude towards the TL and many other potentially influential informant-dependent factors, it is not possible to deduce much from the years of learning English as a variable alone. Moreover, as it has already been shown by, for instance, Waibel (2007), quantitative use of phrasal or phrasal-prepositional verbs does not automatically equate with their correct use by the learner.

The fact that the independent variables *age* and *gender* were not selected in the regression models was less of a surprise. The irrelevance of age to the use of phrasal and phrasal-prepositional verbs may be explained by the fact that the learners' age in the data are in a close range – with the exception of a few outliers, most of them are in their early twenties. As for the variable *gender*, tentatively no impact was expected. No study has, to my knowledge, indicated an effect of this variable on the production of phrasal or phrasal-prepositional verbs; although Kamarudin (2013: 111-119) has shown that gender has an impact on the overall understanding of phrasal verbs, it is only to a small extent. Besides, Kamarudin's result was based on the comparison of means,

showing a monofactorial approach. As mentioned earlier, this approach, though common in LCR studies, is “improverished” (Gries 2015: 175), since it is very unlikely that the differences observed are indeed caused by only one variable – *gender*, in this context.

In the following, the variable *essay topic* will be dealt with. On the basis of observations made during the semantic reading of data, further analysis of this variable seemed necessary. As stated earlier in this section, the binary logistic regression revealed that essay topic was not a significant variable in predicting whether a learner would make use of phrasal and/or phrasal-prepositional verbs. However, this variable is likely to exert influence on the relative frequency of phrasal and phrasal-prepositional verb use for each learner. Since it was neither binary nor metric variable, this variable was not included in the multiple linear regression analysis. Each learner in the data wrote on only one topic, making the variable *essay topic* a speaker-related factor. Some topics were more popular in the data, i.e., they were chosen by more learners. For instance, whereas topics 3 and 10 (‘Sex equality’ and ‘Most university degrees are theoretical and do not prepare students for the real world. They are therefore of very little value’, respectively) were chosen by 57 learners, accounting for 41 percent of all the essays in the data, some topics were chosen by only one learner – namely, topics 5, 6, and 8 (see Table 5 for the list of all essay topics). Indeed, nine out of the eighteen topics were chosen by less than ten learners.

Chapter 5: Multi-word verbs in learner writing – a quantitative approach

Table 17. Descriptive results of the one-way ANOVA test (dependent variable relative frequency – topic ($n > 9$))

Topic	Mean	N	Std. Deviation
3	,0028991	57	,00248921
4	,0027039	17	,00271244
7	,0026854	23	,00252192
9	,0013241	30	,00188222
10	,0019753	57	,00185221
11	,0026126	15	,00213896
12	,0026241	21	,00236453
14	,0049313	12	,00494650
17	,0020657	11	,00153193
Total	,0024752	243	,00249719

In order to find out whether the essay topic is a relevant factor for the relative frequency of phrasal and phrasal-prepositional verb use for each learner, a one-way ANOVA test was performed on the topics which were chosen by at least ten students – nine topics in total (which makes up the 87.4 percent of all essays in the data). There was a statistically significant difference between essay topics in terms of the relative frequency of phrasal and phrasal-prepositional verb use as determined by one-way ANOVA ($F(8,234) = 3.018, p = .003$). That is, some essay topics triggered more use of phrasal and phrasal-prepositional verbs. Full results are to be seen in Table 17. However, a note of caution should be added here to state that according to the results of Levene statistics the homogeneity of variance assumption was not met since the

significance value based on a comparison of medians was greater than .05, suggesting a real difference between topics. Therefore, the ANOVA test result cannot be considered to be robust.

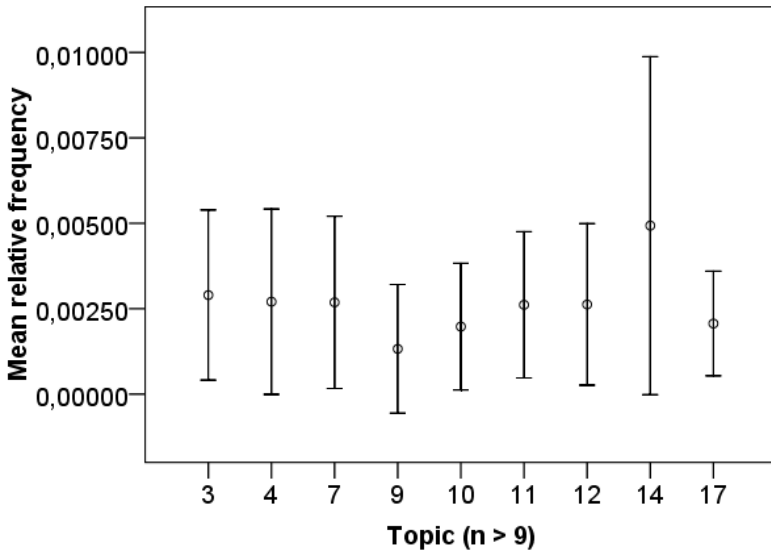


Figure 7. Correlation between topic and frequency of phrasal and phrasal-prepositional verb use in TICLE

As can be seen in the Table 17 as well as in the Figure 7, the mean number of phrasal and phrasal-prepositional verbs used by each learner writing about these topics varies between two to three – with one exception (Topic 14) – while each topic showed different levels of variation between learners writing about that topic. The mean number of phrasal and phrasal-prepositional verbs used by each learner writing on topic 14 (‘Divorce’), which was chosen by 12 learners, is the highest

(nearly five multi-word verbs for each learner) but the variation between the learners writing on this topic is also the highest. That is, whereas some learners made use of high numbers of phrasal and phrasal-prepositional verbs in their essays (as high as 12 or 9 in this context), some writing on the same topic used fewer or none at all. The least variation between the learners' behaviors with regard to the quantitative use of phrasal and phrasal-prepositional verbs is seen among those who wrote on topic 17 ('Animal testing'). Topic 3 and topic 10 were the most popular topics in the data and they were chosen by the same number of learners. Whereas the former triggered more use of phrasal and phrasal-prepositional verbs compared to the latter, i.e., nearly one more verb by each learner and 122 phrasal and phrasal-prepositional verbs in total, there is more variation between the learners who wrote on this topic compared to the latter topic. A closer look at the data shows that out of 57 learners who wrote on topic 3, 12 produced no phrasal or phrasal-prepositional verb in their essays whereas there are learners who produced seven or more verbs in one essay. Topic 9 ('In the words of the old song'), on the other hand, which was chosen by nearly half as many learners and triggered less than half as many phrasal and phrasal-prepositional verbs compared to the topics 3 and 10, shows less variation among learners. The mean number of phrasal and phrasal-prepositional verbs triggered by topic 4 ('Cheating in colleges') and topic 7 ('Great inventions') is very similar to that of topic 3 and the variation between learners is also to a similar extent. With regard to the topics 11 ('Abortion') and 12 ('Euthanasia'), their mean is similar to

the topics 3, 4 and 7, but there is less variation between learners. Taken together, these results mean that although there is some variation between learners writing on the same topic, some essay topics triggered more use of phrasal and phrasal-prepositional verbs. As will be tackled in Section 6.2, a small number of phrasal verbs triggered by a topic (e.g., *grow up* and *bring up*) are recurrently used by some learners.¹²¹

To round off this section, the regression analyses have unexpectedly revealed that the use of reference tools and the length of studying English did not have an impact on the quantitative use of phrasal and phrasal-prepositional verbs. Nor did age or gender. Out of the selected predictor variables, *text length* was found to be the only significant variable in predicting whether a learner would make use of these verbs at all. By means of the regression model formulated, we could explain only 9 percent of the variability observed in the data with regard to the quantitative use of phrasal and phrasal-prepositional verbs. A one-way ANOVA test has shown that *essay topic* is a relevant factor for the relative frequency of phrasal and phrasal-prepositional verbs. In other words, some essay topics triggered more use of phrasal and phrasal-prepositional verbs, with some variation among learners writing on the same topic.

¹²¹ As an aside, one interesting observation related to the predictor *essay topic* is its correlation with the variable *gender*. Some topics were chosen almost exclusively by female learners, for instance the topics 11 ('Abortion') and 14 ('Divorce').

The fact that 91 percent of the variability observed in the data could not be explained by means of the investigated variables suggests that there must be other factors at play that are more relevant to the learners' use of phrasal and/or phrasal-prepositional verbs. If other factors are at play, what might they be? Clearly, the results here support Möller's (2017) claim about the necessity of more extensive data on individual differences, i.e. learner characteristics, especially of the cognitive and affective type such as motivation and anxiety, and of the need of identifying additional relevant variables in order to explain variability in learner corpus data.

5.2.6 Synopsis: The quantitative use of phrasal and phrasal-prepositional verbs

The quantitative analysis of phrasal and phrasal-prepositional verbs in learner writing yielded some informative results. Although the analysis of the most frequent phrasal and phrasal-prepositional verb types in the two corpora has revealed some similarities of usage between learners and their native speaker counterparts, there are remarkable differences in the overall frequencies. In addition to a clear underuse, learners' use of phrasal verbs displayed a Zipfian distribution: The most frequent eight phrasal verb types account for half of the total phrasal verb tokens in the learner corpus. Moreover, the learners make a more restricted use of verb types, i.e. they do not apply them in a wide range of meaning senses as their native speaker counterparts (eg. *bring up*, *go on*). Quite

a few of the attested phrasal and phrasal-prepositional verbs in the learner data were used in inappropriate contexts or they were used idiosyncratically. Some of these deviations from the native speaker norms can be directly attributed to L1 interference. Only very few formal, ‘academic’, phrasal verbs were attested in the learner data, occurring side by side with less formal phrasal verbs, which might be seen as an indication of the learners’ lack of full awareness of stylistic connotations of specific phrasal verbs. On the whole, however, the frequencies of most phrasal verb types in the learner corpus are too low for meaningful conclusions to be drawn. One safe conclusion that can be drawn is that a considerable amount of the attested phrasal verbs show topic-dependency and they are not expected in a formal writing – just as some of the attested phrasal-prepositional verbs.

The use of phrasal and phrasal-prepositional verbs by the learners also differ significantly from that of their native speaker counterparts in one other respect: The learners rely on a more limited range of verb and particle types in the formation of these verbs.

The results of the multiple regression analyses performed to help predict Turkish learners’ potential use of phrasal and phrasal prepositional verbs yielded positive results with moderate correlation among the variables tested, explaining only 9 percent of the total variation in the data set and only one variable has been found to be significant – namely, *text length*. An increase in text length has been found to positively correlate

with the frequency of phrasal and phrasal-prepositional verbs in learners' essays. The use of reference tools during the writing task and the length of studying English exert no influence on the use of phrasal and phrasal-prepositional verbs. Nor does age or gender. According to the results of a one-way ANOVA test, *essay topic* is a relevant factor for the relative frequency of phrasal and phrasal-prepositional verbs in the learners' essays. That is, some essay topics triggered more use of phrasal and phrasal-prepositional verbs. It should be noted that one possibly influential factor – time spent in the target language environment (the variable *months abroad*) – had to be excluded from the analyses due to the very low number of learners who have been abroad.

5.3 Prepositional verbs in TICLE

As mentioned in Section 4.5.1, the extraction of prepositional verbs was based on the verb list of TICLE and was done manually so as to make sure that all prepositions, including the non-spatial ones, are covered in the study. That is, all concordance lines for the entire verb list were reviewed manually in order to not exclude any prepositional verb. Thus, no prelist of prepositions was used. By means of a careful manual filtering process, all instances were eliminated in which the prepositions did not occur within the verb phrase.

The extracted prepositional verbs were quantified independently of inherent semantic differences because there were cases in which the meaning of a prepositional verb in the data deviated from the mean-

ing(s) given in the dictionaries and where a clear-cut semantic differentiation was not possible. Moreover, in some cases learners used prepositional verbs in wrong contexts, assigning them inappropriate meanings in a given context.

In what follows the overall quantitative results of prepositional verb use by Turkish learners is presented first, followed by the aspects of frequency and productivity of the individual constituents of the prepositional verbs attested in TICLE.

5.3.1 Overall results regarding prepositional verb usage

In TICLE 443 prepositional verb types were found to be used a total number of 2,226 times (The entire list of prepositional verb types can be found in Appendix 5). As anticipated, prepositional verbs were used more frequently by the learners compared to phrasal verbs; nearly five times more prepositional verbs were attested. This finding is consistent with previous results (Biber et al. 1999). Accordingly, the number of verb types used in the formation of prepositional verbs is greater. However, although the list of verb types is longer, the degree of repetition is higher compared to phrasal verbs – nearly twice as much. Moreover, the readiness to produce prepositional verbs does not necessarily imply correct and appropriate usage. Quite a few prepositional verbs attested in TICLE deviate from native speaker norms, as can be seen in Appendix 5.

The traces of L1 patterns were expected to be found in the learners' use of prepositions in various ways, partly as a result of their non-existence in L1 but mostly because of the prepositions' very nature – their arbitrariness, polysemy, multifunctionality and semantic complexity. The qualitative analysis of all the prepositional verb tokens extracted from TICLE – the topic of the next chapter – clearly confirms this expectation. As will be discussed at length in the next chapter, a large part of prepositional verb tokens were used erroneously – expectedly learners made wrong choices of prepositions (e.g. *impose to* N for *impose on* N) or verbs (e.g. *adopt to* instead of *adapt to*), used additional prepositions (e.g. **look after to* N, **face with* N), omitted prepositions in some cases where they were necessary (e.g. *consist N* for *consist of* N) and used some ('unrecorded') prepositional verbs unidiomatically in a given context (e.g. *final in* N, *look forward* N).

200 out of 443 prepositional verb types were attested only once in the learner data, viz. nearly 9 percent of the total prepositional verb occurrences in the learner data are hapax legomena. The top ten prepositional verbs constitute nearly 30 percent of the total prepositional verb use.

296 different verb lemmas were used in forming prepositional verbs, combining with 27 different prepositions. Neither the distribution of the base verbs nor the distribution of prepositions in the formation of prepositional verbs was balanced. The most frequent ten verbs used as a base verb in forming prepositional verbs constitute 37 percent of the total prepositional verb tokens. The most frequent five prepositions

used in combination with verbs to form prepositional verbs constitute slightly more than 67 percent of the total prepositional use in the learner data.

5.3.2 Frequency of individual prepositional verbs

After having extracted all prepositional verbs from TICLE, a list of the individual frequency counts was prepared. While going through all the extracted prepositional verbs and commenting on each of them would take us too far afield, in what follows I will highlight a few interesting findings, focusing on the most frequent 25 prepositional verbs – which are listed in Table 18. All the figures given in the table are raw numbers. A detailed interpretation of the qualitative use of prepositional verbs will be provided in the following chapter but there are a number of observations that can be made from Table 18, which I will merely mention briefly here.

By and large, the list of the most frequent prepositional verbs of the learner corpus is predominantly comprised of common prepositional verbs – 15 of the prepositional verbs in the list were reported to be common prepositional verbs by Biber et al. (1999: 416ff). Some of the unexpected prepositional verbs in the list show a clear indication of topic dependency (e.g. *graduate from*, *study for*, *learn about*). Three verbs in the list – namely, *look*, *prepare* and *think* – occur with multiple prepositions.

Table 18. The 25 most frequent prepositional verbs in TICLE

Rank	Prepositional Verb	Raw frequency
1	<i>go to</i>	115
2	<i>prepare for</i>	90
3	<i>look at</i>	72
4	<i>see as</i>	62
5	<i>think of</i>	56
6	<i>think about</i>	55
7	<i>depend on</i>	54
8	<i>look after</i>	50
9	<i>graduate from</i>	48
10	<i>talk about</i>	44
11	<i>deal with</i>	40
12	<i>suffer from</i>	32
13	<i>prepare to*</i>	30
14	<i>apply to</i>	26
15	<i>base on</i>	26
16	<i>look for</i>	22
17	<i>mention about</i>	21
18	<i>speak about</i>	20
19	<i>be against</i>	19
20	<i>benefit from</i>	19
21	<i>agree with</i>	17
22	<i>learn about</i>	17
23	<i>wait for</i>	17
24	<i>come from</i>	16
25	<i>expose to</i>	16

The 25 most frequently used prepositional verbs constitute nearly 45 percent of the total prepositional verb tokens. Table 18 reveals substantial differences in the overall frequencies of these most frequently occurring items. It is headed by *go to* in the meaning of ‘visit, attend’ as in *go to university/college*, with 115 instances. The list ends with *expose to* with 16 instances – used seven times less compared to the most frequent prepositional verb *go on*.

The second most frequent prepositional verb *prepare for* is contained in one of the essay titles given to the learners under scrutiny – namely, topic 10 (‘Most university degrees are theoretical and do not **prepare** students **for** the real world. They are therefore of very little value’). This explains the high frequency attained for this prepositional verb in TICLE (this prepositional verb is not listed as a common one in LGSWE (Biber et al. 1999: 416ff)). In the most frequent prepositional verb list of TICLE, the verb *prepare* is also seen in combination with another preposition: **prepare to*. This erroneous combination was performed by 16 learners and a total number of 30 times, resulting in its rank as the 14th most frequent prepositional verb in the data. This deviation is interesting given the observation that the essay title these learners wrote their essays on contains this very verb with the correct prepositional use – *prepare for* – as well as that the right and wrong choices of preposition with this verb occur side by side in a single essay. This indicates that *prepare for* had not been fully acquired by these 16 learners; their use was influenced by the essay title (see Lorenz 1999: 13, fn. 22).

The third most frequently used prepositional verb by the learners is *look at*. This verb's extreme frequency in conversation and fiction has already been reported by Biber et al. (1999: 416): in LGSWE, it was found to be the single most common prepositional verb overall, occurring in all registers investigated – conversation, fiction, news and academic prose. In TICLE, the verb *look* was also frequently combined with the prepositions *after* and *for* although their frequencies differ to a large extent – the former preposition was combined with *look* 50 times whereas the latter occurred only 22 times. Both prepositional verbs were reported to be common in LGSWE and *look for* was found to be a lot more common, especially in fiction (Biber et al. 1999: 416ff).

Another verb in the list that combines with multiple prepositions to form prepositional verbs is *think*: *think of* and *think about* – interchangeable prepositional verbs according to OALD (online). These verbs feature as the fifth and the sixth most frequent prepositional verbs in the list with nearly the same amount of uses – there is merely one single more use of *think of*. These two mental prepositional verbs were listed by Biber et al. (1999: 417) to be very common – the difference between them lies in the register distribution: Whereas *think of* was found to occur in all four registers (conversation, fiction, news and academic prose), predominantly, however, in fiction, *think about* was not attested in academic prose and its overall frequency was low compared to that of *think of* (ibid.).

There are two other synonymous prepositional verbs in the most frequent prepositional verb list of TICLE: *talk about* and *speak about* – the former featuring as the 10th most frequent prepositional verb with 44 instances, the latter as the 18th with 20 instances. Biber et al. (1999: 417) attested no use of *talk about* in academic prose although in the other three registers investigated it was found to be commonly used, especially in fiction. The verb *speak* was also attested in TICLE in combination with another preposition for the same meaning – ‘mention or describe something’: *speak of*. Native speaker counterparts of the learners, however, produced *speak about* only three times yet *speak of* 13 times in their essays. The preference of the native speakers for *speak of* over *speak about* is supported by Biber et al.’s findings.

As the fourth most frequent prepositional verb in the list comes *see as* with 62 occurrences. This prepositional verb can be considered a formal one since it was reported by Biber et al. (1999: 417) to be common in news and academic prose – especially in the latter. The other two prepositional verbs on the list – *apply to* and *base on* – display a similar register distributional profile in LGSWE. They were attested 26 times in TICLE, ranking as the 14th and 15th prepositional verb in the top 25 prepositional verbs list. *Depend on* and *deal with* are the other ‘formal’ prepositional verbs in the list, reported to be common mostly in academic prose (Biber et al. 1999: 426 and 418); the former’s rank is 7 with 54 occurrences and the latter’s is 11 with 40 occurrences. Other ‘formal’ prepositional verbs were attested in TICLE but their frequencies are not high, e.g. *define as* and *account for*, used eight times and

once, respectively (see Appendix 5 for the entire list of prepositional verbs attested in TICLE).

Apart from the aforementioned erroneous combination **prepare to*, one non-standard prepositional verb is seen in the list: *mention about*. It is used 21 times by 18 learners. It is ranked at 17th position. Although it is considered non-standard, the existence of this prepositional verb, as mentioned in Section 6.2, has already been reported in various ESL varieties as well as in ICE-GB (Nesselhauf 2009).

To conclude this section, the analysis of individual prepositional verbs has revealed remarkable differences in the overall frequencies. The 25 most frequently used prepositional verbs comprise nearly 45 percent of the total prepositional verb use. Compared to phrasal verbs, the learners' use of prepositional verbs displays a higher range of verbs. The results attained for the most frequent prepositional verbs are in good agreement with the previous findings: 15 of the 25 most frequent prepositional verbs in TICLE were reported to be frequent in LGSWE (Biber et al. 1999), most of which are expected to be seen in a formal writing.

As anticipated, the uses of prepositional verbs by the learners often-times deviate from native speaker norms (see Appendix 5). The list of the most frequent prepositional verbs includes one erroneous combination (**prepare to*) and one non-standard combination (*mention about*). The number of deviations from native speaker norms in the formation of prepositional verbs is, however, quite high in the data. More details

on these deviations and their possible sources will be given in the next chapter. Topic-dependency as well as essay title are two determinant factors in the use of some prepositional verb types.

The next section is devoted to the frequencies of the constituents of prepositional verbs, i.e. their productivity.

5.3.3 Productivity of verbs and prepositions in forming prepositional verbs

Prepositional verbs were divided into their parts to investigate their productivity. As noted in Section 5.2.4, by productivity what is meant here is the potential of lexical verbs to combine with different prepositions to form prepositional verbs; in the case of prepositions, the term refers to the potential of a given preposition to combine with different verbs.

The analysis has revealed that the learners used 295 verb types, which were combined with 27 different prepositions to form prepositional verbs.

115 out of 296 verb types in the learner corpus were used only once as the verbal basis in the construction of prepositional verbs, 39 only twice. These verbs have been considered non-productive in the formation of prepositional verbs. Only the verbs that combine with at least three different prepositions qualified as a productive verb in the present study. The list of productive verbs seen in prepositional verb constructions in TICLE contains 28 verbs. These productive verbs along with

the prepositions they were combined with are listed in Table 19. 17 of these productive verbs were combined with three prepositions, as can be seen in Table 19. The overall frequency of these 17 verbs ranges between 4 and 34. The most productive verb in the prepositional verb constructions in the learner corpus is *go*, which is the second most frequent verbal basis in prepositional verbs, with 130 occurrences. It is used with ten different prepositions. The most frequent verb in prepositional verb constructions, *look* is the second most productive verb with eight different prepositions, used a total number of 150 times. As the next productive verbs in the list we see *live* and *care* with five different prepositions although it should be noted that these two verbs are not as frequent as the earlier ones – *live* was attested in 20 prepositional verbs whereas *care* only in 18. The next seven verbs on the list (namely, *talk*, *come*, *speak*, *spend*, *agree*, *be* and *enter*) occur with four prepositions in the data. Despite the fact that they were attested with the same number of prepositions, their overall frequencies vary to a great extent: *talk* has the highest frequency with a total number of 64 occurrences and *enter* has the least with 9 occurrences.

These findings differ to some extent from those of Biber et al. (1999: 422). They also found some lexical verbs that combine with multiple prepositions to form prepositional verbs (namely, *get*, *look*, *work*, *go*, *know*, *hear* and *use*) but noted at the same time that none of these verbs was particularly productive – contrary to the findings attained for phrasal verbs. Only three of the seven verbs in their data were found to

be productive in TICLE: *get*, *look* and *go*. Considering the higher number of productive verbs in TICLE, it should be noted that, as can be seen in Table 19, some of the prepositional verbs formed by the learners are erroneous or non-standard (e.g., *mention about N*, *look forward N*, *face with N*).

The most productive verbs used for the formation of prepositional verbs differ considerably from those used for phrasal verbs. To put it more precisely, only five verbs overlap on the two lists: namely, *go*, *come*, *get*, *be* and *turn*. Compared to phrasal verbs, a large number of verb types were used to form prepositional verbs (83 vs. 290) and the list of productive verbs used in forming prepositional verbs is longer. However, only few verb types are particularly frequent and productive. In contrast, a greater number of the verb types used in forming phrasal verbs display a higher level of productivity. Indeed, out of the 28 productive verbs used to form prepositional verbs, only four were combined with five or more prepositions whereas this number is seven for phrasal verbs; that is, seven verb types were combined with five or more particles to generate phrasal verbs. These findings are consistent with results of Biber et al. (1999: 422).

Chapter 5: Multi-word verbs in learner writing – a quantitative approach

Table 19. The productive verbs in TICLE and the prepositions they combine with

1	go	+	<i>after, around, beyond, by, for, in, into, through, to, with</i>
2	look	+	<i>around, after, at, for, *forward, over, through, *to</i>
3	live	+	<i>in, for, on, through, under</i>
4	care	+	<i>about, for, *of, *on, *with</i>
5	talk	+	<i>about, for, to, wit</i>
6	come	+	<i>across, *into, from, to</i>
7	speak	+	<i>about, of, on, to</i>
8	spend	+	<i>*to, for, on, with</i>
9	agree	+	<i>on, to, upon, with</i>
10	be	+	<i>above, after, against, for</i>
11	enter	+	<i>in, *to, *into, for</i>
12	suffer	+	<i>through, for, from</i>
13	face	+	<i>*with, *to, *of</i>
14	struggle	+	<i>against, for, with</i>
15	decide	+	<i>about, for, on</i>
16	result	+	<i>from, in, with</i>
17	ask	+	<i>about, for, *to</i>
18	fight	+	<i>against, for, with</i>
19	turn	+	<i>into, to, around</i>
20	call	+	<i>as, for, *with</i>
21	direct	+	<i>at, to, towards</i>
22	divide	+	<i>among, between, into</i>
23	get	+	<i>*into, over, to</i>
24	argue	+	<i>about, for, on</i>
25	trust	+	<i>in, *on, to</i>
26	transfer	+	<i>from, into, to</i>
27	die	+	<i>for, from, *through</i>
28	warn	+	<i>about, against, of</i>

As for the productivity of prepositions, as stated earlier, learners used 27 different prepositions. 7 out of the 27 prepositions in the learner corpus were combined only with one verb, 2 with 2 verb types. These prepositions are considered to lack productivity in this study. The productive preposition list of TICLE, therefore, consists of 18 prepositions. Table 20 demonstrates these prepositions (see Appendix 7 for the entire list of prepositions used to form prepositional verbs). The extent the prepositions differ from one another with regard to their productivity in forming prepositional verbs is noteworthy. There are prepositions in the data that were used in combination with as many different verbs as 95, along with other prepositions that were attested in combination with only one single verb (see Appendix 7). In line with Biber et al.'s (1999: 423) findings, the most productive preposition used by the learners is *to*; it was combined with 95 different verbs a total number of 493 times. *To* is followed by *with*, which was used in combination with 68 different verbs in 251 instances.

From is the next productive particle in the learner data, which has not been listed by Biber et al. (1999: 423) as a productive particle. It ranks as the third most productive particle in TICLE, combining with 46 different verbs. *From* is followed by *for*, which displayed a similar degree of productivity to *from*, yet showing a higher overall frequency.

The importance of the four most productive prepositions is verified when the figures obtained for them are added and related to the total number of prepositional verb types: *To*, *with*, *from* and *for* constitute

58.2 percent of prepositional verb types in TICLE. To put it differently, more than half of the prepositional verbs in the learner data were formed with one of these four prepositions.

As the fifth productive preposition in TICLE we see *on* – also one of the most productive prepositions reported by Biber et al. (1999: 423). Despite its close rank to *for*, its productivity and frequency is not as high; it was used with 34 different verbs, a total number of 166 times.

Table 20. Number of prepositional verb types formed with the productive prepositions in TICLE, along with raw frequency of each preposition

	Preposition	Prepositional verb types	Overall frequency
1	<i>to</i>	95	493
2	<i>with</i>	68	251
3	<i>from</i>	46	249
4	<i>for</i>	45	263
5	<i>on</i>	34	166
6	<i>about</i>	27	231
7	<i>into</i>	24	55
8	<i>of</i>	23	96
9	<i>in</i>	17	70
10	<i>as</i>	15	119
11	<i>against</i>	8	33
12	<i>upon</i>	6	14
13	<i>through</i>	6	9
14	<i>at</i>	4	77
15	<i>over</i>	4	5
16	<i>after</i>	3	52
17	<i>around</i>	3	5
18	<i>toward(s)</i>	3	3

The sixth productive preposition in TICLE is *about*. Although its range is narrower than the preceding preposition on the list, it has a higher frequency, pointing to a higher degree of repetition. It was combined with 27 different verbs to form prepositional verbs in a total number of 232 cases.

Further productive prepositions in TICLE are *of*, *into*, *in* and *as* – all listed by Biber et al. (1999: 423) as productive prepositions but in a slightly different order. Their ranges are a lot narrower in comparison to the earlier prepositions on the list, *as* having the highest frequency, i.e., the highest degree of repetition (see Appendix 7 for the details).

The remaining prepositions on the productive prepositional verb list are *against*, *upon*, *through*, *at*, *over*, *after* followed by *around* and *toward(s)*. Out of these prepositions only *at* was listed in LGSWE by Biber et al. (1999: 423) and it is the preposition that shows the highest degree of repetition in prepositional verb constructions in TICLE: It was combined with four verb types a total number of 77 times. A closer look at the data reveals that nearly all uses of this preposition are connected to one verb: *look at* (72 times). The use of the preposition *after* displays a similar degree of repetition but a slightly lower range: It was attested with three different verb types, in total, 52 times. Interestingly, the data shows a similar tendency of learners to link this preposition with the verb *look* (*look after* was attested 50 times). The preposition *around* with the same productivity level as *after* (combination with three different verb types) displays a remarkable difference with regard

to its overall frequency; it was used only five times in total. A similar pattern is found for the preposition *over*: It has the same degree of productivity as *at* (it was combined with four different verb types) but it shows a much lower degree of repetition – it was used only five times in total. The least productive preposition in Table 20 is *toward(s)*, with 3 verb types and 3 tokens in total. It should also be noted that in one of its uses it was substituted for *against* (*defend N *toward N*).

The comparison of the productive preposition list with the productive particle list has revealed that there is no substantial overlap on the two lists: Only *in*, *on* and *over* are productive as both an adverbial particle and a preposition – the first two are more productive as a preposition. Overall, *over* has the least productivity both as an adverbial particle and a preposition. This result is closely in line with Biber et al.'s (1999: 422) findings.

The overall frequencies of prepositions show a considerable difference amongst them, as can be seen in Table 20 – just as in their productivity. We see prepositions in the data that were used as often as 493 times alongside some that were produced only once. The frequencies of the most frequent five prepositions (namely, *to*, *for*, *with*, *from* and *about*) were added and the sum figure attained was then related to the overall amount of prepositional verbs. The results have revealed a heavy reliance on these five prepositions by the learners in forming prepositional verbs. 67 percent of the prepositional verbs attested in the learner corpus consist of one of these five prepositions.

Summing up, the productivity of verbs and of prepositions used in the formation of prepositional verbs differs to a considerable extent. Accordingly, there are also substantial differences in the overall frequencies of verbs and prepositions used. Not only a higher number of verb types were used to form prepositional verbs compared to phrasal verbs, but also the number of productive prepositions used in forming prepositional verbs was higher. However, unlike in the case of phrasal verbs, only very few verb types were particularly frequent and productive in the formation of prepositional verbs. The most productive verbs and prepositions used in constructing prepositional verbs and the ones used in phrasal verb constructions do not overlap significantly. Learners display a high reliance on a few productive prepositions in forming prepositional verbs.

5.3.4 Synopsis: The quantitative use of prepositional verbs

The quantitative analysis of prepositional verbs in TICLE has shown, *inter alia*, that firstly, the learners made use of more prepositional verbs compared to phrasal verbs, and that, secondly, the range of verb and preposition types used in the formation of prepositional verbs is wider. Despite the wide range of verbs and prepositions, learners displayed a high reliance on a limited number of verb and preposition combinations: The most frequent 25 prepositional verb types account for nearly half of the total prepositional verb tokens in the learner corpus, indicating a tendency of the learners to repeat a limited number of verb-preposition combinations that are familiar to them. Accordingly, only very

few verb types were particularly frequent and productive in the formation of prepositional verbs and these do not closely match the ones used in the formation of phrasal verbs. Moreover, as will be mentioned in great detail in the next chapter, the learners constructed a considerable number of prepositional verbs erroneously. Taken together, these results mean that the learners under scrutiny, although they produced a great number of prepositional verbs in their essays, had not successfully internalized even some of the most common prepositional verbs, displaying often difficulty in choosing the right preposition with a given verb.

5.4 Verb-noun collocations in TICLE

In this section some overall results of the analysis are presented, such as the overall frequency of collocations occurring in each of the three subcategories of verb-noun collocations, the most frequent verb types used in the construction of verb-noun combinations and the productivity of these verbs.

Perhaps a brief reminder is in order before we embark on this agenda. As mentioned earlier (see Section 3.2.4), the definition of a verb-noun collocation in the present study goes beyond the combinations of a verb and a noun, including also the central determiners and other elements that are present. The advantage of their inclusion has been shown for instance by Wang (2016) and Nesselhauf (2005) in that they provide a more complete picture of the problems posed by verb-noun combinations. The verb-noun collocations are classified into three groups:

Group I, the most frequent type, has VO pattern (i.e., simple combination of a verb and a noun – with a possible inclusion of determiners (e.g., *make sense, commit a crime, do the washing, take (one's) time*). The only criterion for the noun in these combinations is that it is abstract in nature, which depends on its use in a given context. The VOO pattern (e.g. *do somebody harm*) – highly rare compared to VO pattern – has been considered a subtype and is also treated here. Group II includes verb-noun combinations followed by a preposition (*make use of, catch a glimpse of, pay attention to*), and Group III includes verb-prepositional phrase unit (*come to a conclusion, bring to light, keep in sight, come (in)to (one's) mind*). All verb-noun combinations occurring in these syntactic patterns – in the case of Group I only those occurring in the essays of at least 3 learners – are to be investigated in the present study, regardless of whether they are appropriate in their contexts. Including hapax legomena in Group I would lead to a considerably longer list (for a detailed analysis of the extracted collocations, see the next chapter).

5.4.1 Overall results regarding the use of verb-noun collocations

Altogether, 391 types of verb-noun collocations were attested in TICLE, occurring in 3,380 instances.¹²² In the formation of these collocations 91 verb types were used as the verbal constituent. All the verb-noun collocations attested in the data are distributed over the three groups as follows: Group I, the largest group, comprises 324 verb-noun collocation types, having overall 3,191 tokens; Group II includes 19 types, having 75 tokens; Group III contains 48 types, occurring overall 114 times. Given the freedom of verbs to occur with a wide range of nouns, especially in interlanguage due to the learners' tendency to produce 'atypical' word combinations in addition to the idiomatic ones in their target language, the high number of verb types in the first group – VO and VOO pattern – is not unexpected. In contrast to low level collocational restriction between the verb and its second element in VO pattern, in VOPO pattern (Group II) collocational restriction may occur between the verb and the first object and between the verb and the prepositional object or the adverbial (Nesselhauf 2005: 69), e.g., *put the*

¹²² Note that both the type and token frequency of verb-noun combinations are actually higher in the data since simple verb-noun collocations (Group I type) which were attested only once or twice or those that were used by only one or two learners were not treated in this study; that is, they were not included in the overall counts. One other thing that is worth noting is that in this count, combinations are considered instances of the same collocation if they consist of the same verb and noun. That is, differing complementations and elements (such as determiners, prepositions, etc.) are disregarded, and different senses of the verb-noun combinations are not distinguished (such as between *give a lesson* and *give sb. a lesson* or *take care of sb./sth* and *take care *about/for sb./sth* or *have a chance of doing sth* and *have a chance to*).

blame on N, take N into account. This fact can explain the lowest type (7 verb types) and token frequency attained for the Group II.

While going through all the verb-noun collocations identified in the data and commenting on each of them would take us too far afield, in what follows I will highlight a few interesting observations, focusing on the token frequency of the most common verb-noun collocations in the corpus involved.

5.4.2 Frequency of individual verb-noun collocations

Some of the verb-noun collocations are particularly frequent in the data analysed. Table 21 lists the most frequent 26 verb-noun collocations – their overall instances comprise nearly 33 percent of the total verb-noun collocations. The verb-noun collocation that has by far the highest overall frequency is *have + right* – used 172 times by 73 learners. It occurs in essays on various topics although there are some specific essay topics that prompted its use, as expected – namely, ‘sex equality’ and ‘human rights’. In BNC, this noun has been found to be one of the most common collocates of *have* (see Wang 2016: 223f). What is interesting to note is that the same noun was attested as the most frequent collocate of *have* in the writings of other learner groups in ICLE; in the essays of Swedish- and Chinese-speaking learners of English by Wang (2016) and German-speaking learners of English by Nesselhauf (2005). However, the overall frequency of this combination – *have a right* – is considerably higher in the writings of Turkish-speaking learners compared to those of other learner groups.

Table 21. The most frequent verb-noun collocations in TICLE

	Verb-noun collocations	Frequency (raw)
1	<i>have (a) right</i>	172
2	<i>have a problem (with)</i>	72
3	<i>solve a problem</i>	70
4	<i>find a job</i>	57
5	<i>give importance</i>	46
6	<i>have a chance (of)</i>	46
7	<i>do (one's) work</i>	38
8	<i>have a job</i>	38
9	<i>commit (*a) suicide</i>	37
10	<i>end one's life</i>	36
11	<i>do a/one's job</i>	35
12	<i>get a mark</i>	35
13	<i>make a decision</i>	34
14	<i>have an opportunity</i>	33
15	<i>commit (a) crime</i>	32
16	<i>give right</i>	32
17	<i>find a solution</i>	31
18	<i>have a difficulty (in/of)</i>	31
19	<i>pass an exam/examination</i>	31
20	<i>use (one's) knowledge</i>	31
21	<i>give an example</i>	29
22	<i>give harm (to N)</i>	29
23	<i>have an effect</i>	28
24	<i>have knowledge (about/of)</i>	28
25	<i>have value (of)</i>	28
26	<i>play a role (in/*on/*about)</i>	28

Have a right is followed, with far fewer occurrences, by *have a problem (with)* – a collocation that shares its verbal component, similar to seven

other ones in the list. The third most frequent collocation has also *problem* as its noun collocate (*solve a problem*). The high frequency of the noun collocate *problem* was not unexpected since most of the essay topics assigned to learners were on social and political problems (see Section 4.2 for all the essay topics). *Find a solution* is another noun collocate relevant in this context. Other frequent noun collocates in Table 21 that were combined with more than one verb (namely, *job* and *knowledge*) as well as specific verb-noun combinations (such as *commit suicide*, *end one's life*, *commit a crime*, *pass an exam* and *get a mark*) point to the presence of topic-dependency in the use of frequent verb-noun collocations in TICLE.

Going over the list, one can infer that it is highly probable that most of these collocations are also frequently used in native speaker English since all of them are appropriate word combinations – taking aside the deviations some of them include in the learner data. Indeed, most of the noun collocates of the delexical verbs in the list (*have*, *do*, *make*, *give*, *get*) are attested as common collocates of these verbs in BNC (see Wang 2016: 223-233). Similarly, 17 of the verb-noun collocations in Table 21, the ones in bold, are listed in LTP dictionary; that means, these words combined by the learners are strongly correlated with each other. However, given the tendency of EFL learners, irrespective of their mother tongue, to frequently employ a number of general words which are usable in large number of contexts (Blum & Levenston 1978; 1980) as ‘safe choices’ – what Hasselgren (1994) referred to as “lexical teddy bears” – one could reasonably expect the existence of “collocational

teddy bears” in learner language (Nesselhauf 2005: 69). The results attained by Gilquin and Granger (2011) in their study investigating the (semi-)fixed expressions with *into* in four ICLE subcorpora (French, Dutch, Spanish and Tswana) confirmed this expectation; the learner groups investigated by the researchers displayed a strong preference for certain expressions that are familiar to them and appear to be safe and repeatedly use them in their language. That is, there is a high possibility that some of the frequent collocations in TICLE were chosen by the learner group under investigation as safe options over more specific, less common alternatives. For instance, *exemplify* and *harm* were used only three times by three learners whereas *give an example* and *give harm* were attested 29 times in the data. Other alternatives of the latter collocation – *injure* and *hurt* – have lower frequency as well – ten and twelve, respectively.

In accordance with the tendency of opting for safer word choices, some learners in the data repeatedly used a particular (topic-prompted) verb-noun collocation – that is the reason for the high frequency of a number of verb-noun combinations attested. For instance, *have a job* and *make a decision* were used eight times by learners TRKE2042 and TRCU1173, respectively, whereas *have abortion* was used 13 times by TRCU1093. Therefore, during the frequency counts the number of learners was taken into consideration. 69 out of overall 391 verb-noun collocation types were produced by at least ten learners in the data – these collocations, which feature among the most frequently produced

ones, are as follows¹²³ (see Appendix 8a for a list of all simple verb-noun collocations appearing in the essays of at least three learners; for a list of all Group II and Group III types of verb-noun collocations, see Appendix 8b and 8c, respectively):

answer a question (14 occurrences, 12 learners), *ask a question* (27, 18L), *cause death* (15, 10L), *cause a problem* (23, 16L), *commit a crime* (32, 15L), *commit suicide* (37, 15L), *do one's best* (16, 15L), *do housework* (27, 22L), *do a/one's job* (37, 25L), *do (one's) work* (38, 29L), *face a difficulty* (10, 10L), *find a job* (57, 29L), *find a solution* (31, 24L), *find a way* (11, 11L), *get a job* (13, 12L), *get a mark* (35, 15L), *give birth* (20, 16L), *give (N) a chance* (25, 23L), *give a decision* (18, 14L), *give (an) education* (15, 11L), *give an example* (29, 23L), *give harm* (29, 22L), *give importance* (46, 30L), *give information* (15, 13L), *give knowledge* (20, 13L), *give (sb.) a lesson* (15, 10L), *give life* (13, 11L), *give an opportunity* (19, 18L), *give permission* (11, 11L), *give right* (32, 25L), *have a (dis)advantage* (21, 16L), *have an (in)ability* (14, 13L), *have a chance (of)* (46, 34L), *have a degree* (21, 10L), *have difficulty* (31, 26L), *have education* (16, 10L), *have an effect* (28, 21L), *have experience* (17, 15L), *have a feature* (18, 15L), *have freedom* (12, 11L), *have importance* (12, 12L), *have a job* (38, 23L), *have knowledge* (28, 19L), *have a lesson* (13, 10L), *have a life* (13, 10L), *have an opportunity* (33, 24L), *have power* (23, 19L), *have a problem* (72, 46L), *have a responsibility* (21, 17L), *have (a) right* (172, 73L), *have a role* (20, 19L), *have time* (13, 11L), *have (a) value* (28, 21L), *live*

¹²³ Some instances include deviations (e.g., additional preposition as in *cause *to deaths* (TRCU1096), additional plural marker on the noun collocate as in *have experiences*). All such deviations are disregarded in this list since they are the topic of the next chapter (see Section 6.3) – most of the deviations can also be seen in Appendix 8.

a/one's life (12, 11L), *make a decision* (34, 17L), *obey a rule* (12, 10L), *pass an exam(ination)* (31, 22L), *pay attention (to)* (25, 23L), *play a role* (28, 21L), *solve a problem* (70, 46L), *spend (one's) time* (22, 20L), *take a decision* (16, 10L), *take (a) place* (24, 21L), *take responsibility (of/for)* (11, 10L), *use a method* (17, 12L), *use (one's) knowledge* (31, 14L), *take care of/about/for* (18, 14L), *come in/(in)to (the/one's) mind* (16, 15L), *take (in)to consideration* (14, 14L).

It turns out these 69 verb-noun collocation types were used overall in 1,821 instances, which account for more than half of all the verb-noun collocations in the data – precisely 53,9 percent. As for the verbal units used in the formation of these frequent collocations, we see that only 20 verb types were used – *have* is the most productive one, occurring frequently with 23 different noun collocates. *Have* is followed by another delexical verb *give*, frequently occurring in combination with 14 noun collocates. The productivity of verbal units of collocations will be considered further in the next section.

In contrast to the frequently produced simple verb-noun collocations in the data, there are quite a few that were used by only one or two learners in TICLE. Here are a few example verb-noun collocations (Group I) attested only once in the data to give a flavor of their lexical features:

abort a baby, *abuse a privilege*, *act one's life* (used instead of 'live'), *arouse sb.'s curiosity*, *attain one's goal*, *begin a new start*, *beg the question*, *bridge the gap*, *build a marriage*, *build one's life*, *broaden one's point of view*, *charge tuition*, *commission a report*, *close one's hair/faces/bodies* (used instead of

‘cover’), *experience a method* (used instead of ‘experiment with’), *fulfill one’s task*, *hold a class*, *hold an opinion*, *hold an examination*, *kill time/democracy*, *kill one’s creativity/love*, *question a decision/the reason*, *raise an objection/the living standards*.

There are some other verb-noun collocations that occur more than once in the data but they were produced by less than three students. For instance, *close the telephone*, an idiomatically odd verb-noun combination used instead of *hang up the telephone*, was produced three times by the same learner; *strive effort* was produced twice by one learner or *own a job* was produced three times by two learners. As can be seen from the examples, the verbal elements of the less frequently produced verb-noun collocations in the data are, as expected, less common and more specific, i.e. their collocational range is more restricted compared to the verbs used in the formation of frequently produced verb-noun collocations.

Compared to Group I, type frequency is a lot lower in Group II and Group III. Out of 19 verb-noun collocation types in Group II, eight were used only once, two only twice. In this group only two collocation types were used more than ten times: *make use of* (13 occurrences, 9 learners) and *take care of* (18 occurrences, 14 learners) – the latter was attested with inappropriate prepositions twice: *take care *about them* [family] (TRCU1109) and *take care *for children* (TRKE2005) (The list of all Group II type verb-noun collocations along with information on their frequency can be found in Appendix 8b). As for Group III, more than

two thirds of all verb-noun collocation types in this group were produced only once (see Appendix 8c). Only two collocations in this group feature among the frequent collocations: *come (in)to (one's) mind* and *take (in)to consideration*. Both of these collocations are the so-called 'stretched verb constructions' (see Nesselhauf 2005: 20f) – some uses of both in the data included deviations from the target:

comes your mind (omission of preposition; TRCU1013)

*come **in to** our mind* (preposition was written separately; TRCU1072)

*comes to **the** minds* (additional use of definite article and plural marker; TRCU1125)

*come **in** our minds* (wrong choice of preposition; TRCU1141)

*issue(s) to be taken **to** consideration* (wrong choice of preposition; TRCU1026, TRCU1046)

As for *take into consideration*, the learner group under investigation showed a marked preference for it over its near-synonym *take into account* – as did French and Spanish-speaking learners investigated by Gilquin and Granger (2011) (cf. Nesselhauf 2009). *Take into account* was attested only once in the data. The learners' preference for *take into consideration* could be explained by the existence of its one-word equivalent, namely, *consider*. That is, the learners might be more familiar or feel safer with this collocation compared to *take into account*, for which no one-word verb equivalent exists that is derivationally related.

There is one more collocation in Group III the frequency of which is closest to the former two: *put into practice*. It was produced 11 times in

total by six learners. One of its instances includes a wrong form of the noun collocate (*put into *practical* (TRME3012)), in one instance the preposition choice was wrong (*put *in practice* (TRKE2031), and still in another instance the intended preposition was written separately as **in to* (TRME3015). Despite the general low frequency of the verb-noun collocations in Group III, nearly half of them include a type of deviation. As will be dealt with in detail in the next chapter, the uses of all types of verb-noun collocations in the data oftentimes, as anticipated, deviate from native speaker norms.

The next section is devoted to the productivity of the verbal constituents of verb-noun collocations.

5.4.3 Productivity of verbs in forming verb-noun collocations

Verbs used in the formation of verb-noun collocations were investigated with regard to their productivity to find out their potential of combining with different nouns to form these constructions. Information gleaned from such an analysis along with frequency information can be very revealing about which verbs are used most frequently and productively as the basis for these constructions, what are the most common noun collocates, and whether there are any instances where the learners overextend the collocational possibilities of a given verb, combining it with an unusual collocate. Note that in the measurement of the degree of productivity of verbs, only the nouns they are combined with, not the whole noun phrase, have been taken into consideration.

Table 22. Most productive verb types used in the formation of verb-noun collocations, along with absolute frequencies (percentage they represent)

	Verb type	Productivity	Frequency
1	<i>have</i>	81	1040 (30,7%)
2	<i>give</i>	29	369 (10,9%)
3	<i>take</i>	28	228 (6,7%)
4	<i>make</i>	17	116 (3,4%)
5	<i>come</i>	16	47 (1,3%)
6	<i>get</i>	15	123 (3,6%)
7	<i>put</i>	11	26 (0,7%)
8	<i>do</i>	10	153 (4,5%)
9	<i>find</i>	9	129 (3,8%)
10	<i>use</i>	7	80 (2,3%)
11	<i>learn</i>	7	36 (1,06%)
12	<i>lose</i>	6	29 (0,8%)
13	<i>provide</i>	6	26 (0,7%)

Table 22 lists the most productive verbs involved in the formation of verb-noun collocations, together with absolute frequencies and the percentage they represent in the total use of verb-noun collocations. It has been observed that, among other things, the most frequent verb types in the data are also the ones that are most productive. Moreover, only the delexical ('light') verbs show high-productivity to a great degree and they considerably vary with regard to both their frequency of occurrence and productivity level. Verbs used as the verbal basis in verb-noun collocations in the three different groups differ to some extent – especially the ones in Group III. Compared to Group II, which comprises of exclusively seven high-frequency verbs (namely, *become*, *bring*, *give*, *have*, *make*, *put*, *take*), we also see (slightly) less common

verbs like *stone*, *suffer*, *sink* and *writhe* in Group III used as bases in the formation of verb-noun collocations. Some of the most productive verbs in Table 22 only occur in simple-verb noun combinations (Group I) – namely, *get*, *do*, *find*, *use*, *lose*, and *provide*.

The verb-noun collocations formed with the 13 productive verbs in Table 22 account for 70,4 percent of all verb-noun collocations in TICLE.

As can be seen in Table 22, the delexical verb *have* is by far the most frequent and the most productive verb, occurring in both Group I and II. It yielded nearly three times more instances compared to the next most frequent verb – namely, *give*. *Have* is accordingly nearly three times more productive compared to *give*. It occurs with 81 noun types (80 in Group I and one in Group II) in 1,040 instances.¹²⁴ It turns out that *have* accounts for 30,7 percent of all verb-noun collocations attested in TICLE.¹²⁵ As mentioned in Section 5.4.2, the most common noun collocate of the verb *have* is *right*; it was produced in combination with *have* 172 times by the Turkish-speaking learners under scrutiny (see Table A9.1 in Appendix 9 for all common noun collocates of the verb *have*).¹²⁶ The occurrences of *have (a) right* scattered across essays

¹²⁴ There are other *have*+noun collocations in the data, but they were not frequent enough to be included in the analysis – only the noun collocates that appeared in the essays of at least three learners were included. The excluded verb-noun collocations represent idiosyncratic rather than typical usage.

¹²⁵ As noted earlier, simple verb-noun combinations used by only one or two learners are excluded in this study.

¹²⁶ The common noun collocates of the six delexical verbs can be found in Appendix 9, together with their raw frequencies.

on different topics. The same noun collocate also tops the list of frequent noun collocates of *have* in the Swedish and Chinese subcorpora of ICLE – SweCLE and ChiCLE, respectively (see Wang 2016) and the German subcorpus of ICLE (GICLE) (see Nesselhauf 2005). However, the overall frequency of *have + right* is considerably higher in the writings of Turkish-speaking learners compared to that of other learner groups.

The second most frequent collocate of *have* is *problem* in TICLE with 72 occurrences. The high occurrence of this noun collocate was not surprising since most topics on which the learners wrote their essays deal with social and political issues, such as ‘human rights’, ‘nuclear energy’ and ‘air pollution’ (see Table 3 for the list of essay topics). This noun collocate has the identical rank in SweCLE, but it occurs in far fewer instances (Wang 2016: 235).

The collocation *have a chance* stands out in the data. It occurs in essays on various topics. As will be discussed in the next chapter, the learners showed signs of confusion in some of its uses, as seen in (77):

(77) *There are a lot of foundations for adopting children. If parents don't **have** available **chances** they can give their child to an available couple* <TRCU1091>

Some other common noun collocates of *have*, in contrast, were obviously triggered by one particular topic or another. The repeated occurrences of the collocation *have + job*, and *have + abortion*, for instance, provide a case in point. While the occurrences of the former were mainly produced in relation to two topics – namely, ‘sex equality’ and

‘the value of university degrees’, the latter one was exclusively used in the essays on ‘abortion’.

All these common noun collocates were also attested by Wang (2016) in the earlier mentioned two other subcorpora of ICLE – namely, SweCLE and ChiCLE. As a matter of fact, a comparison of their lists of common noun collocates of the verb *have* reveals resemblance to a considerable degree, especially between SweCLE and TICLE. As some of the essay topics used in the data complication process of different subcorpora of ICLE are the same, the overlap of (high-frequency) noun collocates among its subcorpora can only be expected. Overall, it can be concluded that although there is a fair amount of overlap with regard to the common collocates used in combination with *have* between the abovementioned subcorpora of ICLE, Turkish-speaking learners used the delexical verb *have* far more productively and more often. The tendency of EFL learners, irrespective of their mother tongue, to use a number of high-frequency, ‘core’ verbs (such as *have*, *do*, *make*, etc.) rather than more specific ones has been widely acknowledged and well-documented in a large body of literature (e.g., Hasselgren 1994; Harley & King 1989; Ringbom 1998a, 1998b; Durrant & Schmitt 2009; Wang 2016), just as the fact that they experience confusion among them (e.g., Alternberg & Granger 2001; Lindner 1994; Lennon 1996; Howarth 1996; Zhou 2016; Wang 2016; Nesselhauf 2005; Babanoğlu 2014; see Section 2.2). Some instances of *have* in TICLE verify this confusion, for instance, between *have* and *get* for the learner group under investi-

gation. In TICLE there are instances of *have education* where the appropriate verb would have been *get* instead. The following is an illuminating example of these instances:

(78) *But as we can see only the rich people's children can **have education** with good conditions, others still **have education** in an ordinary school education is with money.* <TRCU1039>

In another example of this type of confusion, *have a mark* was used by a learner and was corrected to *achieve* and *get* by NS judges:

(79) [...] *some students apply to different kinds of cheating ways in the exams to **have a higher mark** to be successful in the competition.* <TRKE2018>

As mentioned earlier, *give* features as the second most frequent and productive verb in forming verb-noun collocations in TICLE; it has 369 occurrences, combined with 29 noun collocates (27 types in Group I – one of which also occurs in Group III, two types in Group II). Its most common collocate is *importance* with 46 instances. This is a strikingly high frequency given the fact that *give importance* has a very low frequency in COCA. Moreover, *importance* is also not among the common noun collocates of *give* (see Wang 2016: 227f). *Give importance (to)* occurs in essays on six different topics, mostly on the topics of ‘nuclear energy’ and ‘sex equality’. Some of its uses, as in (80), were considered not to be ‘elegant’, albeit correct. In Turkish, there is a literal translation of this collocation (*önem vermek*), the equivalents of which correspond to *care (about)*, *value* and *to pay/give heed to* in English. This dissimilarity between L1 and TL might have contributed to the

prevalent use of this particular collocation in TICLE. Furthermore, a learner consulting a bilingual dictionary to look for a translational equivalent of *önem vermek* encounters a wide range of incorrect equivalents, including *notice*, *place emphasis on*, *pay attention to*, *note*, and *place*.¹²⁷ This may be another possible reason for the abundant uses of *give importance*.

(80) *For instance, in the factories, managers are looking for experienced worker (engineer or something else), they are not **giving much importance to** your university degree and [...]*
<TRCU1147>

The second most frequent noun collocate of *give* in TICLE, *example*, may serve to illustrate a strategy of argumentative writing; in writing classes the learners are usually taught to provide examples to support their arguments. As the next common noun collocate of the verb *give* we see *right* in the list (see Table A9.2 in Appendix 9), which, as mentioned earlier, occurred with a strikingly high frequency in combination with another high-frequency verb (*have*). Both of these nouns are common collocates of *give* in the BNC (see Wang 2016: 227f). However, 16 out of the 28 frequent noun collocates of *give* in TICLE do not feature among the frequent noun collocates of the BNC (ibid.). This can partly be explained by the topic-dependency of some items in TICLE, such as *birth*, *salary*, *love*, *punishment*, *permission* and *lesson*, and partly with atypical, unidiomatic collocations formed with this verb by the learners, such as *give harm*, *give knowledge*, *give a decision*, *give*

¹²⁷ <https://www.seslisozluk.net/de/was-bedeutet-%C3%B6nem%20vermek/>

effort and *give a role* (see Section 6.3 for more on these unidiomatic uses). One interesting finding reported by Wang (2016: 86), which should be mentioned here, is that *give* is usually collocated with nouns that have a positive connotation in the BNC. In TICLE, in contrast, words with both negative and positive connotation occur in combination with *give* (such as *happiness, protection, support, love* as well as *punishment, harm, pain, damage*). This may be an indicator of the learners' lack of knowledge as regards to the collocational preferences of this verb.

Give collocates with a large variety of words, some of which were used by only one or two learners. They mostly represent idiosyncratic rather than usual, typical usage, such as *give leadership, give self-esteem, give truth*, and *give duty*. Looking at all the nouns used in combination with *give*, one gets the impression that this verb was used as an 'all-purpose' verb by the learners in accordance with the 'open choice principle' when they lacked a less common verb with a more specific meaning to fit a given context. To give just two examples to illustrate the tendency:

(81) *It is not possible for me to claim that man and woman is [sic] equal to each other. The first reason for why I think so is the social **roles given** to man and woman.* <TRCU1110>

(82) *A mother or a father who isn't happy can not **give** so much **time** to his-her children.* <TRCU1089>

In (81) the learner produced *give* where *assign* would have been one of the more appropriate alternatives; in (82) *devote* would have been a more appropriate verb to collocate with *time*. The high frequency of

give in verb-noun collocations in TICLE can partly be accounted for by uses as exemplified in (81) and (82), i.e. the learners' preference for *give* over more specific, less common verbs, and partly by its abundant use in forming stretched verb-noun constructions in the data (such as *give an example* ('exemplify'), *give a decision* ('decide'), *give punishment* ('punish'), *give education* ('educate') and *give (a) bribe* ('bribe'). Some of these stretched verb constructions are not common in the TL whatsoever. For instance, *give (a) bribe* occurs only 5 times in COCA whereas its one-word equivalent *bribe* was attested 5,678 times. Reasonably, Wang (2016: 38) has suggested that the learners' tendency to produce stretched verb constructions rather than the corresponding unstretched verbs is a developmental feature of learner writing.

The unidiomatic collocations formed by *give*, as *give death* in (83) and the instances in which *give* erroneously substituted another verb, as in (84), added to its overall high frequency in TICLE:

(83) [...] *in our religion it is stated that the birth is given only by God and the **death** is also **given** by God.* <TRCU1984>
(target: 'take life')

(84) *Due to the fact that taking a degree does not mean anyone is ready for the real difficult life which **gives so many responsibilities** or [...]* <TRME3029> (target: 'require')

Interestingly, however, in the case of inappropriate uses of *give* + noun combinations, the correct alternatives provided by NSs were usually also common verbs. This finding confirms the earlier research that showed the EFL learners' confusion between high-frequency verbs (see

Section 2.3.4.1). For instance, *give effort* was used four times in the data where *make an effort* or *put in effort* would have been appropriate. Similarly, *give a decision*, which was attested 18 times in the data, was corrected to *make a decision*. As these examples reveal, similar to what was found above, the high-frequency verbs *give* and *make* were also confused by some learners in TICLE. Babanoğlu (2014) accounted for this confusion experienced by Turkish-speaking learners by L1 transfer – similar instances in the data, as will be pointed out later on, in which other verbs are collocated with the nouns *effort* (*spend/show effort*) and *decision* (*take a decision*) instead of *make*, lend support to Babanoğlu's argument (2014: 46).

Take is the third most productive and frequent verb in the data, occurring with 28 noun collocates in 228 instances. The noun collocate *place* is the most frequent noun collocate, occurring with *take* in 35 instances, bearing two meanings: 'to occur' and 'to replace', as in *take one's place*. The strong collocational link between the two units was reported in Wang (2016: 229). Out of the 23 common noun collocates in TICLE (i.e., those used by ≥ 3 learners), 14 occur in the top list of noun collocates in the BNC, showing an overlap of semantic preferences between the learners and the NSs (see Wang 2016: 229f). Some of these noun collocates form rather fixed expressions in combination with *take*, such as *take place*, *take care (of)*, *take pleasure* and *take revenge*. The noun collocates of *take* that are not listed in the BNC but in TICLE either show topic-dependency (e.g., *life*, *course*, *lesson*) or deviate from the

target norm (e.g. *knowledge, mark, education, right, degree* – see Section 6.3 for on these deviations). As for the topic dependency, the second most frequent collocate *life*, for instance, was used mainly in the essays written on the topic ‘euthanasia’. Yet, the other two topics chosen by some learners – ‘suicide’ and ‘capital punishment’ – also prompted its use. *Life* also features as one of the most common collocates in SweCLE (Wang 2016: 97). Indeed, TICLE resembles SweCLE to a considerable degree with regard to the common collocates of the verb *take* – both corpora share nine collocates. This adds weight to the argument of topic-dependency since some essay topics are shared in the two learner corpora. Another thing that should be noted in this context is that despite the similarity, TICLE shows a higher lexical variety as regards to the frequent noun collocates of *take* compared to both SweCLE and ChiCLE (see Wang 2016: 97). Several of the recurrent noun collocates of *take* are related to education: *take a course, take a mark, take a degree, take notes, take education, take knowledge(*s)* – the latter two were judged inappropriate by NSs. These collocates were clearly prompted by specific essay topics, such as ‘cheating in colleges’, ‘the value of university degrees and the values and consequences of school interaction’ (see Table 3 for a complete list of topics).

The third common collocate of the verb *take* in TICLE is *care* (18 occurrences), followed in two cases by wrong prepositions: *take care *about/*for*. As will be noted later on in Chapter 6, a considerable number of collocations formed with this high-frequency verb, especially the ones in Group III, include a type of deviation from the target norm. The

aforementioned confusion between high-frequency verbs has also been observed in some uses of *take* + noun collocations – *take* was particularly confused with *get*, as in sentences (85) to (87). The confusion between this deictic pair has already been reported by Lennon (1996: 34) for advanced German speakers of English.

(85) [...] *in Europe, most of the people who **take university education** are adults who has his/here own family, has children.* <TRCU1060>

(86) *Moreover, the radiation created by nuclear power can not be **taken under the control** easily.* <TRCU1133>

(87) *Because of this, students just study in order to **take high marks** and in order not to fail in the class.* <TRME3009>

In another example of similar confusion, *knowledge* was mistakenly combined with the verb *take* instead of *give* or *receive*:

(88) *Is the **knowledge taken** at the university enough to find a good job?* <TRCU1070>

In some instances of *take*, we also observe the earlier mentioned reliance on general, ‘easy’ verbs – either due to the learners’ lack of a more specific, appropriate verb in an attempted collocation (e.g., *grab (sb. ’s) attention*) or their wish not to commit lexical errors (Hasselgren 1994). However, the confusion of *take* with the verbs *get* and *receive* can be directly attributed to the learners’ L1: They all respond to one L1 equivalent, namely *almak*.

The fourth productive verb is *make* in TICLE with 17 noun types, occurring 116 times in Group I. Its common noun collocates have a lower

frequency compared to those of the preceding delexical verbs, especially when compared to *have* and *give*. Moreover, less than half of the common collocates are included in the list of the common collocates of *make* in the BNC (see Wang 2016: 231-232). This is, as will be discussed in the following and further in Section 6.3, mostly due to the learners' inappropriate noun choices to combine with *make* and less due to topic dependency (e.g., *invention* occurs with *make* mainly in the essays on 'great inventions').

The noun collocate employed the most in combination with *make* is *decision* with 34 occurrences. As reported earlier, this noun collocate was also attested in combination with *take* – albeit to a lower extent (16 occurrences). There is another delexical verb that was collocated with *decision* in the data, resulting in an unidiomatic collocation: *give a decision* (18 occurrences). These alternatives were expected since the literal translation of both *give a decision* and *take a decision*, in contrast to *make a decision*, are natural and idiomatic in Turkish. Given the existence of their direct equivalents, their lower frequency compared to *make a decision* is remarkable. The fact that not all learners made use of the two collocations that have direct L1 equivalents support Wang's (2016) claims that common collocations in the TL can be entrenched in the learners' mental lexicon and produced as wholes without referring to L1 (see also Wolter and Gyllstad 2011: 442). Other collocations formed with *make* in the data lend further support to this assumption. For instance, in the case of *make a living* ('*hayatımı kazanmak*') or *make a contribution* ('*katkıda bulunmak*'), there is no direct L1 translation to

fall back on. Admittedly, the facilitative effect of the learners' L1 can be observed to a greater extent in the uses of the common collocates of *make*. For instance, *mistake, practice, plan, project, experiment* and *research* are all collocated with a delexical verb in Turkish – namely, *yapmak* – which corresponds to *make/do* (as will be noted shortly, these two verbs have one translational equivalent in Turkish). The similarity between the TL and the learners' L1 might have played a facilitating role in the acquisition of these collocations. Yet, as some examples reveal, the reliance on L1 occasionally resulted in the use of *make* instead of more specific verbs which would be considered more appropriate in the TL, such as *make an experiment* or *research* rather than *conduct an experiment/research*. In other instances, the over-reliance on L1 by the learners resulted in inappropriate collocations:

(89) *As for our country, our great leader Atatürk made an effort for women rights and **made laws** for their selecting and being selected.* <TRKE2037>

(90) *First of all, we kill a living baby. While **making abortion**, the doctor cuts the baby into parts.* <TRCU1074>

In (89) the alternatives for the target item provided by NSs are *draft laws, put laws into place, sign off laws* – all of which include less common and specific verbs. In (90) *perform* would have been the more appropriate verb for the collocate *abortion*. Likewise, *operation* and *euthanasia* were used with *make* by the learners instead of *perform*.

One other observation in this context was that a considerable number of verb-noun collocations constructed with *make* are stretched verb

constructions, e.g. *make a decision* ('decide'), *make a choice* ('choose'), *make a contribution* ('contribute'), *make a plan* ('plan'), *make a deduction* ('deduce'), *make a comment* ('comment'), *make an improvement* ('improve'), *make practice* ('practice'), *make an invention* ('invent'), etc. Some of these constructions, such as *make practice* and *make an invention*, were considered inappropriate by the NSs. Though extremely rare compared to the corresponding unstretched verbs, both collocations were attested in COCA: *make an invention* occurs 9 times as compared to 21,083¹²⁸ occurrences of *to invent* while *make practice* occurs only 4 times as opposed to numerous uses of its one-word equivalent.¹²⁹ As Nesselhauf (2005: 113f) has noted, there are certain constraints governing the use of stretched verb constructions and their inappropriateness in some instances can sometimes be mainly explained by conventions of usage.

In a few instances in the data, the verb *make* was used where *do* would have been more appropriate, e.g. *make exercises* and *make research*.¹³⁰ The confusion between *make* and *do* was also attested by Yıldız (2016b) in the spoken language productions of Turkish learners at intermediate and upper-intermediate levels (e.g., *do clothes* instead of *make clothes*). Given the fact that both verbs correspond to one single delexical verb

¹²⁸ This number includes also nouns. Although the PoS-tag verb was inserted in the COCA search, a quick survey showed that the nominal uses of the word were not excluded.

¹²⁹ See the preceding footnote. *Practice/practise* occurs 199,237 times in COCA and it was not possible to exclude the nominal uses of *practice*.

¹³⁰ This confusion in TICLE was already reported by Babanoğlu (2014: 46).

in Turkish – namely, *yapmak*, this confusion is understandable.¹³¹ However, this confusion does not only pertain to the learner group under investigation. Wang and Shaw (2008), for instance, also reported that both Swedish and Chinese learners confuse these two verbs, with examples such as *do a great effort* and *make damage* (p. 218; see also Juknevičienė 2008: 6 and Gilquin 2007: 279). Nesselhauf's (2005) findings are in contradiction with these findings. In her data these two verbs did not appear to cause much of a problem; she hence asserted that the proficiency level of the learners might be the deciding factor behind this result (Nesselhauf 2005: 77). Similarly, although some confusion between high-frequency verbs were attested in the writings of Japanese-speaking learners of English – for instance between *give* and *make*, the verbs *make* and *do* were not confused by them, either (see Babanoğlu 2014: 46).

Come follows *make* in the list with a very similar productivity level – although with less than half occurrences. In the data *come* was found to be productive in forming Group III type of verb-noun combinations (14 types). However, most of the collocations formed with this verb display a type of deviation from the norm, such as selection of the appropriate noun collocate or 'complementary' preposition (see Appendix 8c for all its uses).

¹³¹ There is one other delexical verb in Turkish, *etmek*, which can replace *yapmak* in rare cases, e.g. *kahvaltı yapmak/etmek*.

With 123 occurrences in combination with 15 noun types, we see *get* as the next – the sixth – most productive verb in forming verb-noun collocations. It is only used in the formation of simple verb-noun combinations (Group I). With a strikingly high frequency, *mark* stands out in the list of common noun collocates of the verb *get* (see Table A9.5 in Appendix 9). *Get a mark* is a topic-prompted collocation; it occurs mainly in the essays written on two topics related to studying: ‘cheating in colleges’ and ‘the value of university degrees’. Accordingly, several other frequent noun collocates of *get* are related to education and studying: *information*, *degree*, *education*, and *knowledge*. Its combination with *knowledge* was considered inappropriate by the NSs and corrected to *receive* and *attain*. One of the learners who produced *get knowledge*, for example, employs it as follows:

(91) *They are valuable if we consider the **knowledge** that we **get** from the university.* <TRME3031>

As a matter of fact, a considerable number of (common) noun collocates were mistakenly combined with *get*. For instance, *get freedom* was corrected by NSs to *gain freedom*; *get one’s right* was corrected variously in different contexts (*have one’s right*, *claim/fight for one’s rights*, *take one’s right away/deny one their right*); *get pleasure* was corrected to *have the pleasure of* or *receive/take pleasure*, and so on. All these and similar instances attested in the use of *get* provide support for the aforementioned findings with regard to the over-reliance on general verbs in collocations and confusion between common general verbs. The latter

issue can partly be accounted for by the fact that *get* and *take* as well as *receive* corresponds to a single Turkish verb (*almak*).

The next productive verb in the formation of verb-noun collocations, *put*, does not occur in simple verb-noun combinations productively but does so in Group II and Group III. Moreover, it has the lowest frequency with as low as 26 occurrences, compared to the preceding verbs in the list. Most of its collocates occur one single time (see Appendix 8b and 8c). It might be worth mentioning that there are instances in the data where another general verb was used in the construction of a collocation where *put* would have been appropriate, e.g. *give an end to* (3).

Put is followed by *do* with a very similar productivity level but the frequency of the latter is five times more: *Do* was produced by the learners in combination with 10 noun collocates 153 times – all of which are simple verb-noun combinations (Group I). Its most common collocate is *work* with 38 instances, as observed also by Wang (2016) and Nesselhauf (2005) in SweCLE and GICLE, respectively. *Work* is followed by semantically related noun collocates: *job* (35 occurrences) and *housework* (27 occurrences). Its combination with some nouns in the data is unidiomatic and inappropriate, such as *do euthanasia* and *do abortion* where the verb *perform* would have been more appropriate. A similar substitution – namely, *do an operation* – was attested in the writings of Swedish-speaking learners of English by Wang (2016).

Looking only at the common noun collocates of *do* in TICLE (see Table A9.6 in Appendix 9), one gets the impression that the learners' use of

this verb tallies well with that of NSs. Out of its ten frequent noun collocates, seven also feature in the list of the BNC (see Wang 2016: 233). Two of the three unlisted ones – namely, *abortion* and *euthanasia* – were induced by two topics, the titles of which include these nouns. Examination of the less commonly used noun collocates of *do* – the ones used by one or two learners – revealed that this verb was indeed used as another ‘all-purpose’ verb. One gets the feeling that it was combined with nouns whenever the learners lacked the appropriate verb. The following two examples illustrate this tendency:

(92) *Fathers and mothers are sometimes **do** their sons’ **desires** while they ignore their daughters’ desires.* <TRKE2033>
(target: ‘fulfill’)

(93) *If I can not **do** my **school experience** perfectly, I would not know how I can behave my students in different conditions or how I can give answers to their questions.*
<TRME3004> (target: ‘gain/gather enough (teaching) experience’)

Similar instances include *development*, *innovation*, *fault*, *crime* (4, 2L), *event* (3, 2L), *attitude*, *process*, and *connections*. As these examples reveal, the learner group under investigation displays a clear over-reliance on the verb *do* in their use of collocations.

Find comes as the next productive verb in the list as the ninth productive verb. It has 129 occurrences, in a combination with nine noun types. It occurs only in the formation of simple verb-noun collocations. It was combined most frequently with the noun collocate *job* (57 occurrences),

followed by *solution* (31 occurrences). Compared to these two collocates, the other seven nouns it occurs with in the data have far fewer occurrences.

Other verbs in the productive verb list – namely, *use*, *learn*, *lose*, *provide* – are not particularly productive; the former two were produced in combination with seven noun collocates, the latter two with six. Among them only *use* has a high frequency (80 instances). *Use*, *lose*, and *provide* occur only in Group I, whereas *learn* in Group I and Group III. The combination of *learn* with its most common collocate in the data – namely, with *knowledge* in seven instances – was corrected differently in the various contexts the collocation was used: *obtain/take on board/gain knowledge*. *Knowledge* is also the most common collocate of the verb *use* (31 occurrences).

The striking difference between verbs with regard to their productivity in forming verb-noun collocations is worth underlining. Along with the highly productive delexical verbs that are used in combinations with as many different nouns as 81, a considerable number of verb types in the data occur with only one single noun collocate. To put it precisely, out of 92 verb types, 53 verbs (43 verbs in Group 1, 1 in Group II, and 9 in Group III) were used in combination with one single noun collocate, 18 only with two. The importance of the first four most productive delexical verbs (namely, *have*, *give*, *take*, *make*) in forming verb-noun collocations is verified for the learners under scrutiny, when the figures obtained for them are added and related to the total number of verb-noun collocation types: these four verbs constitute 39.5 percent of all verb-

noun collocation types in TICLE. To put it differently, 155 out of the 391 verb-noun collocation types in the data have been found to combine with one of these four verbs. With regard to the overall frequency, more than half of the verb-noun constructions (51,8%) in the data were formed with one of these four verbs: 1,753 out of 3,380 verb-noun collocations.

The last point to be taken up regarding the productivity of verbs in forming verb-noun collocations concerns the non-productive verbs, among which there are some very frequent ones – the ones that were recurrently used by different learners. For instance, *solve* occurs in combination with only one single noun collocate (namely, *problem*) but in 70 instances¹³²; *commit* occurs 69 times in combination with two noun collocates (*suicide* and *crime*). The frequency of these non-productive verbs is indeed peculiarly high compared to some productive ones listed in Table 22. The high frequency of these collocations is explicable by the essay topics assigned to the learners – some of which encourage, if not necessitate, the use of these collocations (see Section 4.2). The impact of essay topics on the learners' repeated use of certain delexical collocations was also reported by Wang (2016: 124), with examples

¹³² As stated earlier, the simple verb-noun collocations (Group I) were treated in this study only when the noun was abstract and used by at least three learners. Although in the Table 21 and Appendix 8a the verb *solve* looks like it was combined with one single noun collocate, in the data it indeed also occurs with the nouns *factor*, *secret*, *matter*, *question* – each only once.

such as *have abortion* (in Swedish learners' essays) and *do business* (in Chinese learners' essays).

To sum up the main findings, first of all, the productive list of verbs involved in the formation of verb-noun collocations in TICLE are predominated by delexical verbs, which also feature as the most frequent verbs. One verb that is particularly noteworthy in this context is *have*. This delexical verb was used in combination with a wide variety of nouns. The most productive delexical verbs account for a large proportion of all verb-noun collocations in the data. The difference with regard to the degree of productivity of verbs involved in the formation of verb-noun collocations is remarkable. More than two thirds of all verb types were only combined with less than three noun collocates – some of which have a higher frequency than more productive ones. Secondly, the learners under investigation demonstrated a considerable degree of confusion with regard to the selection of the appropriate high-frequency verbs in collocations, showing a tendency to overly rely on general, less specific verbs when forming verb-noun constructions. These findings converge with previous studies conducted with learner groups from different L1 backgrounds.

5.4.4 Synopsis: The quantitative use of verb-noun collocations

To conclude this section, the analysis of the three groups of verb-noun combinations has revealed remarkable differences among them with regard to the overall type and token frequencies – Group I (simple verb-

noun collocations) having both the most types and tokens. The difference between verb types with regard to their productivity in forming verb-noun collocations is striking.

Despite the wide range of verbs involved in the formation of verb-noun collocations – namely, 92 verb types – Turkish learners under investigation displayed a high reliance on a limited number of verbs (especially on general, high-frequency verbs, most of which are light verbs): the top most frequent 10 verb types involved in the construction of verb-noun collocations account for 68 percent of the total verb-noun collocation tokens in the data, demonstrating the learners' tendency to rely on familiar verbs as safe choices. Accordingly, only few verb types were particularly frequent and productive in the formation of verb-noun collocations. These results tally well with the earlier research reporting EFL learners' preference for general, less specific verb choices and their reliance on a limited number of high-frequency verbs in forming verb-noun collocations.

Moreover, as will be mentioned in great detail in the next chapter, the learners constructed multitudinous verb-noun collocations that were considered inappropriate or clumsy by the NS judges. A considerable number of high-frequency verbs involved in verb-noun collocations in the data reveal the difficulty they pose to learners in their selection. Taken together, these results mean that although the learners under investigation produced a large number of verb-noun collocations in their

Chapter 5: Multi-word verbs in learner writing – a quantitative approach

essays, they had not successfully internalized even some of the collocational restrictions with the most common general verbs, displaying often inappropriate or deviant verbal selections.

Chapter 6: Multi-word verbs in learner writing – a qualitative approach

Do you know what a foreign accent is? It's a sign of bravery.
(Amy Chua)

6.1 Preliminaries

In the previous chapter, which focused on frequency of multi-word verb use, erroneous and unidiomatic usage identified in the learner data was briefly touched upon. This chapter focuses on this aspect of learner language, providing a detailed description of error types and unidiomatic uses spotted in the learner data as well as plausible explanations for them. Before setting out on this enterprise, a few more words on errors in general and the error classification employed are in order here.

The stance taken here towards errors is not purgative. In contrast, they are considered not only an inevitable part of the language learning process but also necessary for learners to learn their TL, i.e. errors are considered to be used by learners in the process of learning the L2 as strategic devices to test their hypotheses about the nature of the language they are learning (Corder 1967: 25). As Lewis (2002: 168) has rightly pointed out, “[a] learner strategy of avoiding mistakes is always counter-productive.”

Heeding the advice of Schachter and Celce-Murcia (1977) to be “extremely cautious when claiming to have identified the cause of a given

error”, in the present study plausible explanations for the deviant uses of multi-word verbs found in the data are intended rather than claiming to have identified the source of errors, since, as explained by Ellis and Barkhuizen (2005: 66), “[a]n error itself can only provide a hint of its source with the result that many errors are ambiguous. [...] In fact, many errors are likely to be explicable in terms of multiple rather than single sources.” Many errors in learner language, according to Lewis (2002), result from learners’ trying to express something for which they lack the linguistic resources and there are three possible linguistic sources of error: interference, lexical deficiency and partial mastery (p. 171). These possible sources of errors as well as some aspects of learner language such as ‘unnaturalness’ and ‘creativity’ will be addressed in this part of the study.

By ‘unnaturalness’ what is meant here is the unidiomatic use of multi-word verbs, i.e. deviant from NS norms of conventionality, which can materialize in learner language in different ways: (a) learners’ use of an existing multi-word verb in an inappropriate context, with inappropriate collocates – the meaning intended by the learner with a given multi-word verb and the meaning senses provided by the dictionaries for that verb do not match, (b) learners make multi-word verb choices in a context that does not accord with native speaker preferences (cf. Sinclair 1991: 6), and (c) learners’ creatively forming ‘new’ multi-word verbs either according to the grammar rules of the TL or using L1 as a resource. It has already been shown that when learners lack or cannot come up with the word they need to express their intended meaning,

they have a tendency to make lexical choices that are not in line with the target norm; they choose ‘unusual’, sometimes even text-type inappropriate words (e.g., Römer 2009). Granger (1996a: 17) calls this aspect of learner language “non-nativeness” or “foreign-soundingness” – two aspects of advanced learner language, the underlying factors of which the ICLE project aims to reveal. In this context, the verb choices deviating from native-speaker norms are of interest along with the collocational aspects of multi-word verbs.

One important thing to note at the outset is the fact that although an error analysis has been conducted for the qualitative part of the study, quantification of errors of any kind has not been attempted. This decision has been taken mainly on the grounds that errors and their possible causes in the context of learner language do not easily lend themselves to taxonomic classification: For one thing, it is difficult to judge whether a given verb is used in an appropriate context and/or with its right collocates in the case of some verb types, such as phrasal verbs, which often have polysemous meanings – making any decision subjective to a great extent. Furthermore, the boundaries between error sources are not always clear-cut, for instance between L1 interference and formal similarity with another verb – in most cases subjective interpretation as well as judgment would be necessary. Therefore, although reference tools as well as native speakers are consulted in the present study for the acceptability judgments (see Sections 4.3.2 and 4.4), an overall quantification of error types has not been attempted. In

what follows, deviations and creativity observed in learners' use of multi-word verbs will be discussed under some general categories, with no intention of being exhaustive. For illustration purposes some selective examples for each error type will be provided.

In the following, an unacceptable or doubtfully acceptable form or usage is preceded by an asterisk (*). For the sake of focus, syntactical problems of the learners in using multi-word verbs, such as tense, aspect, subject-verb agreement and voice, as exemplified in the sentence (94), were not taken into consideration in the present study.

(94) *Some teachers are too insufficient in their areas to educate students [...] They **are adhered to** course books and do not have any other useful things to add.* <TRME3017>
(confusion between active and passive voice)

The chapter is arranged as follows: Section 6.2 presents the deviation types identified in the use of phrasal, prepositional and phrasal-prepositional verbs, followed by a discussion of possible explanations. Results of a similar analysis conducted in the context of verb-noun collocations are presented thereafter. The chapter concludes with Section 6.4 summarizing the main results.

6.2 Erroneous and unnatural usages in the context of phrasal, prepositional and phrasal-prepositional verbs

One of the research questions of the current study concerned the place of the learner's L1 in the use of multi-word verbs. For the Turkish learners, given the earlier mentioned fact that Turkish has no prepositions and particles, in the case of phrasal, prepositional and phrasal-prepositional verbs, the L1 interference was expected mainly in the form of underuse of these verb types, as well as wrong choice or omission of the preposition and/or the particle. The typological distance, resulting in little lexical commonality between the learners' L1 and the TL, was expected to reveal itself in collocational deviations. The collocational deviations and the use of multi-word verbs in inappropriate contexts were expected to be multiplied by the "coarser-gridding" in the learner's L1. These expectations are borne out in the data. An initial observation suggests that the Turkish learners investigated have a general problem with verb use. In some cases nouns were creatively used as verbs, e.g., *duty* and *engineer*, or single verbs were found to be formed into creative combinations, e.g. *face to face (with) N* or *be face to face (with) N* instead of *face N* – mostly using L1 as a resource. Regarding the use of phrasal, prepositional and phrasal-prepositional verbs, the number of deviations is high. In addition to the expected problems with the use of these verb types, some other deviations have been attested in the data: redundant uses of prepositions and particles

with one-word verbs, preference for high-frequency verbs in combination with a particle where a simple verb would be the more appropriate choice – what Waibel (2007) referred to as “simplified use” of multi-word verbs, formation of new multi-word verbs, syntactical problems such as confusion between transitive and intransitive uses of multi-word verbs. It can be concluded that all the deviations attested in the data in the context of phrasal, prepositional and phrasal-prepositional verbs fall into five (sometimes overlapping) categories: 1) redundant uses of particles and/or prepositions, 2) omission of particles or prepositions, 3) substitution errors, i.e., wrong choice of the verb itself or other constituents of the multi-word verb, 4) unnatural and creative uses of multi-word verbs, and 5) lack of syntactical knowledge resulting in confusion between transitive and intransitive multi-word verbs. Table 23 below provides an overview of these deviation categories with subtypes in each. What follows is a detailed discussion of each deviation type with an attempt to find plausible explanations for the errors involved.

It is important to note the fact that while in some cases the verbal deviations from the native-speaker norm were definitely erroneous, some were acceptable (to an extent). In most cases, the native speakers who were approached for acceptability judgments were able to deduce the intended meaning in the given context and to provide an alternative verb for the erroneous one. The sentences (95-98) below exemplify some

cases where the intended meaning by the learner was not clear to the NSs consulted – which did not cause the same problem to the researcher of the current study who shares the same L1 background with the learners:

(95) *Men can do everything, but women can't. [...] Here we see the pressure which is intensive on women again. [...] Only the educated society, educated women and man prevent this situation. [...] But, the most important duty is the women's duty, they should know their place in the life, and they should **come up with** each matter* <TRKE2034>

(96) *But, **covered with** the influence of the exams on graduating, they hardly apply for another way out of cheating in a direct way.* <TRCU1162>

(97) *Because our universities do not make us learn how to live against the world. We think that we know everything, how to communicate, bahave [sic], **stand by** our feet.* <TRME3020>

(98) *Since people can not eat something ,since they can not find a place to stay, they **apply to** different ways. For example; theft.* <TRCU1056>

In (95) the learner aimed at either *cope with each matter* or *come up with solutions for each matter*. The appropriate verb in this context in Turkish is a multi-word verb (namely, *üstesinden gelmek*) and it is erroneously provided as one of the translation equivalents of *come up*

with.¹³³ In (96) the intended meaning by the learner is ‘overwhelmed with exam anxiety’. In (97) we see that the learner made a wrong choice of preposition – the targeted verb was *stand on one’s own feet*. Sentence (98) is an example of direct translation from Turkish – *farklı yollara başvurmak* (= *resort to different ways*). The verb *başvurmak* can be translated by more than one L2 item, including *apply*, *resort to*, *call on/upon* and *consult*, among others – this ‘divergence’ (Dagut 1977) underlies the deviation.

¹³³ See, for instance, www.seslisozluk.net

6.2 Erroneous and unnatural usages in the context of phrasal, prepositional and phrasal-prepositional verbs

Table 23. Types of error and unidiomatic usage attested in the production of phrasal-, prepositional- and phrasal-prepositional verbs in TICLE

Types of error and unidiomatic use		Examples
1. addition	particle	(99) <i>While performing abortion women may *come up across bleeding and infection problems and [...] <TRCU1079> ('come across')</i>
	preposition	(100) <i>But after the money had entered into their lives, unfortunately people started to do the unbelievable things [...] <TRCU1052> ('enter')</i>
2. omission	particle	(101) <i>To sum, money changes human beings very much and even you can not know them if they do not have a well character. <TRCU1053> ('sum up')</i>
	preposition	(102) <i>I think students must work while they are going on their education. <TRCU1073> ('go on with')</i>
3. substitution	particle	(103) <i>[...] but it is not helping his/her friend in a simple way. It is taking over the other friend's right who efforts [sic] and does it [sic] his/her best. <TRCU1111> ('take away')</i>
	preposition	(104) <i>people find out about every good or bad events happened in the world and they began to *react for bad events which give harm[...] <TRCU1033> ('react to')</i>
	verb	(105) <i>this problem is not sourced from the teachers or the students, the real responsibility is in the government's hands. <TRCU1064> ('result from')</i>
	entire MWV	(106) <i>Also the knowledge people took in universities do [sic] not get on with reality so we can name these knowledge as unnecessary. <TRCU1128> ('match/fit with')</i>

Table 23 (cont.) Types of error and unidiomatic usage attested in the production of phrasal-, prepositional- and phrasal-prepositional verbs in TICLE

Types of error and unidiomatic use		Examples
4. unnatural uses	use of a MWV with an ‘unusual’ collocate	(107) <i>What if they go on cheating in other lessons? I mean that you only carry out your precautions.</i> <TRKE2055> (‘take precautions’)
	preference for a high-frequency verb + particle combination	(108) <i>Lastly, when we get older and our body gets tired, some health problems come out.</i> <TRCU1136> (‘appear’)
	creative MWV formation	(109) <i>For example; the people in Turkey believe that women can never play football or men can never knit up.</i> <TRKE2003> (‘knit’)
5. confusion between transitive and intransitive MWVs	(110)	<i>Moreover, growing up children is also woman’s duty in our society.</i> <TRKE2072> (‘bring up’)

The deviation types given in Table 23 occasionally co-occur, causing their separate treatment to be impracticable. For instance, additional uses of particles and/or prepositions cause a multi-word verb to occur in a wrong context, i.e., occur with wrong collocates. Some concrete examples will help to clarify this point:

(111) *If there is an incurable illness and there is no way to **calm** the pain **down**, the euthanasia might be a good alternative for the person.* <TRCU1160>

(112) *The government also announces to the public to invite them to watch the guilty while he **hangs on**.* <TRCU1001>

In sentence (111) we see an additional use of a particle resulting in the use of the multi-word verb *calm down* in a wrong context. The phrasal

verb *calm down* exists as such and can be used either transitively or intransitively yet it needs to be used in relation to animate objects when used transitively. Therefore, it does not seem to fit when used with a feeling or inanimate objects, which is the case in sentence (111). So, the target intended by the learner in this context is indeed *calm the pain*. Example (112) illustrates a similar instance: The intended verb is the one-word verb *hang* rather than the phrasal verb *hang on*. In this sentence we also see syntactical problems (i.e., confusion between active and passive voice) besides the additional particle use.

Come across can be given as a further example for the overlapping deviation types in the data at hand. It can be either a prepositional or phrasal verb in a given context, depending on the meaning. Out of 18 instances found in the data, nine have been found to be followed by the preposition *with* and the rest are instances for its prepositional use. *Come across with* exists as such in English and carries the meaning of ‘give, hand over (money, information, keys)’ but this meaning is not the intended one by the learners in any of the nine instances (see sentences 113-121).¹³⁴ The preposition used in eight cases is redundant resulting in the use of this phrasal-prepositional verb in inappropriate contexts. If the preposition were excluded, in eight out of the nine instances

¹³⁴ The meanings of the multi-word verbs given in this section are based either on Cowie and Mackin (1975) or the Oxford Advanced Learner Dictionary (2000). The suggested correct alternatives to the unacceptable or problematic multi-word verbs were provided by the native speakers approached.

come across would fit the context. Only in one instance *come across with* is not compatible in the chosen context – even after leaving the redundant preposition out. That is, the deviation type in (121) results from substitution; the multi-word verb as a whole was chosen incorrectly. The intended meaning by the learner in sentence (121) could be expressed either with ‘*when you are faced with*’ or ‘*you are tasked with*’.

(113) [...] *women haven't given the same rights for a long time past, they can compete with men in every subject; especially we can *come across with this situation in this period*
<TRCU1003>

(114) *It can be too useful if you are in o [sic] place where there is no communication system or when you *come across with a car accident in an empty way, in that circumstances if you have m. - phone you can just call from it and get help easily*
<TRCU1016>

(115) *We should respect to everyone's thoughts. Because, no one could understand (if he does not *come across with this situation) the person's mood who have to decide to do that*
<TRCU1078>

(116) *Then you *come across with a society portrait who does not want more information, not criticize just satisfied, believe and accept what they see, hear on TV, radio and read in newspapers* <TRCU1126>

(117) [...] *doctors can *come across with patients suffering from points which will remove later on wanting euthanasia, since they are psychologically affected by their malady*
<TRCU1159>

(118) *In some books, we sometimes ***come across with** the information which states that men and women came to the world at the same time.* <TRKE2008>

(119) *When a child comes to school from its own small world, it ***comes across with** both physical, cultural realities of the real world and is to struggle against all of them* <TRKE2014>

(120) *We ***come across with** the term 'equality' since the existence of the world in every aspect of life* <TRKE2052>

(121) *So, knowledge should not only be given theoretically, It should not be forgotten that no matter how much you read, study or memorize, you can not be successful when you ***come across with** the application of what you do if you have never tried it* <TRME3023>

The same verb and preposition/particle combination (*come across*) was attested in another instance in TICLE with an additional particle, resulting in an unusual (i.e., unrecorded) phrasal-prepositional verb: **come up across* (see sentence (99) in Table 23). Here the correct verb intended by the learner is either *come across* or *come up against* – so, the learner either used an additional particle or made an incorrect choice of preposition.

In TICLE, out of 17 learners who produced the prepositional verb *come across* in their essays, only seven managed to do so correctly. The difficulty caused by this polysemous multi-word verb for Turkish learners under scrutiny might be related to its Turkish equivalents in the meaning of 'to meet someone or discover something by chance': *karşı*

karşıya gelmek/karşılaşmak or *ansızın bulmak*¹³⁵ – they all necessitate the use of postposition *ile* with their object (which equates to the English preposition *with*). *Gelmek* is the Turkish equivalent of *come*, and *across* can be translated into Turkish in more than one way depending on the context: *karşı(sında)*, *karşıdan karşıya*, etc. Looking at the contexts where learners used this verb, we can clearly see that most of the learners translated this multi-word verb directly from Turkish into English because in sentences (113) to (121), the use of the postposition *ile* would be necessary after its direct object in Turkish: *bir durum ile karşı karşıya gelmek* [lit. trans. **come across with a situation*].¹³⁶ The same L1 influence was attested in TICLE in the use of another multi-word verb with a similar meaning: *run across*. Unlike in the case of *come across with*, **run across with* does not exist as such – in this case, the additional preposition results in an unusual phrasal-prepositional verb (sentence 122):

(122) *In our country, as the women haave [sic] got no economic freedom and they depend on their husbands, they can ***run across with** some difficulties.* <TRCU1165>

There are some other instances of an additional use of prepositions where the influence of the learners' L1 is clearly discernible. For instance, the additional use of *with* in a similar fashion to the two former

¹³⁵ <https://dictionary.cambridge.org/dictionary/english-turkish/come-across-sb-sth>

¹³⁶ The verb *come across* was reported to cause similar problems for Iranian learners by Mazaherylaghab (2013: 142f), who accounted for the problems by L1 interference.

multi-word verbs has been attested in the use of other verbs (see sentences 123-125). Although the verbs *marry*, *meet* and *encounter* are one-word verbs in English and can directly be followed by their objects, their equivalents in Turkish – *evlenmek*, *buluşmak*, *karşılaşmak*, respectively – necessitate the use of the postposition *ile* after their direct object. Carrying this L1 pattern into English results in the redundant use of the preposition *with*, as in (123) to (125) (see also sentence (53) in Section 5.2.2). In a similar manner, verbs the objects of which require an ablative or dative ending in Turkish were used by the learners with the equivalents of these endings in English – namely, the preposition *from* and *to*, respectively. For instance, the verbs *divorce* [‘boşanmak’] and *respect* [‘saygı göstermek’] necessitate ablative ending (-*den/dan*) and dative ending (-*e/-a*) in Turkish, respectively (see sentences 126 and 127).

(123) *Every one says that the happiest day of a person's life is his marriage, because most people *marry with person they love, [...]* <TRCU1088>

(124) *For example I *met with a man who don't [sic] like his wife and wants to divorce.* <TRCU1089>

(125) *Of course, in some parts of the country it is likely to *encounter with some parents who still use this way but [...]* <TRCU1034>

(126) *As far as I know, almost all religions don't want people to *divorce from their husbands or wives, but [...]* <TRCU1154>

(127) *We should *respect to everyone's thoughts.* <TRCU1078>

Examples (123) to (127) are typical cases of direct translations from the learners' L1.

The instances in the data where the preposition *to* was used redundantly as a counterpart of the dative ending are plentiful (e.g., *begin to one's job*, *look after to children*, *be against to N*, *resemble to N*, *attack to N*, *help to N*¹³⁷, *join to*, etc.). In some other cases, the preposition *in* is used redundantly in a similar manner by the learners (e.g., *join in*, *attend in*, *enter in*, etc.).

There are few cases where we see additions of different prepositions to the same verb. The verb *face*, for instance, which has proved to be a very problematic verb for Turkish learners, was attested with two different redundant prepositions, namely *with* (13 times) and *to* (12 times), as illustrated in (128) and (129). This variety in the addition of units probably results from the fact that the verb *face* has different equivalents in Turkish that require uses of different postpositions or case endings (e.g., *göğüs germek*, *karşı karşıya kalmak*, *yüzyüze gelmek*, *yüzleşmek* – the first equivalent requires the dative case ending (-e/-a) whereas the latter three require the use of postposition *ile*). Given this fact in the learners' L1, the choice of redundant prepositions seen in (128) and (129) – *zorluklara göğüs germek*; *bir problem ile yüzyüze gelmek/karşı karşıya kalmak* – is not surprising:

¹³⁷ The prepositional verb *help to* exists as it is but in the meaning of 'serve sb [guest, diner, oneself] with fish, meat, etc.].' yet the use of the preposition in combination with the verb *help* in the data is redundant: *help to (the) patient* (4), *help to the person* (1).

6.2 Erroneous and unnatural usages in the context of phrasal, prepositional and phrasal-prepositional verbs

(128) *First of all, when a lot of new graduated people start a new job without work experience, they **face to** a lot of difficulties in real life because they have theoretical knowledge about their job, yet, [...]* <TRCU1124>

(129) *[...] married women felt themselves free, when they **face with** a problem in their marriage, they found divorce as the easiest way of rescuing.* <TRCU1088>

Although not all the additional uses of prepositions and particles attested in the learner data can be accounted for by L1 interference – for instance, the erroneously added *up* in *come up across* (sentence 99 in Table 23) and the prepositions in the examples (130) to (132), there are in general some clear patterns in the use of additional particles and prepositions among Turkish learners, showing heavy reliance on their L1.

(130) *We can **call for** these years matriarchal.* <TRKE2002>

(131) *[...] the literal meaning of intelligence is the person's level of capacity to find solutions for the problems that they ***are faced up with**, [...]* <TRCU1006>

(132) *If I can not give answers to their questions truly, they [students] could not ***appreciate with** me.* <TRME3004>

Moreover, in the formation of 'new' prepositional verbs by means of adding redundant prepositions to one-word verbs, we find similar types of nonstandard forms to the ones reported in the literature for other (learner) varieties. This can be illustrated with analogous examples of *discuss about* as in (133) from Nesselhauf's (2009) data, and (134) from

TICLE – in which *discuss about* was found to occur seven times.¹³⁸ Finding similar examples across different varieties, Nesselhauf (2009: 19) claimed that the existence of a derivationally related noun, which requires the preposition in question, is an influential factor in the formation of new prepositional verbs – in the case of *discuss about*, the derivationally related noun is *a discussion about* N (see also Schneider and Gilquin 2018).

(133) *I want to move on next to **discuss about** tertiary treatment systems.* <ICE- Sing>

(134) *I only want to **discuss about** the inequality between these two gender [sic].* <TRCU1137>

In a similar manner, it can be claimed that *care (v) of* in (135) might arise from an analogy with *take care of someone* or *for care of someone*.

(135) *Generally, woman has the responsibility to do all the works in the house such as, **caring of** children, doing the shopping [...]* <TRCU1135>

One other impact suggested by Nesselhauf (2009: 19f) in the formation of new prepositions by adding a “non-L1” preposition to a verb is supported by the results obtained: According to her, the additional preposition “is used in the meaning assigned to it in similar constructions, for example with semantically similar verbs, or even more generally”

¹³⁸ *Discuss about* is listed as a common ‘error’ of foreign learners of English in the *Longman Dictionary of Common Errors* (Turton & Heaton 1996) – a learner-corpus based dictionary that deals with common errors attested in the written productions of learners who are at intermediate learner proficiency and above.

(ibid.). In the case of, for instance, *discuss about* and *mention about* – both of which were attested in TICLE more than once, as in (134) and (136), their formation is “licenced” by the existence of other verbs with similar meaning that are used with the preposition *about* (e.g. *talk about* or *speak about*) and the more general use of *about* in the sense of ‘on the subject of sth’ (Nesselhauf 2009: 20).¹³⁹

(136) *I have mentioned about the balance between the quality of education and development of a country at the beginning of my essay.* <TRCU1065>

Oppose to is another example that can be given in this context. The additional use of *to* in combination with *oppose* (attested three times in the data) may arise either from analogy with the semantically similar verb *object to* or from confusion with the standard form *be opposed to* – passive construction with the same lexical word and preposition. This is indeed the third influence Nesselhauf (2009) mentions in the formation of new prepositional verbs: “the existence of the lexical word + preposition combination either with other meanings or in other constructions” (p. 20). As seen in (137), the learner used *be opposed to* correctly but not the verb *oppose* in the active voice. Sentence (138)

¹³⁹ *Mention about* is also listed as a common ‘error’ of foreign learners of English in the *Longman Dictionary of Common Errors* (Turton & Heaton 1996). Accordingly, Nesselhauf (2009: 18) attested this prepositional verb in various ESL varieties as well as in ICE-GB. She has noted that this prepositional verb is not recorded in most dictionaries and is usually regarded as non-standard yet it occasionally occurs in British English (mainly spoken).

shows a similar confusion experienced by another learner with the use of the verb *constitute* – most probably due to the existence of *constitute of* in passive (*be constituted of*).

(137) *As a conclusion, there are some people who **are opposed to** the idea of euthanasia and there are others who defend the right of patient's wish of death. [...] On the other hand, I ***oppose to** the idea, since a young person should be supported by medical treatment and s/he is given the pleasure of life.* <TRCU1134>

(138) *I also believe that such societies ***constitute of** people who are uneducated because an educated society does not divide people into [...]* <TRCU1120>

Confusion was also attested in the use of the text-structuring device mentioned in the preceding section: In (139) the learner mixed two similar text-structuring devices used to conclude essays – namely *in sum* and *to sum up*, resulting in an unusual blend:

(139) ****In sum up**, there are something different except the presence of men and women in life that changes the fate of the human beings.* <TRKE2002>

In TICLE, additional particles were attested even with verbs that are high in frequency, as the following two sentences illustrate, where the correct verb should merely be *find*:

(140) *Are men and women equal? I think this one of the most difficult questions to which human being have not **found out** a satisfactory answer yet.* <TRCU1110>

(141) *He **found out** 'conditioned learning' method.* <TRKE2032>

(142) *So, when they do this experiments, they have to use microscopes; otherwise they could not **find out** that virus.*
<TRCU1021>

The additional particle used with *find* in these contexts (140-142) cannot be explained with L1 interference. Besides, similar instances of additional particles have been reported to exist in other learner varieties (e.g., Mondor 2008, Gilquin 2011). After having chosen the correct verb, why would the learners use an additional particle? One possibility is considered to be the learners' knowledge of the aspectual meanings the particles contribute to the verbs with which they combine, i.e. knowledge of existing patterns of forming aspectual phrasal verbs (see Section 3.1). As one of the most productive particles in forming phrasal verbs by means of combining with common verbs such as *be, come, go, do, make, take, put* and *give* (see Section 5.2.3), *out* adds several aspectual meanings to the verbs it is attached to. Being one of the particles forming the so-called "completive phrasal verbs", in *find out*, *out* adds durativity to the punctual achievement verb *find* (Celce-Murcia & Larsen-Freeman 1999: 433). This fact may have triggered its use by the learners in order to stress the process aspect of looking for *an answer, method* or *a virus* in sentences (140) to (142). In sentence (143), we have a very similar case – the semantic contribution of the particle *up* is completive in this context and similarly adds durativity to the punctual achievement verb *end*. Akin to *find out* in the aforementioned sentences, *end up* is not the appropriate choice in this context; *end* as a one-word verb would be the appropriate choice.

(143) *Every body [sic] should know that each individual human life begins at the beginning [sic], at fertilization and it ends up at the normal time of death.* <TRCU1091>

Whereas the phrasal verbs *find out* and *end up* exist as such in English, some ‘new’ phrasal verbs, i.e. unrecorded in common phrasal verb dictionaries¹⁴⁰, were attested in the data that were creatively formed by the Turkish learners, such as *knit up* in sentence (109) in Table 23 and *heighten up* in sentence (144), where the appropriate verbs should be *knit* and *heighten* or *raise*, respectively. In these two cases we are dealing with aspectual phrasal verbs again – the particle *up* adds the aspectual meaning of completion of an action: It either turns an activity verb into an accomplishment, as in the case of *drink up*, or it reinforces the sense of goal orientation in an accomplishment verb, as in *close up* (Celce-Murcia & Larsen-Freeman 1999: 433). Following these patterns in forming aspectual phrasal verbs, the learner in (109) creatively turned the activity verb *knit* into an accomplishment verb. Although *knit up* is not listed in the phrasal verb dictionaries consulted, it would be unfair to call this use erroneous since such phrasal verbs are created by analogy fairly frequently by NSs (cf. Prodromou 2005). The productivity of the particle *up* has been expressed most cogently by Bolinger (1971): Claiming that *up* is nearly as free as the prefix *re-*, he has stated that if we were to hear a sentence such as “*Let’s barter up*”, we would not tend

¹⁴⁰ Gilquin (2011) refers to these nonstandard phrasal verbs as “innovations”. Based on data derived from ICE and ICLE, she has shown that they exist in both ESL and EFL varieties and some of these “innovations” are shared by the two varieties.

to react to it as deviant but rather “as something unknown” (p. 101). Asserting that nonce-formations with the particle *up* are normal, Bolinger has explained that the only occasion where the combination of *up* and a verb will be perceived as deviant is “if the verb proper does not admit of any relevant directional or aspectual meaning”, as in **He disgusted up his friends* (ibid.). For the learner who produced sentence (109) it is reasonable to claim that s/he relied on her/his knowledge of this existing pattern in the TL.

(144) *People use the animals to **highten** [sic] the standarts of their living up* <TRKE2038>

Similarly, in (144) the learner uses *up* with an accomplishment verb but here, unlike in the earlier instances exemplified, we see that the particle carries a meaning that already exists in the verbal element, resulting in semantic redundancy. The existence of “redundant” particles in forming phrasal verbs has already been acknowledged and documented in the literature: For instance, Bolinger (1971: xii), claiming that many Latinate ‘equivalents’ of phrasal verbs themselves picked up “redundant” particles, stated the following:

[...] we would not feel comfortable now without the *through* in *With a little pressure the needle finally perforated through*. And forms like *extend out*, *refer back*, and *proceed forth* are common, though some, such as *retrieve back*, *extract out*, and *include in* may be regarded as non-standard (Bolinger 1971: xii).

The notion of semantic redundancy has also been reported in learner language, for instance, by Mazaherylaghab (2013) for Iranian learners:

He, with good reason, claimed that the use of superfluous particles resulting in semantic redundancy may result from, besides the L1 interference, learners' intention to be on the safe side with the meanings they are trying to convey (Mazaherylaghab 2013: 142ff). Other instances for the use of redundant particles were also attested in TICLE, two of which are displayed in (145) and (146):

(145) *It gave very by damages to Turkey, many people died, the buildings were **collapsed down**.* <TRCU1019>

(146) *No one deserves to be killed whatever the crime is because it's impossible for the dead to **return back**.* <TRCU1105>

Collapse down was unexpectedly found to occur both in COCA and BNC¹⁴¹ in various registers – 27 times in COCA (magazine, fiction, spoken, news) and 3 times in BNC (fiction and academic writing). As mentioned in Section 2.3.4.3, additional uses of prepositions have been reported to occur not only in EFL varieties but also in ESL varieties (e.g., Nesselhauf 2009; Mukherjee 2009; Gilquin & Granger 2011; Schneider & Zipp 2013; Edwards & Laporte 2015; Mwangi 2004). Therefore, GloWbE was consulted as the next reference corpus and 73 instances of *collapse down* were found occurring in several varieties (in both inner and outer circle varieties): namely, in American, British, Canadian, Irish, Australian, New Zealand, Indian, Bangladesh, Singapore,

¹⁴¹ The reason for also referring to the BNC corpus in addition to other reference tools is triggered by the differences seen in the native speaker judgments in the present study. Besides, the fact that there are differences with regard to the use of phrasal verbs between varieties of English has been noted in various dictionaries and shown in studies (e.g., Oros 2006).

Malaysia, Philippines, Nigerian, Tanzanian, Jamaican, Sri Lanka, Hong Kong and South African English. Two of the NSs approached for acceptability judgments found the combination erroneous; another commented with “duplication”.

Unlike in the case of *collapse down*, *return back* may be an L1-induced error besides the possible motive of the learner for being on the safe side; in Turkish it is conventional to say *return back* (*geri dönmek*), in which the particle is optional. The existence of *return back* in COCA has already been noted by Mazaherylaghab (2013: 144f): The researcher has, however, stated that in most cases the particle *back* is a part of a time or place adverbial phrase, as in *return back home* or *return back to the airport*. However, as can be seen in sentence (146), the particle *back* forms a unit with the verb in the learner’s production, giving rise to redundancy. Having attested the very same multi-word verb along with similar ones (e.g. *enter into*, *approach to*) across both ESL and EFL varieties, Nesselhauf (2009: 20) has suggested that such redundant uses may indicate the learner’s tendency to make the direction expressed in the verbs of movement more explicit. This “hyper-explicitness”, to use Edwards and Laporte’s (2015) terminology, has been observed in other ‘new’ prepositional verbs attested in TICLE, e.g., *enter to*, *enter into*, *enter in*, *reach to*.

The instances of redundant uses of prepositions and particles exemplified from the learner data so far can be seen as an indicator of, firstly, the learners’ underlying awareness of the prepositions/particles, i.e., their role in forming multi-word verbs, and secondly, a wish to employ

them in their language. However, contrary to additional use of particles and prepositions, their omission in the use of multi-word verbs is also a common deviation spotted in the learner data – an expected deviation for the learner group under scrutiny due to the absence of these units in their L1. For instance, quite a few uses of the most frequently produced phrasal verb in the learner corpus – namely, *go on* – show deviance from the target norm. For instance, in some cases, as mentioned in Section 5.2.2, the intended verb by the learner was actually a phrasal-prepositional verb (e.g., *go on with*) with the meaning of ‘continue’ but the verb’s third constituent – the preposition – was missing and the phrasal verb *go on* was directly followed by a noun phrase. The following sentences from the data illustrate this deviation type:

(147) *As a result some will study and can not find a job and others will not be educated to **go on** a living.* <TRCU1061>

(148) *Because they do not need to **go on** their marriage, they have their own money, [...]* <TRCU1088>

(149) *In conclusion, each country has to keep its regularity and prevent anarchy to **go on** its existence.* <TRCU1119>

Sentence (101) in Table 23 illustrates a case of particle omission; this time in the use of second most frequent phrasal verb in the learner corpus – the one used as a text-structuring device to conclude essays: *sum up*.

In sentences (150) to (152), we see that the prepositions of the prepositional verbs are missing – *as*, *on* and *to*, respectively:

(150) [...] *we are not regarding the people who are in pain so it may be **defined** a kind of egoism.* <TRCU1081>

(151) *To be alive in the hospital room with **depending** some machines like a plant with feeling nothing or feeling just an incredible and unstandable pain mean for me dying every day.* <TRCU1078>

(152) *If they **apply** the same job of course Hacettepe University graduate will get the job.* <TRCU1068>

The number of instances in the learner data where the prepositions of the prepositional verbs are omitted as in (150) to (152) is high, including the uses of *go to*, *concentrate on*, *object to*, *study for*, *talk about*, *divide into*, *look at*, *participate in*, *swear at*, *decide on*, among others. All these cases of omission can be seen as an indication of partial lexical knowledge of the learner.

Related to the notion of omission, there is one other type of erroneous use of multi-word verbs, which fall under the last category in Table 23. It results from either confusion between transitive and intransitive multi-word verbs or lack of lexical knowledge in this respect. In some cases, objects of transitive multi-word verbs were left out, resulting in incorrect uses of transitive verbs as intransitive ones, as in (153), whereas in some other cases intransitive multi-word verbs are used as transitive ones, as can be seen in (154) and (155). In both cases, the result is the use of an existing multi-word verb inappropriately in its context.

(153) *A girl student says All I do is write all of the answers under my skirt and when I'm taking the test, I **hike up** a little so that I can read the answers of my legs <TRKE2062>*

(154) *While as a human being, no one can't **stand by** this kind of illness, how can we think that a doctor, who is also a human, wait and watch such an end. <TRCU1156>*

(155) *Before people were respectful for each other and they **stand** every conditions **together**, but now the power of tolerance of the people is less and less and they hurt each other quickly. <TRCU1085>*

Stand by as a phrasal verb in the meaning it carries in sentence (154) (namely, ‘observe events which require an active response without doing anything’) cannot be followed by a NP. Instead, it would need to be phrased along the lines of ‘*no one can **stand by** idle while this kind of illness takes its course*’. In (155) the selection of the multi-word verb is not incorrect semantically (‘be united’ (e.g., in the face of some outside threat); ‘stick together’) yet its use as a transitive phrasal verb is – the proper collocate of *condition* in this context would be, for instance, *endure*.

This type of deviation has also been spotted, as mentioned in Section 5.2.2, in the use of the third most frequent phrasal verb in the learner data: *grow up*. It was used numerous times as a transitive verb – instead of *bring up* (see sentence 60). Here are some further examples illustrating this confusion:

(156) *If the baby is born, it will be lack of a real family and affection of them. As it **is grown up** in an unhealthy environment*

[sic] *it may have some shortcomings in its character when it become an adult.* <TRCU1100>

(157) *Today, still most of the babies are born with many diseases and most of the plants that **could grown up** in that land in the past, now, **may not grown up** in the present.* <TRCU1117>

(158) *To sum up, universities are very important units which constitute a society's educational system. They purpose to **grow up** individuals well-educated and useful for their community. [...] Universities with all areas interact with industry, private and public foundations and social sciences. They **grow up** the students according to their demands and needs.* <TRCU1129>

(159) *According to people who **are grown up** under the effects of the traditional beliefs thinks that, [...]* <TRCU1174>

(160) *Turning back to the main point, when the lessons generally depend on theory, universities **can not grow up** qualified individuals.* <TRME3027>

One can speculate that perhaps this confusion is teaching-induced – students are not always explicitly taught the difference between transitive and intransitive verbs. Another possible, more plausible, reason behind this confusion is the formal similarity of the verbs *grow up* and *bring up* in Turkish: Their base is the same and the only difference is the causative suffix in *bring up*, which makes the verb transitive in Turkish. The common Turkish equivalent of *grow up* (in the meaning of ‘to become older or an adult’) is *büyüme*, and that of *bring up* is *büyütmek*. Even though it is unconventional to say ‘*universities bring up students/people*’ (intended verb meaning here is ‘educate’) in English, it is

the correct verb choice in this context in Turkish. Lack of necessary contextual knowledge of *bring up* compounded by the L1 influence may have triggered the learners to extend the verb's scope to inappropriate contexts. The transitive use of *grow up* has also been identified in the writings of other learner groups; for instance, in Iranian and Italian learners' writings by Mazaherylaglab (2013: 147f) and Riguel (2014: 13f), respectively.

Incorrect choice of multi-word verb itself or one of its constituents, i.e. substitution errors, is another deviation attested in the learner essays. Although the number of the incorrectly chosen multi-word verbs as a whole is high in the data, it is usually one of its constituents that is chosen incorrectly in the relevant context. In sentence (161), for instance, the learner made an incorrect choice of the entire multi-word verb since *pass over*, although it exists as a phrasal verb with various meanings in English, does not fit the context – its given dictionary meanings are different from the one intended in (161) – *go through* or *complete* would be the appropriate verb in that context. Looking at the noun phrase following the verb in the learner's production, the chosen verb *pass* by the learner can be related to the Turkish verb *geçmek*. The Turkish equivalents of the noun *period* [*dönem*, *süreç*] collocate habitually with the verb *geçmek*: *bir süreçten/dönemden geçmek* [lit. trans. *pass from a period*]. Whereas the verbal part of the incorrectly chosen multi-word verb can be explained by L1 interference, the particle cannot. Nor the verb choice as a whole in (162) – which as a phrasal and prepositional verb exists as such in English but took on a meaning in the sentence

produced by the learner that diverges from the two meanings provided in the dictionaries ('to increase in extent and quantity; supplement' and 'make sense, present a reasonable picture; lead to an obvious conclusion'). The meaning intended with the multi-word verb used in (162) can, however, be understood with the help of the larger context; what the learner probably aimed for should alternatively be expressed as *form an opinion on*. A similar case is observed in the use of *get on with* in (106) (see Table 23); the meaning intended with *get on with* diverges from the dictionary meanings of this phrasal-prepositional verb – 'have a harmonious relationship with; continue to do [work, job; homework, etc.]'. In this context the learner's intended meaning could be conveyed with *match with* or, alternatively, *fit with*.

(161) *Most of the time the students **passed over** their training period without doing extra studies [...]* <TRCU1064>

(162) *After a while, they decide to which way they should follow, which technique can be used thanks to continuity of training period, but just as they **add up** how a good teacher should be in their mind, the training period ends.* <TRCU1067>

In example (163), where the intended verb by the learner is probably the prepositional verb *cope with*, we see that the learner not only mistakenly chose an inappropriate multi-word verb but also got confused

with the order of particle and preposition – a deviation type attested only once in the data.¹⁴²

(163) *While they are walking around the street, or while driving a car they are felt it, with some rude words and other similar bad behaviours. In order to prevent from it women try to be strong but it is really very hard to **come with up**.*

<TRKE2034>

Sentence (164) illustrate where the learner got the verbal part right but had difficulty in selecting the appropriate particle – the particle intended is probably *out* in this context. Sentence (165) illustrates the difficulty experienced by another learner with the use of particle *over*, again in combination with the verb *pass* (see sentence (161)). As understood from the context, the learner indeed aimed for the phrasal-prepositional verb *pass on to* – the equivalent of which in Turkish is a one-word verb *aktarmak*, the indirect object of which necessitates the use of dative ending (the counterpart of the preposition *to* in English).

(164) *Another point is our society is always looking for a quick way out instead of **sticking it through**.* <TRCU1176>

(165) *Insofar as the fact that human life is transcient and the weight of the knowledge to **pass over** the next generation is incredibly huge, we are left with only two alternatives.*

<TRME3022

¹⁴² Gilquin (2011) has shown that *cope up (with)* is a shared ‘innovation’ in ESL and EFL varieties.

The substitution of particles in these cases cannot be accounted for by L1 interference. That is, the learner for no obvious reason chose to combine verbs with these particles. There is, however, one interesting type of deviation attested in the data where we clearly see the L1 patterns transferred to L2. As pointed out in 5.2.2, in four instances the one-word verb *overcome* was confused with the phrasal verb *come over* and used as a transitive phrasal verb, collocating with *problems* in three instances and with *life* in one instance (see sentences 166-169). This confusion can easily be accounted for by L1 influence. The Turkish equivalent of *overcome* is a multi-word verb (*üstesinden gelmek*) and its literal translation is *come over NP* [e.g., *a problem* or *a difficulty*]. Carrying over L1 patterns into the TL resulted in an entirely incorrect use of the phrasal verb *come over*:

(166) *So then, we can **come over** this problem, as well.*
<TRCU1123>

(167) *In order to solve this problem there are a lot of things to do. As a first step; education is the most efficient way while trying to **come over** this problem.* <TRKE2071>

(168) *As a conclusion, the thing which we call 'real life' is something really difficult to cape [sic] with to learn how to **come over** it no university programe [sic], even the best one, is efficient enough on its own.* <TRME3019>

(169) [...] *because they would be skilfull [sic] enough to find alternative solutions to problems even the ones that seems to be impossible to **come over**, then the university degree becomes the most valuable thing to passess [sic] on the earth.*
<TRME3019>

In the case of another one-word verb – namely *overtake* – we see the same process: A learner divided the verb into two units and used it as a transitive phrasal verb – as *take over (men)* in sentence (170). Yet, here the Turkish equivalents of *overtake* (namely, *sollamak, yetişmek, geçmek*) cannot account for the erroneous use. A confusion between two verbs due to formal similarity (i.e., sharing the same units as their constituents) is a more probable reason behind this substitution error.

(170) *Even though men seems [sic] to be the winners of these arguments, women can easily **take over** them in those categories when they are allowed to do their best* <TRCU1102>

Substitution errors in the data occur most often with prepositions. The following few examples are taken from the learner data and are to give a glimpse of the situation:

(171) *Today when we **look over** the world, we can easily see that the consciousness of sex equality is improving step by step.* <TRKE2060>

(172) *If you remember some teenagers who were less than fourteen were **arrested of** stealing baklava in G.Antep and fined many years* <TRKE1053>

(173) *In fact, their real aim is just to **impose** their own believes [sic], politics, rules **to** the society and make them blind and [...]* <TRCU1126>

(174) *People are so addicted to money in recent days that they can do anything for it, they can **give up from** their relatives, **from** their friends, they can tell lie, [...]* Even it is good or bad,

*I think people will never **give up from** money, they'll always work for it.* <TRCU1044>

(175) *Lastly, students do not make any effort to improve their knowledge, skills or any hobbies to **add** something new **into** their personality.* <TRME3029>

In the case of *look over* in (171), the wrong choice of the preposition might have been triggered by the meanings of this word combination in English [‘survey, inspect, go over’]. In sentences (172) and (173), we see that substitution results in non-standard prepositional verbs – a deviation type seen very often in the data, including **benefit for*, **care on*, **persist on*, **die through*, among others (See Appendix 5 for all prepositional verb types attested in the data). Another instance of preposition substitution resulting in a nonstandard form, yet this time a phrasal-prepositional verb, is seen in (174) – which, unlike in the earlier cases, can be accounted for by L1 interference: The Turkish equivalents of *give up on* (*vazgeçmek*, *umudunu/ümidini kesmek*) require the use of ablative suffix with their objects – the correspondence of which is the preposition *from* in English. Therefore, the preposition *from* was used rather than the correct preposition *on* in this context four times by two learners. As for the sentence (175), the use of *add into* rather than *add to* has been reported to occur across different ESL varieties by, for instance, Edwards and Laporte (2015: 158f) and this unconventional use has been interpreted as an emphasis put on making explicit the direction/movement involved (see also Gilquin and Granger 2011 for the nonstandard uses of *into* in different (learner) varieties).

Just as prepositions and particles, the verbal part of the multi-word verbs has occasionally been substituted by the learner, as exemplified in (176) and (177). In (176) the target verb aimed at by the learners appears to be *adapt to*. This substitution error was attested eight times in the data and it is highly likely that the formal similarity of the two verbs caused the problem. The intended meaning with *donate with* in (177) can be expressed with *equip with*. One possible explanation for this substitution is that the Turkish equivalent of the intended verb in this context shows a formal similarity to the actually used verb in this context – *donate* vs. *donatmak* – the latter can collocate with the noun *knowledge* in Turkish (*bilgi ile donatmak* [lit. trans. ‘equip with knowledge’]).

(176) *As we are human beings of course we **adopt to** a new society subject to constant change and to contribute to bring about these changes. Thus, continuing adaptation is going on.*
<TRKE2011>

(177) *This shows clearly that people are not improved in universities or in other words the ways that improve people’s personality, point of view and ability of practical usage of knowledge are not shown in universities. They are only **donated with** theoretical knowledge* <TRME3030>

Somewhat more creative instances of verb substitution attested in the learner data are exemplified in (178) and (179). In (178) we see that *source from* was used by the learner in the meaning of *result from* – that is, in a context that does not match the definition provided in OALD: ‘to get sth from a particular place’. We could speculate that it might

have arisen from analogy with the targeted verb, the choice of the verbal part triggered by L1 equivalent: *result from* can be translated into Turkish as ‘*kaynaklanmak*’ – the base of which (*kaynak*) equates to *source* in English. As for *final in* in (179), there is no entry of *final* as a verb in dictionaries – neither in COCA nor in GloWbE. The context makes it clear that the intended meaning with this combination by the learner was *end in* or *result in* – a semantically similar verb.

(178) *The first reason which do not give any right to prepare the students to the real life is **sourced from** the state universities in Turkey have very low aid from the government. They do not take enough money to provide [sic] real environment to the students. [...] this problem is not **sourced from** the teachers or the students, the real responsibility is in the government's hands.*
<TRCU1064>

(179) *Of course she cannot give birth to a child on her own. But she spends almost a year in action in this process which **finals in** pain while her partner [...]* <TRCU1115>

In the case of some prepositional verbs, the choice of the entire multi-word verb was inappropriate in the context it was used, as exemplified in the following examples:

(180) *Not much will be achieved by making it illegal, as that will not prevent people from committing [sic] suicide or **appealing to** euthanasia.* <TRCU1075>

(181) *Some sorts of studies have to be carried out with strict cohesion to theoreticality. To name just a few branches which have no other alternatives but to **consort to** a theoretical [sic] based framework of education are philosophy and theology.*
<TRME3022>

(182) *The students are surrounded many unnecessary knowledge. [...] When will the individual **apply to** these [sic] unnecessary knowledge? <TRCU1065>*

In (180), we see that the learner aimed for *turn to* rather than *appeal to*. As already mentioned a couple of times in the earlier sections, Turkish vocabulary is coarser compared to English. If one then only consults (online) bilingual dictionaries where no information is provided on the contextual use of words or semantic distinctions between synonyms or semantically close words, it is inevitable to commit errors as in (180). A quick look at a commonly used Turkish-English online dictionary¹⁴³ is enough to understand the source of the substitution error in (180). As equivalents to one verb in Turkish that would fit the context in (180), namely *başvurmak*, many verbs as equivalents in English are provided, some of which are unrelated. These include *consult*, *appeal*, *apply*, *refer*, *resort (to)*, *look to* and *turn to*. Therefore, it is no surprise to see the learner group under investigation having problems that involve a grasp of the semantic distinctions and overlaps between verbs. In the case of (181), it can be speculated that the learner got confused due to formal similarity between *consult* (the target verb) and *consort* – since these two verbs are not semantically close to one another. In (182) we see an L1-induced unidiomatic use of *apply to*: *apply to (one's) knowledge* is a conventional collocation in Turkish.

¹⁴³ <https://seslisozluk.net>

Semantic prosody is a related notion that is worth of a mention in this context: Occasionally, as expected, native speaker judgments differed on the acceptability of multi-word verbs attested in the data. The judgment on one use of the phrasal verb *bring on* was such an instance; although one NS considered it acceptable, two did not:

(183) *This brings on equality between women and men.*
<TRKE2042>

One of the meanings of the transitive phrasal verb *bring on* provided in Cowie and Mackin (1975) is ‘cause, produce, lead to’ – which is most probably the intended meaning by the learner in sentence (183). Having gone through most of the instances of *bring on* in COCA, we can state that this verb has overwhelmingly negative semantic prosody since it primarily occurs in contexts of war, crisis, stress, (heart) attacks, headaches, pain, etc. Unaware of this fact, the learner in (183) collocated *bring on* with a word that have a positive meaning, i.e. association.

This last point brings us to the category of collocational deviation – the phenomenon of ‘collocability’, already touched on briefly. One fairly common deviation attested in the data is not related to the choice of the multi-word verb itself or one of its constituents but to its collocates. As a result of overlooking collocational restrictions, multi-word verbs are

combined with inappropriate collocates, resulting in unidiomatic combinations. Two instances of the phrasal verb *carry out* from the data can be given here to illustrate this type of deviation:

(184) *This results in the way not **carrying out the usage of their knowledge** in real life. [...] the education programmes completely based on theory, don't give permission for this. Many graduates have difficulties in **carrying out the knowledge**; they learnt it before as theoretical.* <TRCU1069>

(185) *One of its negative dimension is that it is very difficult and impossible to **carry out equality** between sexes in a nation like us which has a patriarchal national structure.* <TRKE2052>

None of the nouns combined with *carry out* by the learners in sentences (184) and (185) were attested with this verb in COCA (see also sentences (61) to (64) in Section 5.2.2 and sentence (107) in Table 23). The most common collocates of *carry out* in COCA are *attack*, *mission*, *order*, *task*, and *duty*. Lacking the contextual knowledge, i.e. range, of this phrasal verb, the learners extended its scope to different contexts; they combined it with unusual collocates such as *knowledge*, *occupation*, *value* or *precautions*. In (184), what the learner aimed at with **carry out the/their knowledge* was ‘use the knowledge’ or ‘put the knowledge into practice’; in (185), ‘achieve equality’. Sentence (107) in Table 23 shows that the learner is not aware of the fact that *precautions* are always *taken* – the set phrase *take precautions*. In a similar vein, the collocational restrictions of *build up* were also overlooked by two other learners, as seen in (186) and (187):

(186) *To prison him, to **build up** many handicaps in front of him in order not to unite with his readers and then after his death to declare him as a hero?* <TRCU1126>

(187) *As a first step, physical features **built up** differences between women and men.* <TRKE2063>

Neither in COCA nor in GloWbE do *handicap* and *difference* co-occur with the phrasal verb *build up*. Its common collocates include *confidence, strength, picture, military, forces, layers, confidence, muscle(s), trust, etc.*

Collocational deviations in the learner language can be taken as further evidence in support of the claims of the learner's reliance on the 'open choice principle' (see Sections 2.1 and 2.3.1). This reliance becomes more dangerous when the L1 of the learners makes too few semantic distinctions compared to the TL: "one-to-many semantic relationship" (Dagut 1977: 227), already discussed in Section 2.2. This is the case at hand with the uses of both *carry out* and *build up*. If one references the earlier-mentioned online dictionary, it is not difficult to see why the learners had problems with the collocational range of these phrasal verbs. As an equivalent to *carry out*, it provides many verbs in Turkish, even a very general one such as *yapmak* (= *do* and *make*). For instance, in Turkish the correct verb in (184), where the learner most probably aimed for 'use the knowledge' or 'put the knowledge into practice' with **carry out the/their knowledge*, would be *uygulamak* or *hayata geçirmek* (the second and seventh meaning provided in the online dictionary, respectively). If we imagined that the learner just wanted to use

another verb instead of *use* and therefore checked for another equivalent of *uygulamak* and *hayata geçirmek*, *carry out* was one of the first options s/he would get. It is believed that a similar thing was the case in the earlier mentioned inappropriate uses of *build up*. In both (186) and (187), the Turkish verb *oluşturmak* would apply to the context, which corresponds to many English verbs, including *form*, *constitute*, *create*, *build up*, *generate*, all of which are provided as ‘equivalents’ in the online dictionary. As mentioned earlier, this online dictionary, like similar ones, provides no other information to the learner to support a verb’s correct use, such as its (in)transitivity, collocational restrictions or common collocates. There are occasional example sentences given with their translations but these examples themselves can indeed trigger collocational deviations in the TL. Therefore, it comes as no surprise why the learners extended the collocational range of these multi-word verbs inappropriately.

Similar collocational deviations were attested also with prepositional verbs and phrasal-prepositional verbs, as exemplified in sentences (188) and (189):

(188) *Now a days [sic], the couples show severe incompatibility as a reason getting divorced when we look at the records. Some couples **apply to law**, because [...]* <TRCU1085>

(189) *If he doesn’t highlight that he couldn’t grasp the subject, he should inform this to his teacher as soon as possible, otherwise the subject grows too big and he can’t **catch up with the lesson**, and finally find the solution in cheating.* <TRKE2051>

The deviation observed in (188) can be accounted for by L1 interference since this combination is a conventional one in Turkish. The inappropriate use in (189) is very close to being accurate. Most probably the difficulty caused by this verb to learners lies in the difficulty of determining the contexts where this verb could not be used since it can collocate with both animate and inanimate objects, e.g. one can *catch up with work, peers, course material or technology* but not with a *lesson* (see Turton and Heaton 1996: 65). In COCA, this verb co-occurs mostly in the context of animate objects (i.e., personal pronouns, friends, competitors, etc.). Although such unusual combinations do not generally fail to communicate the intended meaning, they leave an unidiomatic impression on the reader.

The learner productions exemplified here so far verify the difficulties multi-word verbs cause to the learner group at hand. Despite the difficulties they cause, however, some deviations in the data, i.e., redundant use of prepositions or particles, can be seen as an indicator of a wish on the learners' side to use multi-word verbs. As mentioned earlier, there are instances in the data where a 'simple' (i.e., one-word) verb, would be preferable by a NS yet the learner unsuccessfully employed a multi-word verb (as seen in sentences (140) to (142)). Sentence (190) can be given as a further example for the learner's tendency to opt for a multi-word verb, where NSs would preferably instead use the one-word verb *endure*, rather than *stand together*. We see in (190) that the learner used this intransitive phrasal verb erroneously as a transitive verb.

(190) *Before people were respectful for each other and they **stand** every conditions **together**, but now the power of tolerance of the people is less and less and they hurt each other quickly.*
<TRCU1085>

A word needs to be added here on what Waibel (2007) called “simplified use of phrasal verbs”. What she referred to was the tendency of learners to express their intended meaning in very general words in a context where a different expression would indeed be more fitting, as in *my interest falls down* rather than using the one-word verb *decrease* (p. 137f). According to Waibel, such “simplified” uses occur due to a lack of relevant vocabulary knowledge and their usage is one of the factors contributing to the unnaturalness in learner language. If the use of common, high-frequency verbs with particles rather than less frequent and more explicit one-word verbs can be called a simplification process, then, so, too the use of *stand together* in (190) can be seen as a “simplified” use since the single verb *endure* (the targeted verb) is lower in frequency compared to *stand* (along with other verbs with similar meanings that would also fit the present context, for instance *bear* and *withstand*). A similar case can be seen in (191); rather than using a verb like *promote* to express the intended meaning, the learner opted for a common verb and particle combination, namely *bring on*. As for the phrasal verb *come out* in (108) in Table 23, it exists in English with various meaning senses, one of which is ‘appear, become visible’ but when used in this sense, what can occur in the subject position is rather limited: *the sun, moon, stars* (Cowie & Mackin 1975). Unaware of this

limitation, the learner mistakenly combined *come out* with *health problems*, rather than *appear*, which would probably be the choice of a NS in this context.

(191) [...] *animal testing is considered superior than other clinical researches, and it seems as if it was more scientific. This image can **bring on** doing researches which are not based on any important question, and are entirely unnecessary.* <TRKE2029>

The foregoing has presented an account of the deviation types spotted during the analysis of TICLE. Before ending the section, there are three other things that are worth a mention here. Firstly, the learner variety investigated in the present study displays the lack of register awareness reported in earlier studies (e.g., Gilquin 2011; Gilquin & Paquot 2008; Römer 2009; Mondor 2008; Waibel 2007). Occasionally there are some informal verb choices in the data (e.g., *build up* rather than *create* or *look for* rather than *ask*, *hike up* rather than *lift*) along with some verbs that are not expected in academic or argumentative writing (e.g., *wake up*, *get up*, *sit down*). Since the language of novice native speakers has also been shown to have the same feature – though to a lesser extent – insensitivity to register variation has not been captured under error types.

Secondly, some learner essays include both erroneous and correct uses of a given multi-word verb. For instance, in the case of *grow up*, a learner erroneously used it as a transitive verb twice along with its two other correct uses in the very same essay (see sentence (192) and (60) – the latter is reproduced here as (193):

(192) *This child will **grow up** with the lack of a father; she or he will never have the chance to taste the father love [...] These children will **grow up** being deprived of many things they want.* <TRCU1094>

(193) *Giving a birth to a child is not difficult; however, **growing** him or her **up** as a conscious, educated, healthy and beneficial individual is difficult. A family should have children as many as they can be in charge of and they can afford to **grow up** perfectly.* <TRCU1094>

In a similar fashion, there are instances in the data where a learner used a prepositional verb correctly in one sentence but went on to use it in another without the necessary preposition, with an additional one or with a wrong one, as exemplified in (194) to (196):

(194) *Men were working and earning money for their family. They had jobs and **looked after with** their wives and children. [...] On the other hand, women were staying at home all day, they were cleaning the house, cooking, and if they had children, they were **looking after with** their children. [...] They [women] cook very delicious meals, they clean the house, and they **look after** their children. [...] They should **look after** their children. Although they have jobs and they work, they shouldn't forget that they are wives and mothers, and they should **look after** their husbands and children.* <TRKE2042>

(195) *When we **look at** men, their physical power is stronger than us. [...] When we **look** other countries, sexuality inequality is not so striking as in our country.* <TRCU1137>

(196) *There are three basic, traditional functional universities; giving theoretical knowledge about our jobs in the future, providing to grasp our mental and creative skills and teaching us to*

*prepare ourselves to life. [...] The third aim of universities is to teach students to **prepare** themselves to life. It is true that most universities do not **prepare** students for the real world because that is impossible. However, universities teach pupils to **prepare** themselves to their lives. [...] A person himself learns what life means and **prepare** himself the difficulties that he will be confronted with. Universities, friends in fact even families can not **prepare** us to the real world. <TRME3026>*

The interesting thing with the last example is that the prepositional verb *prepare for* is a part of the essay title the learner chose – namely, topic 10 (‘Most university degrees are theoretical and do not **prepare** students **for** the real world. They are therefore of very little value’). The learner in (196) used *prepare for* correctly once – only in the case of the sequence given in the topic title. In the rest, the learner wrongly combined *prepare* with *to* four times, carrying L1 pattern into English, and in one case, s/he forgot the preposition, indicating that this multi-word verb is indeed not in his/her active vocabulary.

Thirdly, what has also been noticed during the analysis is that a semantic evaluation of the context in which multi-word verbs occur is necessary for the investigation of unidiomatic uses since, as pointed out by Schneider and Zipp (2013), some combinations are standard, i.e., they are listed in dictionaries, making the detection of their unidiomatic uses in a given context difficult by automatic retrieval methods. Occasionally a larger context than sentence level is necessary to judge a given verb’s appropriateness.

To round off this section, on the basis of all of the above findings, one can conveniently propose that the verbs dealt with in this section – namely, phrasal, prepositional and phrasal-prepositional verbs – appear to be a major hurdle for Turkish learners under scrutiny. Their use by the learners shows deviation from the native speaker norms to a great extent. With regard to the deviation types, the broad patterns of our results align well with earlier studies conducted on multi-word verb use by learners with other L1 backgrounds. In TICLE these three verb types were used in contexts that do not match the definitions provided in dictionaries or do not apply to the context; they were used with redundant or inappropriate prepositions and particles; they were combined with unusual collocates – although these unnatural combinations did not always hinder communicating the intended meaning, they contribute to a lack of nativeness by leaving an unidiomatic impression on the reader. In the data there are instances of transitive uses of intransitive multi-word verbs or vice versa, indicating a lack of syntactical knowledge of learners. In addition to these deviations, the learners sometimes creatively formed ‘new’ multi-word verbs based on either analogy with existing L2 patterns – in the case of aspectual phrasal verbs – or direct translations from L1. Although the deviation types have been divided into five categories, it should be noted that it has not been easy to neatly classify the deviations and deal with them entirely separately due to common overlaps between the categories.

The correct and incorrect uses of the very same multi-word verb in an essay side by side raise the question of whether some of the erroneous

uses attested might only result from learners' carelessness rather than their lack of complete vocabulary knowledge (i.e., performance vs. competence errors). Since the essays in the data were not written under time-constraints, this should be a rather slim possibility compared to two other more plausible causes: L1 interference and partial, i.e. deficient, vocabulary knowledge.

L1 interference has revealed itself noticeably in nearly all deviation types. A considerable number of similar deviations were shared by different learners in the data, which can be regarded as a support for the underlying influence of L1. The case endings or postpositions required in L1 equivalents of English verbs were translated into L2 and used as additional prepositions. Redundant uses of particles and prepositions in the data are high in frequency, which is interesting given the fact that these units do not exist in the learners' L1. However, not all redundancy observed in the data can be explained by L1 – especially the creation of 'new' phrasal and prepositional verbs. The results in relation to the processes behind this creativity are in line with earlier studies: intralinguistic analogy with existing patterns and irregularity. Tentatively, one could regard a learner's additional particle and preposition usage as conscious attempts to use them, being aware of these units' prominence in forming multi-word verbs. The existence of formally and semantically similar verbs has occasionally been identified as a source of substitution errors.

As for the unusual combinations, we could say that a combination of the "coarser-gridding" in L1 and incomplete (contextual) knowledge of

the L2 lexicon resulted in the learners' extending of a given (multi-word) verb's collocational range inappropriately – lending support to Hasselgren's (1994: 251) claim: 'spreading' of one word's semantic terrain goes hand in hand with divergence.

All the deviations reported in this section, regardless of their source, demonstrate the learner's partial mastery of the L2 lexicon. Their insufficient or incomplete lexical knowledge prevent them to form and/or produce standard, i.e. native-like, multi-word verbs in appropriate contexts.

6.3 Erroneous and unidiomatic usages in the context of verb-noun collocations

Before presenting the qualitative results of the analysis, a brief discussion regarding the definition of verb-noun collocation adopted in the present study may be in order. Except the verb *to be*, all verb-noun collocations, including the interrupted ones, were extracted from TICLE. The only criterion for the noun in these combinations was that it be abstract in nature, with abstraction relying heavily on its use in a given context (see Sections 3.2.5 and 4.5.2 for more detail). The analysis has revealed 3,380 instances of verb-noun collocations in total. This section provides a detailed analysis of these collocations.

In line with the results reported for phrasal, prepositional and phrasal-prepositional verbs in Section 6.2, a high number of deviations in the use of verb-noun collocations were observed in the data, indicating the

difficulty they cause to the learner group under investigation. Verbs are quite commonly confused and combined with inappropriate collocates – most probably as a result of the scarcity of lexical commonality between Turkish and English due to the typological distance and the “coarser-gridding” in Turkish compared to English, as mentioned earlier. The issues found to occur in the use of prepositions and particles – namely, addition, substitution, omission and unnatural uses – were also identified in the production of verb-noun collocations. In the following the different types of deviations identified in the data will be investigated in detail. Some of the problems attested did not appear in large numbers. Therefore, most of the findings concerning individual collocations should be regarded as preliminary.

As stated in Section 5.4.1, Group I comprises 3,191 verb-noun combinations. Out of 324 types, 160 are listed in the LTP dictionary (see the verb-noun collocations in bold in Appendix 8a). That is, these verb-noun combinations are conventional and their constituents are strongly correlated. The rest of the combinations produced by the learners are either very common verb-noun combinations, components of which have low restricted collocational relationships, and thus are not listed in the dictionary (e.g., *have importance*, *give feedback*, *find information/time*, *end a marriage*, etc.), or are idiosyncratic or deviant in some way from typical native speaker production.

The deviations identified in the data in the context of verb-noun combinations seem to fall into four (sometimes overlapping) categories –

the first two being grammatical and the latter two lexical deviations: 1) redundant uses of prepositions, determiners as well as plural markers, 2) omission of prepositions, determiners or a noun collocates, 3) substitution errors, i.e., an existing collocation as a whole is not appropriate in a given context since its dictionary definition does not match the context or one of the constituents of the collocation is inappropriate, 4) unnatural (i.e., unconventional) uses of verb-noun collocations, i.e. use of an existing collocation in a context that does not accord closely with NS preferences (idiomaticity issues), as well as use of atypical, unrecorded (stretched) verb-noun combination where another (unstretched) verb would be more appropriate. Table 24 provides an overview of these deviation categories, accompanied by examples from the data. What follows is a detailed discussion of each deviation type, with an attempt to find plausible explanations for the deviations attested.

6.3 Erroneous and unidiomatic usages in the context of verb-noun collocations

Table 24. Types of error and unidiomatic usage attested in the production of verb-noun collocations in TICLE

Types of error and unidiomatic use		Examples	
1. addition	determiner	(197)	<i>Because although we try to put a distance between us, we need each other. <TRCU1038></i>
	plural marker	(198)	<i>Machines do the works and housewives get tired. <TRCU1018></i>
	preposition	(199)	<i>This means we face to a big problem that time. <TRKE2054></i>
2. omission	determiner	(200)	<i>Firstly, it influences psychology of the people. If someone does not have enough money to live, she/he has tendency to do bad events. <TRCU1144></i>
	plural marker	(201)	<i>Because of these, I support that euthanasia is not a good issue to take side on. <TRCU1159></i>
	preposition (Group II)	(202)	<i>She can get benefit all the good things in the world. <TRCU1056> ('get benefit from')</i>
	(Group III)	(203)	<i>If the society want [sic] their children to gain the social values, the [sic] must take account some other effects. <TRKE2046> ('take into account')</i>
3. substitution	verb (Group I)	(204)	<i>In eastern regions of Turkey, families take people's attention by the great number of having children. <TRCU1154> ('grab people's attention')</i>

Chapter 6: Multi-word verbs in learner writing – a qualitative approach

Table 24 (cont.). Types of error and unidiomatic usage attested in the production of verb-noun collocations in TICLE

Types of error and unidiomatic use		Examples
3. substitution	verb (Group III)	(205) <i>Moreover, the radiation created by nuclear power can not be taken under the control easily. <TRCU1133> ('get under control')</i>
	noun collocate	(206) <i>In order to decrease these reasons both students and teachers must give equal important to both theoretical [sic] knowledge and application. <TRME3015> ('to give importance to N')</i>
		(207) <i>Also both sex have an opportunity to be a president or a prime minister, there is no law that prevents men and women from getting these stages. <TRKE2066> ('jobs/positions')</i>
	preposition (Group II)	(208) <i>When we have a general look in the universe we can witness to see that every living creature is created with its opposite, female and male. <TRKE2028></i>
	(Group III)	(209) <i>There are also social issues to be taken to consideration; for example [...] <TRCU1026></i>
4. unnatural uses	entire collocation	(210) <i>On the other hand they don't take care the increasing number of mosquitos which can not be stopped if frogs are killed continuously. <TRKE2054> ('care about')</i>
	unusual V+N collocation	(211) <i>[...] especially fathers blame mothers when the children do something wrong but when they do something right or good they say he is my child or she is my child. They take good sides and give bad sides to mothers. <TRKE2031> ('attribute the good traits (to themselves)')</i>
		(212) <i>Money also gives harm to people's relationships. <TRCU1043></i>
	choice of V+N col. does not accord with NS preferences	(213) <i>if that means becoming pregnant and creating a life, then [...] <TRCU1092> ('get pregnant')</i>

The first category of deviations in the use of verb-noun collocations includes additional uses that occur in three subtypes. The first type comprises additional usage of determiners – especially of articles: In the data, there are some instances where noncount abstract nouns, which usually have no article when used generically (Quirk et al. 1985: 274, 286), occur with an additional article, as in *do *the household jobs*. Here are some further examples of deviations involving an additional article:

(214) *Women feel guilty regretful **having *the abortion**.*
<TRCU1095>

(215) *Because they do not **have *a homework** or an exam related with the repair of a machine* <TRME3018>

(216) *Women grow the new generation, determined the taboos, solved and knew the secrets of nature. Therefore, women **had *a big respect *on the nature**.* <TRKE2002>

Similar uses to the latter example include *get *a power, have *a confidence, have *a knowledge, get *a success, give *a big pain*, to name just a few.

In addition to the definite and indefinite article, other determiners were inappropriately added to uncountable, abstract nouns in the data:

(217) *Most people who **have *many necessary knowledge** migrates since they are not valued in that country.* <TRCU1066>

The article systems of Turkish and English differ: In contrast to English, which has two articles, use of which depends on the feature [\pm definite],

Turkish has only one indefinite article – unstressed *bir*¹⁴⁴ (Kornfilt 1997) and the range of its distribution differs from that of the indefinite article in English. Whereas singular countable nouns must take an article in English, plural countable nouns and mass nouns are used with zero article if they are indefinite (Goad & White 2009: 203). In Turkish, on the other hand, the articles can be omitted in certain contexts and “[t]he interpretation of a bare noun as definite or indefinite depends on a number of factors, including word order and case marking” (ibid.). Trenkic (2009) has observed that omission of articles in English in contexts where their use is obligatory is a pervasive feature of the speech of learners whose L1 lacks articles. Given the lack of a definite article in Turkish, the additional use of articles in verb-noun collocations are even more interesting. When learning English, the learners are introduced to the articles at a very early stage. In the following years of study, learners are repeatedly reminded of the contexts in which articles are required. The learners’ use of articles in non-required contexts may be a type of ‘hypercorrection’. While trying to integrate these L2-specific units into their language use, the learners may, as stated by Dose-Heidelmayer and Götz (2016: 242), “overshoot the mark”.

The phenomenon of addition is also attested in the use of the plural marker in the data; uncountable abstract and mass nouns, such as *harm*, *homework*, *information* and *knowledge*¹⁴⁵, were attested with the plural

¹⁴⁴ Turkish does not distinguish the indefinite article and the numeral ‘one’.

¹⁴⁵ Some of these abstract nouns can be used both countably and uncountably (e.g., *knowledge*) but their plural use in the data was judged to be incorrect by the NSs, e.g.

marker in the essays of more than one learner in the data, as in (217) and (218). This deviation type can be directly related to the learners' L1: Turkish makes no clear distinction between count and mass nouns (Göksel & Kerslake 2005: 146). That is, *furniture*, for instance, which behaves grammatically as a mass noun in English and does not inflect number, can appear with the plural marker in Turkish. The difficulty caused by this crosslinguistic difference for the L1 speakers of Turkish has been shown by, for instance, Yazıcı and Irtes (2014). In line with the current study, they attested plural markers used in combination with the uncountable nouns *homework*, *information*, *help*, *hair* and *happiness* in their data (Yazıcı & Irtes 2014: 163).

(218) *The problem with this argument becomes evident when the patient is not able to present a desicion [sic], whether he is unconscious or **has** other **inabilities** of communication or thought processes.* <TRCU1173>

The plural marker was also attested in NPs in fixed verb-noun collocations. For instance, the noun collocates in *take place* and *commit suicide* were produced in the plural form by three and two learners, respectively. Sentences (219) to (221) illustrate the same deviation – the latter includes an additional quantifier.

(219) *Obviously, learning **takes** ***places** in many ways and forms.*
<TRME3016>

use one's knowledges. In COCA, the only possessive pronoun used with *knowledges* is *their* and *their knowledges* occurs only four times as opposed to *knowledges* alone (that occurs 245 times).

(220) [...] *in this way the teacher will **take** *parts in the development of the students' capacities negatively,* [...] <TRKE2018>

(221) *They **spend many** *times watching TV.* <TRCU1020>

Such incorrect uses of the plural marker in the data may be a performance error (such as a typing mistake) but recurrent wrong usage in a single learner's essay renders such an interpretation unlikely – at least for some of the redundant instances encountered here.

The next type of deviation in the category of addition is the redundant usage of prepositions in verb-noun collocations. In this subtype, the lexical elements are correct and only the removal of the preposition is required to make the verb-noun collocation acceptable. As reported in Section 6.2, additional prepositions were produced in the construction of phrasal and prepositional verbs – albeit not all, most instances could be traced back to the learner's L1. In the data, quite a few one-word verbs that should directly be followed by their objects, forming verb-noun collocations, were instead interrupted by a superfluous preposition, as seen in the example (199) in Table 24. *Face a problem* was attested 11 times in the data and one of these instances included *to* between the collocator and the collocate – *face *to a big problem*. One possible explanation for the redundant use of the preposition *to* in this and similar instances is that it was produced as a counterpart of the dative ending in Turkish. Similar instances are plentiful in the data: *begin to one's job, study to N, attend to the classes/to the school, cause to*

death, face to difficulties, etc. *To* is not the only preposition used redundantly in these constructions; there are instances where the learners carry L1 patterns into their L2, using other prepositions. For instance, in Turkish the literal equivalent of *stay a secret* would necessitate the use of postposition *olarak* – the English equivalent of which is *as* – and in (222) we hence see that the learner produced this construction with the additional *as*. The additional *in* in *run in a marathon*, however, cannot be accounted for by the carrying-over L1 patterns since the literal equivalent of this collocation in Turkish – *marathon kořmak* – is entirely the same; the NP does not require a suffix. The formation of *suffer to death* in (223)¹⁴⁶ might have been triggered by analogy to existing, formally similar collocation *stone to death*.

(222) *Hense [sic], these bad aims against a nation or a country can not get clear and stay as a dangerous secret.* <TRKE2064>

(223) *Therefore, i think if a couple cannot afford to meet all an infant's physical, emotional, psychological needs thoroughly, they ought to in advance prefer to make use of abortion so that the baby subsequent to the birth shouldn't suffer to death at all.* <TRCU1155>

A more expected error type to be committed by the learner group under investigation with regard to prepositional use is indeed their omission due to their absence in Turkish. This expectation is borne out in the data; there are some instances where obligatory prepositions in verb-

¹⁴⁶ One of the NSs suggested *suffer and die* and *suffer pain/death* as alternatives for this unconventional collocation, stating that one suffers while s/he dies, not suffers to die.

noun constructions were left out by the learners – both in Group II and III verb-noun collocations (see (202) and (203) in Table 24 and the following examples (224) and (225)). However, the number of such instances is far lower compared to additional uses of prepositions. In most cases, the missing preposition is *to* – the very same preposition used redundantly in many cases:

(224) *Whatever comes your mind, you get information about it*
[...] <TRCU1013>

(225) *Do we learn real life in the universities or not?*
<TRCU1066>

In (224), the learner's omitting the preposition *to* is remarkable since the Turkish equivalent of this collocation, which is basically the word-for-word rendering (namely, *aklına gelmek*), necessitates the dative suffix (-E) – the functional equivalent of *to* – to be attached to NP [=akıl + -a]. Also remarkable is the missing *to* in *come to an understanding* and in *pay attention to* (which was attested six times without the preposition). If the learners had relied on their L1 during the production of these collocations, they would not have omitted the necessary preposition. As shown in earlier studies (e.g., Yamashita & Jiang 2010; Wang 2016), L2 learners do not always rely on their L1 during the acquisition and use of L2 collocations. The missing preposition *about* in (225) is, on the other hand, less remarkable, i.e. more expected, given the Turkish equivalent of the collocation – which does not necessitate a preposition (namely, *hayati öğrenmek*).

As mentioned earlier in this section, Turkish lacks the definite article and the indefinite article, which is not differentiated from the numeral ‘one’, has a different distribution compared to English. Considering these facts, omission of both the definite and the indefinite article in English productions of Turkish speakers was expected. As anticipated, the data reveals ample instances where the indefinite article was left out. For instance, *get a divorce* was used eight times by seven learners without the article, as in (226).

(226) *In the old days, while wives and husbands respected each other, today, they always quarrel because of unnecessary reasons and then they apply to court in order to **get divorce**.*
<TRCU1165>

In seven instances, the verb-noun collocation *commit a crime* was used without the indefinite article by four learners. It is interesting to note that *commit suicide* was produced five times by four learners with the indefinite article. The case of *commit *a suicide* might arise from analogy with *commit a crime* and perhaps also other conventional collocations such as *commit a robbery/theft*.

In some other instances, the definite article was missing – a more common deviation type than the omission of indefinite article. Two illustrative examples are displayed in (227) and (228):

(227) [...] *the important thing is **minimizing opportunity** to cheat by means of some prevention.* <TRKE2050>

(228) *This inequality in society occurs [sic] by **coming to fore** that a man is a hitting, breaking person as his physical [sic] feature.* <TRCU1157>

One other repeatedly missing unit in the verb-noun collocations extracted from the data is the plural marker, as seen in (229)-(231). Unlike in English, which is a mass/count language, in which the use of plural morphology is required in the context of numerals greater than one, and also zero and decimals (e.g., *zero books, 1,5 books*), in Turkish, numerals greater than one obligatorily combine with bare nouns (Alexiadou 2019: 124). The cases in (229) and (230) might be accounted for by this crosslinguistic difference. However, the omission of the plural marker with the noun collocates in *take sides* (see 201 and 231), which always occurs in plural form meaning ‘to express support for somebody in an argument’ (OALD), indicates the learner’s incomplete lexical knowledge.

(229) *First of all, they will **have** all **feature** they need, [...]*
<TRCU1070>

(230) *S/he can go best schools which **has much opportunity**, and qualified education.* <TRCU1056>

(231) *The Media tries to bridge the gap between citizens and administrators. It functions as a messenger. [...] While working it never **takes side**.* <TRKE2001>

These results are in line with the findings of Yazıcı and Irtes (2014), who also attested omission of the plural marker in the writings of Turkish EFL learners at three different proficiency levels.

Before we move on to the third category of deviations, there is one more thing related to the notion of omission that is worth mentioning. In one single case, a learner left out the noun collocate of a collocation:

(232) *He tried to rob a bank but he could not achieve his goal.*
[...] *Some people are **committing** for they really need, but some*
[...] <TRCU1107>

This could simply be a performance error or may be another instance of confusion between the transitive and intransitive verbs (see sentences (156) to (160) in Section 6.2). The use of intransitive verb *grow* as a transitive one in (233) lends support to the latter assumption.

(233) *A housewife does cleaning, washing, ironing, cooking and **growing child**.* <TRKE2053> (target: 'raise a child')

Thus far the deviations reported included either an additional element or omission of one. The third deviation type is the most common one in the data and it involves the choice of inappropriate units in collocations – the so-called substitution errors. Although there are instances where verb-noun collocations as a whole unit do not fit the context they were used in, it is mostly one constituent of a given collocation that is chosen inappropriately, particularly verbs. In sentence (234), for instance, the only inappropriate unit is the verb: *Give* was used instead of *put* in *put an end to* [one's life/pain]. This inappropriate choice can be directly attributable to the influence of the collocation's L1 equivalent: *bir şeye* [birinin acısına/hayatına] *son vermek* [lit.= 'give end to sth']. This confusion was experienced by three learners in the data. In (235) we see two inappropriate verb choices. The learner first produced *give* instead

of *make* in combination with the noun collocate *effort*, and then *take a place* where *gain a place* or *work hard for a place* would have been more appropriate for the intended meaning by the learner. In both cases it is not possible to claim that L1 exerted influence on the inappropriate verb choices.

(234) *Firstly, the aim of this method is to **give an end to** patients' pain.* <TRCU1174>

(235) *The education in university is important for many people all around the world and so many people are **giving effort to take a place** in university classroom* <TRCU1072>

Below a range of typical examples from the data is presented. They display instances in which an inappropriate simple (i.e., one-word) verb needs to be replaced by another (simple) verb in order to be (more) acceptable (the first number given in the brackets indicate the number of occurrences, the second the number of learners; singular and plural occurrences in the data have been merged). The suggestions for the target verbs were provided by the native speaker judges.

<i>do an operation</i> (1, 1L)	(target: 'perform')
<i>do euthanasia</i> (8, 6L),	(target: 'perform')
<i>do abortion</i> (6, 4L)	(target: 'perform')
<i>feed one's needs</i> (1, 1L)	(target: 'satisfy')
<i>get (a) power</i> (5, 4L)	(target: 'come into power')
<i>get the revenge</i> (1, 1L)	(target: 'take')
<i>get freedom</i> (4, 4L)	(target: 'gain')
<i>get (a) success</i> (2, 2L)	(target: 'achieve')
<i>give (sb) knowledge</i> (20, 13L)	(target: 'provide/offer/pass on')

6.3 Erroneous and unidiomatic usages in the context of verb-noun collocations

<i>give a decision</i> ¹⁴⁷ (18, 15L)	(target: ‘take/make’)
<i>give effort</i> (4, 3L)	(target: ‘make’)
<i>give experience</i> (3, 3L)	(target: ‘gain/have’)
<i>grow a child</i> (3, 3L)	(target: ‘raise’)
<i>have a mark</i> (3, 3L)	(target: ‘achieve/get’)
<i>know knowledge</i> (2, 2L)	(target: ‘have’)
<i>make a project</i> (3, 3L)	(target: ‘do a project’)
<i>make abortion</i> (10, 2L)	(target: ‘perform’)
<i>make euthanasia</i> (3, 2L)	(target: ‘perform’)
<i>make an operation</i> (3,3L)	(target: ‘perform/carry out’)
<i>make laws</i> (5, 4L)	(target: ‘draft/put into/sign off’)
<i>make an invention</i> (6, 6L)	(target: ‘come up with, invent’)
<i>make a marriage</i> (3, 2L)	(target: ‘enter into’)
<i>meet a problem</i> (4, 4L)	(target: ‘encounter’)
<i>perform knowledge</i> (2,1L)	(target: ‘exhibit, show, demonstrate’)
<i>practice knowledge</i> (6, 4L)	(target: ‘apply/put into practice’)
<i>provide equality</i> (3, 3L)	(target: ‘ensure’)
<i>spend effort</i> (9, 8L)	(target: ‘exert’, ‘put in effort’, ‘give’)
<i>take a mark</i> (9, 6L)	(target: ‘achieve/get’)
<i>take a degree</i> (7, 3L)	(target: ‘achieve/get/have’)
<i>use (one’s) right</i> (12, 8L)	(target: ‘claim’)
<i>win an exam(ination)</i> (5, 4L)	(target: ‘pass an exam’)

A note of caution should be added here to say that not all the collocations where the verb was judged deviant are deviant as such. In some instances a verb-noun collocation exists and is acceptable yet only in a certain context. For instance, *have a role* exists with a theatrical meaning but in the data this combination was used 20 times by 19 learners out of the theatre context where *play a role* would have been more appropriate. Here are some of these uses:

¹⁴⁷ In one instance the noun collocate was erroneously written as ‘decesion’ by the learner TRKE2019.

(236) *These inventions **have an important role** on people's life, [...]* <TRCU1017>

(237) *On the other hand, Media **has a big role** on our family structure.* <TRKE2012>

(238) *All around the world women **have almost the same and difficult role** as a mother, as a spouse and [...]* <TRKE2005>

Similarly, the phrases *take care* and *take care about* N exist but their meanings do not fit the contexts in (239), nor in (210) (see Table 24) where the learners apparently targeted *care about*.¹⁴⁸

(239) *They [rich people] **don't think about their family**. They **don't take care about them*** <TRCU1109>

These examples demonstrate that formal and semantic similarity between verbs in the TL is one factor causing substitution errors in IL. Semantically related nouns may also influence the verb choice of learners. For instance, the inappropriate verb choice in (240) may result from the fact that the verb *cover* can collocate with semantically related nouns to *lesson*, such as *subject* and *course material*.

(240) *Also, the lecturer **covers these lessons** ignorantly; they **do not give any effort to make lessons more bearable** [...]* <TRME3003>

Occasionally, an inappropriate choice of a collocation combined with poor vocabulary choices in other elements in a given sentence failed to communicate a clear meaning; NSs could not construe the intended

¹⁴⁸ The formal similarity of *take care of*, *care about* and *care for* has been reported to pose a general problem for EFL learners (see Turton and Heaton 1996: 63).

meaning. This is the case in (241), where it is apparent from a larger context than a sentence that the learner tried to render the fact that women appear in every aspect/facet of life, doing all kinds of jobs. Although the context does not match the dictionary definitions of *take place*, its direct L1 translation (namely, *yer almak*) fits the context in Turkish. The targeted verb in English is *appear*.

(241) *Women start to **take place at each type of life**. They work outside the home. They show their achievements at different skills. By this way, people understand that both men and women have the same skills.* <TRKE2034> (target: ‘appear in every aspects/facets of life’)

As the above examples reveal, whereas some of the verbs in the data were used inappropriately in combination with only one particular collocate (e.g., such as the confusion of *win* for *pass* in *pass an exam*), some are involved in various combinations, e.g., *do euthanasia/abortion/operation*, *close the telephone* (with the meaning of ‘hang up’) or *close one’s hair/faces/bodies* (with the meaning of ‘cover’), *give experience/effort/decision/knowledge*; *take knowledge/attention/a mark/revenge*, *make laws/abortion/euthanasia/marriage/operation*, *have a mark/education/a role*, etc.

The findings of the present study converge with previous studies that showed that the EFL learners display a tendency to rely on general, ‘easy’ verbs that can be employed in a large number of contexts rather than more specific ones in collocations (see Chapter 2). For instance,

the learners preferred using *do* and *make* over *perform* or *carry out*¹⁴⁹ to collocate with the nouns *operation*, *euthanasia* and *abortion*. In a similar vein, as mentioned in Section 5.4.3, *give* was chosen to combine with *roles* rather than *assign*, *give* rather than *devote* to combine with *time*, and *take* rather than *grab* to combine with *attention*. These results do not seem to confirm Nesselhauf's observation: The researcher reported that there were only a few cases in her data where a more specific, somewhat less common verb was not available to learners (2005: 86). The reliance on general verbs in TICLE is in contrast strong. The general high frequency of these verbs in the data can be seen as a confirmation of the high reliance on them (see Section 5.4.2 and 5.4.3).

In the context of high-frequency verbs, what also stands out in the data is the common use of stretched verb constructions. As mentioned in Section 5.3.3, learners produced stretched verb constructions where a NS would prefer their one-word, unstretched corresponding verbs – see, for instance, the example (242). *Make* was particularly common in the formation of these constructions, e.g. *make torture* (2, 1L) instead of *torture*, *make an invention* (6, 6L) for *invent*, *make an interview* (1) for *interview*, *make a relation between N* (2, 1L) instead of *relate*, *make a behaviour* (1) instead of *behave*, *make a quarrel* (1) for *quarrel*, *make*

¹⁴⁹ As mentioned in Section 6.2, the collocational use of *carry out* is problematic – it was frequently combined with unusual collocates. Only two learners combined it with *euthanasia* – three times in total.

a comment (1) for *comment*, *make a judgment* (3, 2L) for *judge*, *make a classification* (1) for *classify*.

(242) (*Recently, the rate of divorce is **getting increase** in Turkey in the west countries.* <TRCU1165> (target: ‘increasing’)

Some of the stretched verb-noun constructions formed by means of high-frequency verbs are considered inappropriate by the NSs (e.g., *get inheritance*, *make practice* and *make an invention*, *give harm* and *do harm*). This leads us to the next observation: the high number of inappropriate uses of collocations that involve common verbs.

Despite the learners’ reliance on high-frequency verbs – the ones that are used in forming delexical verb-noun collocations (namely, *take*, *do*, *make*, *get*, *have*, and *give*) – the learners under investigation committed errors mostly in their use. These verbs were frequently confused with one another. This confusion has been widely noted in ELT materials (e.g. Turton & Heaton 1996: 104, 208) and also reported in earlier works conducted on language of learners with various L1 backgrounds (e.g., Lennon 1996; Nesselhauf 2005; Källzkvist 1999; Men 2015; Wang 2016; Altenberg & Granger 2001; see Chapter 2 for more on this topic). For instance, in (243) we see that the learner inappropriately produced *have* twice where *get* would have been more appropriate:

(243) *But as we can see only the rich people's children can **have education** with good conditions, others still **have education** in an ordinary school education is with money.* <TRCU1039>

As noted in Section 5.4.3, *make* and *do* – two of the most common light verbs in English – were confused in some instances, as exemplified in (244) and (245):

(244) *But in Engineering Department students **make** every **pro-**
jects on a paper.* <TRCU1072>

(245) *Teachers say how to present a lesson. **make** **exercises**.
draw attention of students and sometimes we **make** **presenta-**
tions in our classes.* <TRCU1071>

Make a project was attested three times in the data – all in the context of education as in (244) – and the NS judges, who regarded this collocation as inappropriate in the context where it was used, provided as alternatives *do a project* and *complete a project*. In a similar manner, two other learners produced *make exercises*, where *do exercises* would be the appropriate choice – see the sentence (245). There are other hapax legomena in the data where we see this confusion: *make one's job*, *make housework*, *make business*, just to name a few – the confusion in the latter case might result from the fact that *make* is frequently used in expressions related to *money*, such as *make profit*, *make a lot of money*.¹⁵⁰

Similar confusion between high-frequency verbs in delexical verb-noun collocations is abundant in the data. *Take*, *give* and *get* are the most

¹⁵⁰ Although *do (a/some) business* occurs a lot more often in COCA (9,267 hits in total), *make (a/some) business* is also fairly common (548 hits), which raises the question of whether one should really consider it deviant.

susceptible to deviation. *Give* was consistently used instead of *make* in combination with *effort* and *decision* and instead of *gain* in combination with *experience*. It was also substituted for *put* in the collocation *put an end to*. Occasionally *make* was used instead of *give*, as in *make presentations* in the earlier given example (245). *Take* was recurrently used in combination with *knowledge* and *degree* where the targeted verb was indeed either *give*, *receive*, *gain* or *have*. Similarly, it was frequently collocated with the nouns *education* and *mark* where *get* would have been the appropriate verb choice. Most instances of *get* in verb-noun constructions were judged inappropriate and corrected with various verbs depending on the context; these verbs include *gain*, *receive*, *attain*, *have*, *exercise*, *fight for*, *claim*, *take away* and *deny* – the last five corrections were for the instances where *get* was combined with the noun *right*.

Collocational deviations in the data are not restricted to high-frequency verbs. As the above examples reveal, there are a considerable number of instances where less common and more specific verbs are involved in substitution errors. Despite their lower frequency, most of these verbs are introduced fairly early in ELT classes. The verbs used inappropriately by more than one learner include *reach*, *achieve*, *come*, *grow*, *attend*, *face*, *enter*, *catch*, *become*, *perform* and *use*. The verbs not produced although they would have been appropriate by more than one learner include *put*, *require*, *provide*, *receive*, *gain*, *raise*, *achieve*, *perform*, *carry out*, *claim*, *enter into*, *apply* and *exert*. Nesselhauf (2005: 87f) has pointed out that less common verbs, which she found to cause

German-speaking learners greater difficulties, have not been the focus in learner collocation studies. The results of the present study conform to the suggestion of Nesselhauf not to focus only on the use of high-frequency verbs of learners in collocation studies. There are consistent misuses in the data that do not involve a high-frequency verb. Collocations that involve *reach* and *achieve* can be given as an example in this context. *Reach* occurs with 19 noun collocates¹⁵¹ in TICLE: *information* (3, 2L), *knowledge* (3, 3L), *data* (1), *aim* (8, 6L), *profession* (1), *point* (3, 3L), *competence* (1), *goal* (3, 3L), *level* (4, 4L), *result* (1), *happiness* (1), *position* (3, 3L), *decision* (1), *solution* (2, 2L), *agreement* (1), *conclusion* (4, 4L), *success* (1), *climax* (1), and *youth* (1). The first five of these nouns (namely, *information*, *knowledge*, *data*, *aim*, and *profession*) were judged to be inappropriately collocated with the verb *reach* and the next two – *result* and *happiness* – were suggested to be better combined with *achieve* and *find*, respectively. Although *reach a point* is an appropriate collocation as such, in one instance the verb was substituted for *make* in the context where it was used:

(246) The **point** that I want to **reach** is the fact that [...] <TRME3028>

The verb *reach* has already been reported to pose problems to other learner groups with different L1s. For instance, Nesselhauf (2005: 78 and passim) observed that it was consistently used instead of *achieve*

¹⁵¹ There are five more nouns *reach* co-occur but they are not treated in this study since they are not abstract: **publishes* (target: ‘publications’), *daily newspapers*, *the water*, *money* and *stair* – all but the latter were preceded by an additional *to*: *reach to* + NP.

by German learners, collocating with *aim*. The same collocation (*reach an aim*) was produced eight times by six learners in TICLE. The NSs consulted considered this verb-noun combination deviant and provided *achieve* and *work towards* as appropriate verbs to combine with *aim*. While the latter alternative was not attested in combination with the noun *aim* in the analysed data, the former (namely, *achieve an aim*) was attested only twice. COCA results confirm the native speaker judgments: *Reach an aim* was attested only 13 times in COCA, as opposed to 236 hits attained for *achieve an aim* and 1,816 hits for *reach a goal*. The latter was produced only three times by three learners in TICLE. Interestingly, the verb *reach* is listed in the LTP dictionary as one of the collocators of the noun *aim*. The fact that the nouns *aim* and *goal* can be collocated with the Turkish equivalents of both *reach* and *achieve* interchangeably may explain the confusion experienced by the learners in their use. However, *reach* was substituted for other verbs in TICLE: e.g., for *gain*, *access* and *acquire* in *reach knowledge* and for *find* in *reach information*. In the case of *reach a profession* in (247), *follow* or *have a profession* would have been more appropriate.

(247) *Again oppositions claim that their working as a grocer not as an engineer is not the fault of university degrees but individuals in the aspect of their inability of using their degrees and knowledge in order to **reach a profession** of their own.*
<TRME3030>

Achieve proved nearly as problematic as *reach* for the observed learners. Quite a few of its noun collocates were judged as inappropriate,

e.g., *difficulty* (targeted verb: ‘overcome’), *acquisition* (target: ‘foster/enable’), *comfort* (target: ‘find’), *duty* (target: ‘fulfill’), *job* (target: ‘handle/manage’), among others. The learners’ L1 cannot always explain this ‘spread’ to cover the semantic area of other verbs (see Haselgren 1994: 251). If we take but one example, in the case of the erroneous collocation *achieve a duty*, it becomes apparent that the learner experienced a problem in distinguishing semantically related verbs. Both *achieve* and *fulfill* share the meaning of ‘succeed/success’ but they differ from each other in the sense that the latter denotes ‘doing something that is required or necessary’ – which is the sense necessitated by the noun collocates ‘duty’. The findings of *reach* and *achieve* taken together demonstrate the partial acquisition of verb semantics by the learners. This finding tallies well with Men’s (2015) findings.

Another problematic verb that is worth mentioning is *perform*. Out of the 16 noun collocates with which it was attested in the data, eight were inappropriate. In other words, the verb *perform* was inappropriately combined with eight nouns (namely, *capital punishment*, *capacity*, *ability*, *behavior*, *belief*, *responsibility*, *education* and *knowledge*). These deviations in the use of *perform* might be explained by the fact that it can be translated into Turkish with various verbs, including *sahnelemek*, *çalışmak*, *icra etmek*, *performans sergilemek* and even a very general one, namely, *yapmak* (=do/make). Its semantic web in Turkish covers a wide range and overlaps with some other verbs. This incongruency in lexical gridding explains erroneous combinations

such as *perform a (childish) behavior* or *perform one's knowledge* – both of which are conventional in Turkish (*sergilemek*).

Thus far, all the substitutional errors reported have been related to verb choices. In the data there are also inappropriate choices of the noun collocate in verb-noun collocations. Their number is yet a lot lower compared to verb substitution errors. As shown in Table 24, there are two types of inappropriate noun choices: While in the first type only the noun form is inappropriate (e.g., confusion of *affect* and *effect*)¹⁵², in the second type the produced noun by the learner is derivationally not related to the target one, such as the confusion between *stage* and *position/job* in sentence (207) in Table 24. In some instances, the confused words – albeit not derivationally related – are semantically related. For instance, it is apparent from the context in (248) that the learner targeted *household* rather than *homework*. Similarly, *knowledge* and *information* have been confused in combinations with *give* and *take* – most probably because these two nouns have one translational equivalent in Turkish. In some cases, it looks like formal similarity of words caused the confusion, as seen in (249). A more general reason for these substitutions is, however, simply insufficient or deficient lexical knowledge (as seen in sentences (250) to (253)).

¹⁵² *Affect* and *effect* are the most commonly substituted word pair, which are derivationally related, in the data – they were confused 11 times by 5 learners. This is a common error of L2 learners, as noted by Turton and Heaton (1996: 112).

(248) *After dinner, she must wash the dishes. After that, she must start to do ironing. [...] The woman continuously [sic] **does homework**.* <TRKE2003>

(249) *If the women don't obey the rules determined by the leader of the family **get a trashing**. A newspaper says that %83 percent of the women in our country **get a trashing**.*
<TRKE2039> (target: 'get a thrashing')

(250) *An important reason why there are many unemployed people graduated from university is that they don't **have enough feature** to find a good job.* <TRCU1070> (target: 'skills')

Further examples of substitution errors involving a derivationally related inappropriate noun choice are:

(251) *[...] when a patient has great pain he may **lose his conscious** at that time [...] It can be wanted unconsciously by the patient who **lose [sic] his conscious** because of his pain.*
<TRCU1140> (target: 'to lose his consciousness')

(252) *Some people were thinking that this woman should have been killed to **end her suffer** [...] No one can decide one's death; even if he wants to **end other one's suffer**.* <TRCU1166> (target: 'to end her suffering'/'end another one's suffering')

(253) *To be able to be successful, one should **put into practical** what he learns but [...]* <TRME3012> (target: 'put into practice')

It is interesting to note in this context that in some instances such deviations occur side by side with appropriate uses of the very same collocation. For instance, the learner in (253) produced *put into practice* correctly three times in the very same essay.

One other thing to note is that some nouns produced are not entirely wrong but native speakers noted that another noun would have been more common, e.g, rather than saying *getting high results* in an exam, it would be more common to say *get high grades* (in AmE), *get high marks* (in BrE) or *get high test scores*.

The next substitution error type attested in the use of verb-noun collocations is relevant to Group II and III: substitution of prepositions. As reported in earlier sections, the learner group under investigation has a general problem in selecting appropriate prepositions. As the sentences (254) to (257) reveal, the deviation can be rectified by changing only the preposition in the verb-noun collocation (e.g., *pay attention *on/*in* to *pay attention to* or *have respect *on/*in* to *have respect for*). In some instances the inappropriate choice of preposition might be triggered by analogy to existing, semantically related verb. For instance, *take care for someone* in (258) might have been produced in analogy to *care for someone* –confusion they cause to L2 learners have already been noted in Turton & Heaton (1996: 63).

(254) we should *pay attention *on* importance of human life too.
<TRKE2032>

(255) *Therefore, they do not thoroughly pay their attention *in that kind of course.* <TRME3003>

(256) *Therefore, women had a big respect *on the nature.* <TRKE2002>

(257) *Some students are really lazy. They do not have respect *in themselves* <TRKE2047>

(258) *They are also supposed to do all houseworks and to **take care** *for children <TRKE2005>*

Further examples of preposition substitution include *play a role *on/*to, come *in one's mind, come *in a point, have a look *in, take care *about* and *take part *at*, among others. Some deviations are shared by the learners; for instance, *take *to consideration* was produced by two different learners. Yet, there are no patterns as to which prepositions are confused particularly often. The same learner occasionally uses a given verb-noun collocation with two different inappropriate prepositions in the very same essay, as in (259), or the very same collocation reveals different deviations in different learners' essays – for instance, *pay attention to* was used six times without the necessary preposition, in two cases with different inappropriate prepositions, as in (254) and (255). The only pattern that emerges is that the pairs *in* and *to* and *in* and *at* are confused by the learner group under investigation. The confusion of the latter pair can be explained by the fact that *in* and *at* are not differentiated in Turkish; i.e., both are expressed with the locative suffix (namely, '-dE'). However, in order to provide a possible explanation regarding the confusion between *in* and *to* (both of which require different case endings in Turkish), one would need to investigate the learner data with regard to the use of these two prepositions in general

(259) *This bad attitudes [sic] of their father and mother affects them very much, plays very important role **on** their characteristic*

features. [...] the divorce plays a very important role in community life. <TRCU1158>

The last category of deviations consists of unnatural uses of verb-noun collocations – one other fairly common deviation attested in the data. There are two types of unnatural uses: (a) unconventional, unrecorded verb-noun combinations – what Wang (2016) referred to as “awkward combinations”, (b) use of a conventional (existing) verb-noun combination in a context that does not accord closely with NS preferences. In the case of the latter type, the native speakers approached judged them only acceptable but provided the alternative verbs they would have used. The example provided in Table 24, namely, *become pregnant*, for instance, is an entirely correct, existing collocation but for some, unknown, reason *get pregnant* was the preferred option of the NSs. COCA results support this preference. The frequency of *get pregnant* is over three times more than that of *become pregnant* (5,820 vs. 1,921 hits). *Have (an) education* versus *get (an) education* is another pair where we see such disparity – the first was attested 353 times in COCA while the latter 1,000 times. Some collocations produced by the learners are, however, less acceptable. Here are three more examples from the data:

(260) *Second reason, why I support the idea that most university degrees do not prepare us for real world has **become a clear point** in my mind [...]* <TRCU1058> (target: ‘become clear in my mind’)

(261) *university degrees **become of little value** since they don't reach the aim of preparing students for the real world.* <TRCU1131> (target: ‘lose value’)

(262) *People may make extra business or **give a start to** different working areas just for fun.* <TRCU1118> (target: ‘give a go’ (coll.)/ ‘try out’)

The last example (*give a start to*) has only six occurrences in COCA, which may explain the native speakers’ negative judgment.

Unlike these acceptable but not preferred uses, the data is full of unconventional verb-noun combinations – some of which are more ‘creative’ than the others. In (263), for instance, we see that the learner blended two existing multi-word verbs – namely, *get along well* and *get cross with*. In (264) the learner did not take a risk and showed the earlier reported over-reliance on high-frequency verbs. The examples (265) and (266) display direct translations from Turkish. The combinations in (267) and (268) are unconventional because the meanings of the noun collocates are contained in the verbs. All the substitution errors reported earlier could indeed also fall under this category since an inappropriate choice of a component results in an unnatural verb-noun collocation.

(263) *Media shouldn’t interfere in where and with whom the models go at night or which artist **get a well** and **cross with** which artist, or what they wear whether they go out naked or with long, very long dresses, [...]* <TRKE2012>

(264) *On the other hand, some teachers don’t **make their responsibilities** properly. Some may not come to lessons regularly and some only come, tell the lesson and go.* <TRCU1131> (target: ‘live up to’ or ‘follow through with’)

(265) *Although we seem to **come to good positions** in working life, we are not thought to do a work successfully.* <TRCU1137> (target: ‘get good positions’)

(266) *After the birth maybe there would not be good conditions for it, baby's life would not be **taken under guarantee** or [...]* <TRCU1095> (target: ‘would not be guaranteed’)

(267) [...] *but he/she does not **strive any effort** in order to get [sic] good mark* <TRCU1111> (target: ‘strive to get’ or ‘put in any effort to get’)

(268) *Therefore I believe that if a marriage isn't just working out it's best to let go of the union and **begin a new start** in life before it's too late.* <TRCU1171> (target: ‘start over’, ‘begin’)

In most cases the unusual collocations produced by the learners can be traced back to their L1 – to put it more precisely, they resulted from the earlier mentioned “coarser-gridding” of Turkish lexicon compared to English. For instance, the polysemous verb *put* has the same core meaning with its Turkish counterpart. Yet, the Turkish equivalent *koymak* covers a larger semantic web, including *place*, *lay*, *set*, *pour*, among others. This “one-to-many semantic divergence” may explain the unusual verb noun-combinations seen in (269) – at least the first one. The unusual collocate *win an exam(ination)*, which was used five times by four learners, can similarly be accounted for by the fact that one verb in Turkish (namely, *kazanmak*) is used in the meaning of all these three verbs. That is, the verbs *win*, *gain* and *earn* share an equivalent in Turkish.

(269) *In many hauses [sic] the leader is the father. [...] Usually his rules are valid at home. He goes to work, He earns money,*

and he puts the rules. The males puts the last decision although the females have the idea. <TRCU1149>

It appears that other languages known by the learner, not only their L1, may also influence the use of verb-noun collocations. The inappropriate verb choice in (270), for instance, can be accounted for by the learner's second FL – namely, German – in which a formally similar verb to *become* would be the right verb choice: *eine Stelle bekommen*.

(270) When we look around us we notice that the academicians, lecturers are educated in universities. So has can they become these positions. <TRME3020> (target: 'get/obtain')

The deviations attested in the data were categorized and presented separately, and the examples provided so far mostly included only one type of deviation in the context of verb-noun collocations. Nevertheless, quite a few collocations in the data deviate from the target norms in more than one way. The sentence (271), for instance, includes both an additional determiner and an inappropriate choice of preposition while (272) includes an additional use of a preposition and an omission of the indefinite article. The (273), on the other hand, displays an inappropriate choice of both a preposition and a noun. The inappropriate noun choice in (273) can be accounted for by a direct translation of its Turkish equivalent (*gün ışığı-n-a çıkmak*). Another instance where L1 could be an influential factor in the use of a deviant collocation is (274): The collocation does not only include an additional determiner but it is as a whole unit inappropriate in the context.

(271) *It creates a chance for everybody to **take a part at** a discussion.* <TRKE2001> (target: 'take part in')

(272) *The equality between men and women doesn't mean that if women can work or if men look after a child at home they are equal. It's not as simple as it's heard. We have to **go to great distance**.* <TRCU1038> (target: 'go a great distance')

(273) *the social and psychological [sic] troubles which **come into daylight** and [...]* <TRCU1110> (target: 'come to light')

(274) *It is doubtless that planning **takes an important part** in teaching process [...]* <TRKE2014> (target: 'makes up')

(275) *I'm going to tell about an event from a film in which while **making interview**, speaker asks to the men [...]* <TRCU1005> (alternatives: *interview, perform an interview*)

In (275) we see more than a simple deviation. It displays also the tendencies of the learners we have reported: reliance on simple verbs and preferring a stretched verb-noun construction over a one-word unstretched verb.

In addition to the four deviation categories presented, there are some instances where the deviation in the produced verb-noun collocation cannot be fixed by changing or omitting only one component of the collocation; a more major adjustment is instead necessary. For instance, the unconventional verb-noun combination *take model* in (276) needs to be rephrased as 'taking the previous generation as a role model'.

(276) *This idea came to present day by way of **taking model** from generation to generation.* <TRCU1157>

In some cases, as exemplified in (277) and (278), the intended meaning by the learner is ambiguous, making adjusting or rephrasing unfeasible:

(277) *When a marriage ends, the children, the family and the society influence its effect.* <TRCU1085>

(278) *University degrees are theoretical and do not prepare students for the real life is a common lament of students which exactly represents human beings desire to find everything has been prepared by someone else for them to take the pleasure of.* <TRME3019>

Before moving to the summary of the main findings, one more remark is in order regarding the acceptability of the verb-noun collocations attested. Albeit rarely, the native speaker judgments differed from one another now and then; for instance, in the case of *practice knowledge*, while one of the NSs consulted considered this collocation correct, another stated that one cannot *practice knowledge* but *put knowledge into practice*. Similarly, *sink into one's subconscious* was considered odd by one of the NSs, albeit not entirely incorrect. Some collocations that were judged inappropriate were attested in COCA. This variability in the use of language and discrepancy between NSs with regard to acceptability judgments may be seen as an indication that there is a grey zone between what is acceptable and what is not.

In this section, a detailed analysis of the deviant verb-noun collocations was presented. The aim was to identify the major types of difficulties in the use of verb-noun collocations as well as those of individual collocations and elements of collocations that appear to be particularly

challenging. The findings of the present study, mostly in line with earlier research on the use of collocation usage in IL, show that the learners lack collocational competence, relying on the ‘open choice principle’ in the use of verb-noun collocations to a great extent. This reliance results in an impression of non-idiomaticity.

Earlier studies (e.g., Nesselhauf 2005; Wang 2016; Howarth 1996, 1998a) have shown that although most deviations in the use of collocations lie in the verb, all constituents of collocations can be deviant. These findings are confirmed by the present results. Most of the deviations attested in verb-noun collocations involved an inappropriate choice of a verb. Other deviations included prepositional errors, determiner errors, inappropriate choice of nouns or their correct grammatical form as regards number marking, unnatural uses, and use of existing collocations in an inappropriate context.

One other clear finding of the present study is that the learners displayed a tendency to rely on high-frequency verbs (i.e., light verbs) in their use of verb-noun collocations instead of lexical verbs with a more specific meaning – a tendency reported in a myriad of other studies (see Section 2.3.4.1). Although the learners opted for the high-frequency verbs, they confused them with one another and produced a considerable number of deviant as well as unusual collocations – two features contributed to unnaturalness of the IL under scrutiny. Occasionally the reliance on verbs with general meanings failed to convey a clear meaning (e.g., *take skills*). These findings confirm Lennon’s (1996: 28) claim, namely that

the learners lack information as to the collocational possibilities of common verbs. One other noticeable feature of the IL under investigation in this respect was the common use of stretched verb-noun constructions instead of their derivationally related one-word corresponding verbs. This finding is in line with Nesselhauf (2005), who also observed this tendency in the writings of advanced German-speaking learners. Some of the stretched verb-noun constructions attested in the data were unusual/unrecorded (e.g. *make inheritance* and *give harm*). Considering all the unusual combinations attested in TICLE, it can be concluded that the highly polysemous high-frequency verbs, which can be combined with a wide range of noun collocates, are used by the Turkish learners as “lexical teddy bears” – most probably either because a more specific verb was not in their repertoire or they did not feel safe and/or certain about the more specific alternatives. As observed in earlier research (e.g. Nesselhauf 2005; Men 2015), the results have shown that in addition to light verbs, other (semantically related) verbs which are quite common and learned at a fairly early stage (e.g., *enter*, *become*, *achieve* and *reach*) also pose difficulties to the learners.

The influence of L1 was clearly discernible in all types of deviations in the extracted verb-noun collocations from the data, particularly in the inappropriate choice of verbs. Although not all the deviations attested can be accounted for by L1, the learners in general displayed a heavy reliance on L1 in their use of verb-noun collocations. As reported in earlier studies (e.g., Nesselhauf 2005; Yamashita & Jiang 2010; Wolter & Gyllstad 2011), incongruent collocations, whose lexical components

differed in two languages, caused the most problem. Furthermore, the incongruencies in lexical “gridding” between the learners’ L1 and L2 (e.g., one translational equivalent of *win, earn, gain* in Turkish) have proven to be as a source of difficulty in lexical learning, as proposed by Dagut (1977). Yet a closer examination of the data has shown that L1 is only one factor in the use of L2 collocations; there are other, more subtle factors underlying the inappropriate collocational uses in the data. Formal as well as semantic similarity between verbs (e.g., *experience methods* for *experiment with a method*), existence of semantically related noun collocates (e.g., *cover lessons* in analogue to *cover a course material/subject*), and incomplete grasp of (fine-grained) semantic distinctions between semantically similar verbs (*meet* vs. *encounter a problem*), analogy from known expressions, influence of other known languages to the learner¹⁵³ are among potential sources of the deviations.

Taken together, the results reported here have firmly ascertained the difficulty the learner group under scrutiny experiences with regard to verb-noun collocations. L1 has proven to be an important error-inducing factor on the use of collocations in L2. The reliance on it usually led to inaccuracies and inappropriate uses by the learners. Regardless of the

¹⁵³ As mentioned in Section 4.2, all the languages known by the learners are not specified on the ICLE-CD. Therefore, the influence of other languages known to the learner on their use of verb-noun collocations could not be investigated to the full extent.

source of the difficulty, all the deviations presented in this section display the learners' incomplete and insufficient collocational knowledge, distinguishing their language from that of NSs.

6.4 Synopsis: The qualitative use of multi-word verbs

The main aim of this chapter was to provide information on the Turkish learners' qualitative use of multi-word verbs. By-and-large, the results align with earlier research investigating the use of multi-word verbs in IL of learner groups with various L1 backgrounds.

To recapitulate the main findings, the study notably found an abundance of clear erroneous uses of multi-word verbs (in all four categories investigated) along with questionable, unconventional uses, resulting in a lack of idiomaticity in the learners' written production. Most of the deviations discovered in the multi-word verb use involved an inappropriate choice of a verb. Other deviations included prepositional and particle errors (addition, omission, or substitution), determiner errors (addition or omission), inappropriate choice of noun collocates or their correct grammatical form in regard to number marking, and the use of existing multi-word verbs in inappropriate contexts and unconventional (i.e., unrecorded) combinations.

Significantly, the study also uncovered Turkish learners' over-reliance on verbs with general meanings (high frequency verbs). It is clearly seen in the data that the learners mostly operated on the open choice principle in the use of multi-word verbs. In the construction of verb-noun collocations, the learners frequently resorted to delexical verb-

noun collocations instead of making use of a lexical verb with a more specific meaning (e.g., *give pain to N* instead of *hurt*). The learners displayed a general tendency to prefer stretched verb constructions rather than using their corresponding one-word equivalents (e.g., *make an invention* instead of *invent*). In the case of phrasal and prepositional verbs, the most frequently exploited verbs were also high-frequency verbs. In some instances, the learners' clinging onto the high-frequency verbs was considered inappropriate by the native speakers due to register issues (e.g., *look at patient's illness* instead of *observe patient's illness*). Although high frequency verbs were used as safe choices – in Haselgren's terminology as “lexical teddy bears”, the learners often displayed confusion among these verbs and combined them with unusual noun collocates. Even the very frequent and fixed collocations that consisted of a common verb and a noun (such as *take care of*) were attested with incorrect complementary prepositions or used in inappropriate contexts by the learners. Similarly, improper uses of early-learned phrasal and prepositional verbs (such as *grow up* or *look at*) were abundant in the data. This finding confirms Lennon's (1996: 28) assertion that L2 learners have difficulty understanding meaning and usage boundaries among some very common verbs, and the learners appear to lack information as to their collocational possibilities/restrictions. Not only high-frequency verbs but also less common but early-learned, semantically-related verbs (such as *reach* and *achieve*, *meet* and *encounter*) posed difficulties to the learner group under scrutiny. These findings converge with previous studies (e.g. Men 2015; Nesselhauf 2005).

It is believed that, as claimed by Dagut (1977), the generally ‘coarser’ gridding of the lexicon of the learners’ L1 compared to that of their TL (‘one-to-many semantic relationship’) poses difficulty for the Turkish learners under investigation. There were clear instances in the data where the learners carried over their L1 patterns in their use of multi-word verbs or directly translated L1 expressions into their L2, which mostly resulted in the use of additional prepositions and substitution errors. The deviated, unconventional uses of some specific multi-word verb types were shared by some learners. That is, these deviations do not represent individual learner’s usage, a claim that adds to the validity of the assumed L1 interference. Since this chapter focused on the erroneous uses of multi-word verbs, we mainly talked about the negative influence of L1 during the analysis of the results. Clearly, there are instances where the learners’ L1 had a facilitative influence on their multi-word verb use, but, as already pointed out by Waibel (2007: 128), it is “much more difficult to decide whether the correct production of a target language (TL) feature is due to positive transfer or to the fact that the TL feature was mastered during the acquisition of the foreign language without any NL influence.”

Some of the unconventional, i.e. non-standard, uses of prepositional verbs noted in TICLE (such as *return back*, *enter into*, *discuss about*) have been reported for other EFL and ESL varieties. In the case of the latter varieties, they have been considered ‘innovative’ or ‘creative’ uses rather than deviations (see Section 2.3.4.3). Similar ‘creative’ uses were also found in the use of phrasal verbs. For example, aspectual

meanings of the particle *up* were used by individual learners to form new phrasal verbs (such as *knit up* and *heighten up*). Furthermore, redundant uses of prepositions – in contrast to our expectations – were more frequent compared to omission of them. On one hand, the creative uses of prepositions and particles, as well as having more redundant uses of prepositions in the data, suggests that the learners under investigation are aware of the abundance and importance of these units in their TL and consciously attempt to use them. Conversely, the data may indicate that unconventional and erroneous uses of multi-word verbs are only developmental features of this IL variety; that is, the learners have not yet internalized the collocational links between the units of multi-word verbs.

L1 is obviously only one factor in the use of multi-word verbs; there are other, more subtle factors underlying the inappropriate multi-word verb uses in the data. Just as formal similarity between verbs (e.g., *come over a problem* instead of *overcome a problem*, *adopt to* instead of *adapt to*), semantic similarity between verbs [i.e. incomplete grasp of (fine-grained) semantic distinctions between semantically similar verbs (*meet* vs. *encounter a problem*, *reach* vs. *achieve*, *behave* vs. *treat*)] and confusion due to existence of semantically related noun collocates (e.g., *cover lessons* in analogue to *cover a course material/subject*) prove to be potential sources of deviation, so, too, do closely related prepositions (*in* vs. *into*), analogy from known expressions (*final in* in analogue to *result in*), and influence of other known languages to the learner prove to be potential sources of the deviations.

Although some prepositions (namely *with*, *to*, *from* and *in* – the equivalents of the instrumental, dative, ablative, and locative postpositions in Turkish) were recurrently used with one-word verbs as an additional unit by means of carrying over L1 patterns into L2 (e.g., *date with*, *join in*), no clear pattern has emerged with regard to the erroneous uses of other prepositions and particles. This lack of a clear pattern calls for further investigation of the processes involved in the use of these items in the formation of multi-word verbs.

Lastly, it can be stated that the presence of several multi-word verb types in the data was triggered by some essay topics. The topic-induced multi-word verbs were repeatedly produced by different learners. In accordance with this repetition and the learners' reliance on the high-frequency verbs, the use of multi-word verbs in TICLE lacks variety.

On the basis of the above reported findings, it can clearly be concluded that the present study lends support to the reports in the SLA literature of L2 learners' deficient phraseological knowledge.

Chapter 7: Summary and conclusions

The more languages you know, the more human you are.
(Tomáš Garrigue Masaryk)

7.1 Introductory remarks

The present study investigated the use of multi-word verbs in the essays written by intermediate to advanced level of Turkish learners of English. Aside from being the first large scale corpus-based study to be carried out on the IL performance of this group of English learners with regard to the multi-word verb usage, the significance of the present study lies in the fact that four different categories of multi-word verbs – namely, phrasal, phrasal-prepositional, prepositional verbs and verb-noun collocations – were investigated (rather than restricting the research to only one category of multi-word verbs). Furthermore, the notions of over- and underuse as well as the interrelation of possible factors in the quantitative use of phrasal and phrasal-prepositional verbs (or lack thereof) were also examined. Analysis of the use of these two multi-word verb categories was deemed critical as they were reported to be avoided and underused by many learner groups.

This study aimed to illuminate the complexities of multi-word verb learning in an L2 by focusing upon the influence of L1. In order to gain a better understanding of the phraseological competence of the learners,

both the qualitative and quantitative aspects of the learners' performance were investigated. More specifically, the focus was first on the aspects of frequency and lexical choices in the use of multi-word verbs and second on an error analysis on the extracted multi-word verbs.

This chapter is arranged as follows: In Section 7.2, the major results of the present study are summarized. In Section 7.3, some potential implications of the results for foreign language teaching are provided. The chapter concludes with Section 7.4 offering suggestions for future research in the study of multi-word verbs in learner language.

7.2 Summary of the main findings

Corroborating findings from previous studies investigating the multi-word verb use in learner language (see Section 2.3.4), both quantitative and qualitative deficiencies were identified in the written productions of Turkish-speaking learners of English. Quantitative deficiency was exemplified in the low number of phrasal verbs in TICLE compared to LOCNESS. The regression analyses performed with the learner-related and external variables to help predict Turkish learners' potential use of phrasal and phrasal prepositional verbs could explain only 9 percent of the total variation in the data set, and only one of the variables – namely, *text length* – was found to be significant. *Text length* displayed a positive correlation with the frequency of phrasal and phrasal-prepositional verbs. As for the qualitative deficiencies, a considerable amount of the multi-word verbs extracted from the learner corpus (6,129 tokens in total) included deviations from the target norm. The deviations occurred

in all components of the multi-word verbs (verbs, noun collocates, prepositions, particles, determiners), and a considerable number of existing multi-word verbs were used in contexts where they did not match the meanings found in dictionaries. Furthermore, the learners' multi-word verb repertoire (as can be judged from their output) was not very rich. A limited set of multi-word verbs – mostly those that contained a common, high-frequency verb – were employed recurrently and in a limited range by the learners, indicating clearly that the learners lacked diversity in their multi-word verb repertoires. Accordingly, only very few verb types were particularly frequent and productive in TICLE.

Deficient collocational knowledge of the Turkish learners was clearly discernible in all of the four multi-word verb categories investigated. In essence, multi-word verb use in the TL posed a general problem for the learner group under scrutiny, regardless of whether or not a similar multi-word verb structure exists in the learners' L1. The learners lacked knowledge of collocational possibilities and restrictions of verbs (even very common ones). It appears that this lack of knowledge left the learners no choice but to combine verbs with inappropriate nouns and/or prepositions and particles to communicate the intended meaning – either applying what Sinclair (1987, 1991) referred to as the 'open choice principle' (combining units according to grammatical rules in TL; see Section 2.3.1) or relying on the translational equivalents of L1 expressions and grammatical patterns. The learners' common use of high-frequency verbs revealed their general confusion in their proper use. Furthermore, the learners displayed difficulty differentiating semantically

and, albeit rarely, formally similar verbs in their L2. This finding can be accounted for by the way our mental lexicon is organized. It is now well established that words in our L1 and L2 mental lexicons are linked to each other by multiple types of relations, including formal (mainly phonological), semantical and collocational as well as by association, and it is also known that words with similar meanings are stored close to each other (e.g., Stella et al. 2018; Channell 1988; Wolter 2001; Jiang 2002). This organization in the mental lexicon surely affects the way we access and retrieve words. Language production involves the selection of appropriate words to convey intended meaning – that is, coming up with a word occurs according to meaning (e.g., Channell 1988: 85). The appropriate word selection is a challenging task for L2 learners, especially when the choice is to be made among a group of semantically related words. Even in native speakers' online language production, the retrieval of an intended word can fail for a period of time ['tip-of-the-tongue (TOT) state'¹⁵⁴] or fail due to error in word choice. In the case of an inappropriate word choice, the produced and the targeted words can be either semantically related (semantic errors) or phonologically related (the so-called 'malapropisms' (Fay & Cutler (1977), cited in Channell 1988: 87)). Such speech errors can (and do) occur for various reasons, such as a lapse in memory, in online language production.

¹⁵⁴ TOT experiences have been shown to reflect temporary failure of phonological activation rather than failures of semantic activation or lemma retrieval (see e.g., Lovelace 1987; Brown and McNeill 1966).

There are, however, two other interdependent factors causing such errors (especially semantic ones) in the language production of L2 learners: first, the role of L1 knowledge (both lexical and conceptual) on the structure of L2 mental lexicon and, second, incomplete and/or incorrect acquisition of words – especially that of semantically related words. Shown by Men as recently as 2015, the increase in the number of semantically-related words ('synsets') in the L2 learners' mental lexicon causes interference ('semantic interference') in the word selection process. The fact that L2 learners draw heavily upon their L1 lexicon and conceptual knowledge in structuring their L2 mental lexicon has been widely accepted. The results of the present study clearly confirm the work of Wolter (2006: 741f) in demonstrating that existing knowledge – although useful for building L2 lexical networks at times (especially at the beginning stages of L2 learning) – provides L2 learners with misinformation about acceptable combinations of L2 words. Due to the existing links between words in the L1 mental lexicon and limited as well as less rich input L2 learners typically receive, the links between words in the L2 mental lexicon are weaker and therefore more liable to confusion (Wolter 2006). Meara (1982, 1984) showed that the semantic links between the lexical units in L2 learners' mental lexicon are different from and less stable than those of native speakers. So, too, Meara proved that phonology plays a more important role in the structuring of links between units in the L2 mental lexicon. The weaker and different (inappropriate) links between lexical units in the L2 mental lexicon can account for the high number of deviant, idiosyncratic uses of multi-

word verbs seen in the data. Similarly, the occasional correct and inappropriate uses of a given multi-word verb side by side in a learner's essay adds support to the proposal of weaker links between units in the L2 mental lexicon. That means, these multi-word verbs as wholes had not been firmly entrenched in the learners' mental lexicon yet.

This above discussion is related to one of the main aims of the present study – namely, the extent of L1 influence in the use of multi-word verbs. On the basis of the abundant deviant and idiosyncratic use of multi-word verbs in TICLE, we can clearly claim that the Turkish learners displayed a high reliance on their L1. In all four categories investigated, L1's negative influence on choices of specific multi-word verbs (i.e., traces of direct mapping from corresponding L1 expressions) was clearly discernible. The use of a low number of phrasal and phrasal-prepositional verbs in the learner data can be seen as an indirect influence of L1, i.e. the lack of structural equivalent of verb+particle construction. The analysis of the one-word equivalents of some phrasal verbs showed that learners had a tendency to prefer them over phrasal verbs. In addition to this indirect influence, similar L1 interference was also seen across the multi-word verb categories investigated: redundant uses of prepositions and/or particles, substitution errors (inappropriate choice of multi-word verbs as a whole in a given context or of its components), omissions of one of the components of multi-word verbs, and unidiomatic (unconventional) combinations. L1 interference was more frequent when the English multi-word verb included a verb different

from the one in the learners' L1. It should, however, be noted that, although multi-word verbs with word-for-word translational equivalents were found to cause less difficulty in the choice of verbal element, deviations in the case of congruency also occurred on occasion as L1 equivalents of some L2 verbs or nouns cover more semantic range ('one-to-many semantic relationship').

. That is, as proposed by Dagut (1977), incongruencies in lexical grid-
ding between the learners' L1 (Turkish) in comparison to their TL (English) posed difficulty for the learners.

The learners' reliance on their L1 as a communicative strategy was also supported by the fact that the production of the same deviant multi-word verbs that were direct equivalents of their L1 counterparts (e.g., **come over a problem/difficulty*, **come across with*) was observed in essays of different learners. The deviant multi-word verbs in these essays might have resulted from the learners' combining individual units while writing. However, as proposed by Nesselhauf (2005: 248), it is also possible that these units might have been stored as wholes in the learners' mental lexicon as a consequence of frequent use by the learners themselves or other Turkish learners around them.

Along with L1 influences (carrying over L1 grammatical patterns and word-for-word translation of L1 expressions), the data identified other, more subtle, sources of error accounting for the learners' inappropriate and unidiomatic use of multi-word verbs. These sources include formal

as well as semantic similarity between L2 verbs, confusion due to existence of semantically-related noun collocates of the targeted verb and the produced verb, incomplete grasp of (fine-grained) semantic distinctions between semantically similar verbs (i.e., near-synonyms) and closely related prepositions, analogy to known expressions in L2, blends of L2 expressions, and influence of a learner's other known languages. Furthermore, in quite a few cases, the learners used existing multi-word verbs for meanings different than the ones provided in dictionaries. And, occasionally, the learners displayed more creativity in their multi-word verb use, forming new aspectual phrasal verbs and prepositional verbs (indicating 'hyper-explicitness').

The high occurrence of redundant uses of prepositions with simple verbs is related to the creative uses observed with phrasal and prepositional verbs. Whereas some of these redundant uses are interference errors (use of equivalent prepositions of the necessary case endings or postpositions in Turkish expression in English), some are 'hypercorrection' errors. Both the hypercorrection errors and the creative formation of new phrasal and prepositional verbs might indicate that the learners are aware of the importance and abundance of these 'small' units (at least in the formation of multi-word verbs), but they have just not internalized their correct use yet, supporting Hemchua and Schmitt (2006).

On the basis of all of the findings, we cannot conclude that one multi-word verb category is easier or more difficult than the other(s). One of the questions raised in this study was what factors contribute to the difficulty of using multi-word verbs. On the basis of the findings, we can

draw a few tentative conclusions. First, in the case of phrasal, prepositional and phrasal-prepositional verbs, the lack of the L1 equivalent of these structures in addition to their nature (i.e., (occasional) arbitrariness of their grammatical units and idiomaticity) might have caused difficulty for the learners. Second, the mostly transparent nature of verb-noun collocations causes little difficulty in perception (see Sections 2.3.4 and 2.3.4.1) and this relative ease in perception might have resulted in the collocations going unnoticed. So, too, other difficulty-inducing factors may be at play: a lack of overlap or a partial overlap between L1 and L2 meanings of lexical units and/or the existence of more than one translational equivalent in L1 for an L2 item. Lastly, the semantic and formal similarity between words (especially semantically-related verbs) and polysemous nature of (high-frequency) verbs pose a problem for the learners in their multi-word verb use due to a learners' incomplete lexical knowledge.

The ultimate aim of the study was to shed light on the question of what learner behavior tells us about the storage and acquisition of multi-word verbs and lexical chunks in general. Unfortunately, no clear pattern of multi-word verb acquisition and storage has emerged in the present study. Some assert that collocations with a high degree of fixedness, for instance, are easier for learners to acquire and less susceptible to L1 interference (e.g., Wang 2016; Nesselhauf 2005). Notably, the results of the present study do not fully support this suggestion. While some common, relatively fixed multi-word units (such as *keep in mind*, *take advantage of*, *pay attention to*, *put up with*) were produced without

much variation by the learners, others with similar features included deviations (e.g., *put in(to) practice/*practical*, *commit (*a) suicide*, *take (*a) place(*s)*, *take care of/*for/*about*, *take (in)to consideration*, and *come (*up) across*). That is, the (strong) collocational link between the units and the relatively stable grammatical patterns did not always enable their easy learning for the learner group under investigation. We can tentatively conclude that the multi-word units produced correctly by the learners (especially the ones without a word-for-word L1 translation) might have been stored as holistic units in the learners' mental lexicon and treated as such during their production (cf. Durrant & Siyanova-Chanturia 2015; Wolter 2006).

Related to the issue of storage, creative uses of aspectual phrasal verbs described in Section 6.2 show that some learners were aware of the meanings particles add to the combination. In these cases, it is possible to presume that the meanings of the particles were stored as separate entries in the mental lexicon of the L2 learners. These separate entries then allowed the learners the flexibility to combine the particles with verbs to form new aspectual phrasal verbs (see Waibel 2007: 163).

One last observation in this context is that the learners displayed a high tendency to opt for a verb-noun collocation rather than its one-word equivalent; i.e., learners commonly used stretched verb constructions instead of their non-stretched equivalents. They even formed unusual combinations instead of using the available one-word verbs (see Sections 5.4 and 6.3). As mentioned in Section 2.3.3, Wray (1999: 216, 2002: 75) proposed that the uses of longer multi-word units instead of

shorter alternatives available in the native language function as ‘time buyers’ while speakers/writers plan what to say next. Although the essays in TICLE were not written under a time constraint, it might be a possibility – albeit a slight one – that stretched verb constructions were used for the same reason by the learners.

As a considerable amount of multi-word verbs were found to be topic-induced, the list of particularly problematic individual (multi-word) verbs identified here cannot be considered exhaustive. Similarly, more potential factors correlating with the difficulty of the multi-word verbs may be found in future research. Nevertheless, the results of the present study have a number of pedagogical implications which will be presented in the next section.

7.3 Pedagogical implications

The present study bears a few important pedagogical implications for vocabulary instruction, which will be discussed in the following.

The central observation in the present study was that the use of multi-word verbs as a general phenomenon poses a significant problem for the Turkish learners of English investigated, whose proficiency level ranged between intermediate and advanced (see Section 4.2). The difficulty experienced in the use of these constructions, which revealed itself in various forms of deviations (see Chapter 6), is a clear indicator of the inefficiency and weakness in the way they are taught. The remarkable degree of error in the use of multi-word verbs is likely due, in

part, to the traditional focus on words as single items in vocabulary instruction. This single-word focus inevitably encourages the formation of questionable (i.e., unnatural) word sequences in L2, by relying on the so-called ‘open choice principle’ (see Section 2.3.1). Judging from the high number of deviations and questionable combinations observed in TICLE, clear evidence points to the failure of the methods and exercises used in the teaching of multi-word verbs in the context of the Turkish EFL setting. These methods and exercises need improvement as they fail, to a great extent, to provoke successful use of multi-word verbs. Of note in the present study, the number of years the learners had English instruction was not relevant for the accurate quantitative and qualitative use of phrasal and phrasal-prepositional verbs. This fact supports the need for greater, more dedicated, attention to multi-word construction in teaching, even at more advanced levels. As is obvious here, increased time learning English does not mean increased knowledge of appropriate use of multi-word verbs. Based on the overall results of the present study, I will add my voice to the voices of many others who have strongly argued for moving the focus in vocabulary instruction from single units to larger units. Research hitherto has shown that adult L2 learners tend to analyze language input more and store smaller chunks (see Section 2.3.3). Therefore, it is of paramount importance to teach multi-word verbs as wholes, focusing on their collocational habits (i.e. common collocates) and usage patterns.

In the present study, not only lexical units but also grammatical units were found to pose problems for learners in the formation of multi-word

verbs. This dual problem indicates, first and foremost, that all units in multi-word verb constructions deserve attention. Secondly, the errors indicate that learners need increased awareness of both the formulaic nature of ordinary language use and the difficulties presented in the production of multi-word units. As research has shown, the learners do not pay attention to the lexical and grammatical patterns in the input as they are not aware of the collocational aspects of lexicon. Men (2015: 173) has recently suggested that “[i]t is necessary for learners’ attention to be diverted from single lexical items to habitual word combinations.” With this necessary diversion, learners may start ‘noticing’ the patterns and/or (less salient) items (i.e., light verbs, prepositions and particles) that tend to escape their attention, and they might avoid creating unnatural and deviant combinations (see Schmidt 1990 for the role of ‘noticing’ in learning learning). Apparently, learners cannot always derive the collocational and syntactic restrictions (e.g., determiner use and number as in *commit a suicide* and *take places*) by themselves. It seems that learners struggle partly because they are not given enough input. That is, they have fewer encounters than necessary to recognize a multi-word unit as a ‘legitimate’ whole and store it in the mental lexicon (see the three steps proposed by Yamashita and Jiang (2010: 262) in learning collocations).

The abundance of learner errors is all the more remarkable given the fact that half of the learners in TICLE were allowed to use reference tools and that all the essays were written without time pressure (i.e.,

untimed). In the case of the learners for whom reference tools were allowed, the high number of deviations in the multi-word verb use might indicate a) inefficient use of available reference tools to find the targeted multi-word verbs and their usage (i.e., lack of dictionary skills), b) lack of awareness that they are experiencing difficulty – most probably as a result of not being cognizant of the fact that there are collocational restrictions between lexical units. Without awareness, the learners might have thought it unnecessary to use reference tools. As shown in earlier studies, learners are mostly unaware of their deficient phraseological knowledge; mere exposure to multi-word units is not enough to improve the learners' performance in their use (e.g., Nesselhauf 2005; Waibel 2007; Granger 1998a; cf. Webb et al. 2013). Having observed the same lack of awareness, Nesselhauf (2005) correctly suggested informing L2 learners of the fact that there are word combinations which are “neither freely combinable nor largely opaque and fixed (such as idioms) but that are nevertheless arbitrary to some degree and therefore have to be learnt” (p. 252). On the basis of the results, teaching multi-word verbs explicitly—focusing on all their components, meaning(s) and usage patterns—is believed to be indispensable to improve error-free production of L2 learners (see Nesselhauf 2005: 269).

Aside from learner cognizance, language teacher awareness must be addressed. Teachers must recognize the formulaic nature of language use as well as the importance of multi-word units in language acquisition. Moreover, teachers should realize that learners might not even be aware of the difficulties they experience in the use of multi-word verbs (in

fact, multi-word units in general). As stated by Arabski (1979: 103), language teachers need to grasp how the error-making mechanism works and know where to expect errors. After all, the teachers are the ones who can initiate and control the acquisition of these units as well as accelerate their acquisition rate. Having the same L1 background as the learners would be of great help for teachers to anticipate where the potential difficulties lie, i.e., which verbs or verb constructions might cause particular problems for a given learner group. In fact, Armstrong (2004: 213) has asserted that the difficulty phrasal verbs pose to learners may be partly due to insufficient conscious awareness of language teachers. The researcher posits that teachers may be impaired with a deficit in understanding the systems that underlie these multi-word units. A semantic and systematic analysis of multi-word verbs would aid in abandoning the common idea of multi-word verbs as units to be memorized or units without any system and, in turn, as suggested by Armstrong (2004: 213), may improve the teachers' effectiveness in teaching them (see Torres-Martinez 2015).

Learner and teacher awareness alone is not enough to entirely solve the problem at hand. One of the reasons proposed for the failure of EFL learners in acquiring formulaic language to native-level competence has been lack of sufficient and rich enough input (see, for instance, Durrant and Schmitt 2010). Given the fact that textbooks are the main source of language instruction in some EFL settings—for instance in Turkey (Kartal 2018: 545), they are of utmost importance to support the acquisition of (multi-word) verbs. Therefore, these books need to repeatedly

present verbs in natural and rich linguistic contexts so as to increase the chances of learner retention and to enable the learners to understand the full meanings of verbs (i.e., multiple meanings or shades of meaning as well as registerial characteristics) (Judd 1978: 71ff; Carter & McCarthy 1988: 45). The findings of the present study can be taken as a proof of the need for improvement in the ELT materials used in Turkey. As seen in the high number of collocational errors in TICLE that resulted from confusion between semantically similar verbs, most of which have one L1 equivalent in Turkish (such as *reach* and *achieve* or *behave* and *treat*) and high-frequency verbs, instances encountered by the learners were not enough for them to differentiate these items from each other, i.e. to derive the specific features that differentiate a given verb from those with similar meaning(s) and/or to internalize verbs' collocational patterns. These findings converge with previous studies that support the proposal of repeated exposure in vocabulary instruction (e.g., Webb et al. 2013; Yamashita & Jiang 2010). As pointed out by Granger (1998a: 159) with regard to the language-specific nature of multi-word units, the fact that most EFL materials are designed for learners coming from various L1 backgrounds (due to commercial reasons) constitutes a major problem that needs to be solved in order to provide learners the most efficient learning aids. One solution could be to complement these general EFL materials with extra materials prepared specifically for a learner group with a particular L1 background, taking into consideration the potential difficulties this learner group may encounter.

To be clear, repeated exposure to a word in itself does not guarantee memory¹⁵⁵, let alone its production or its proper usage. As Snellings, Van Gelderen and De Gloper (2004) have shown, L2 learners will not use the words for which they have receptive knowledge in their writing unless they receive focused and repeated practice with those words. Moreover, learning is not a passive process; active coding is necessary during the learning process (Reisberg 2013: 180ff; see the references therein). That is, exercises created to assist the acquisition of new verbs (in this case multi-word verbs) should be more than just an exercise in a rote, mechanical fashion; the exercises need to trigger deep processing. They need to prompt learners to think about the verbs in terms of meaning, to analyze their components, to relate them to the words and collocates they already know since deeper processing ensures better recall (see, for instance, Hyde and Jenkins 1969; Craik and Tulving 1975). Compelling evidence proves that effects of intention to learn are indirect; the quality of the memory depends on the strategy chosen by the learner in the learning process (Reisberg 2013: 184). Therefore, teachers should be familiar with learning mechanisms, i.e., how our memory functions. Teachers need to be aware of the need for active as well as deep and elaborative encoding in order for the new words to enter a student's long term memory. Teachers should know appropriate learning strategies so as to implement them and teach them to learners.

¹⁵⁵ How remarkably bad our memory is even for an object we see on a daily basis (namely, penny) was shown by Nickerson and Adams (1979).

The direct application of corpus analysis methods in the classroom (i.e., data-driven-learning (DDL)) is a practice that allows deep processing by means of exploration of language structures. Just as this practice enables deep processing, so, too, it motivates learners to (further) discover their TL by self-driven research interests or questions. The idea of a learner as a researcher, as Johansson (2007: 25) pointed out, clearly aligns with thoughts about language learning as an active process. It has been asserted that DDL is particularly suitable for the instruction of usage aspects of multi-word units to raise learner consciousness, thus allowing learners to observe and internalize repeated patterns and meanings (e.g., Meunier 2002: 130; Nesselhauf 2005: 269f; Xiao & McEneyr 2006: 126). The importance of corpora in language teaching has been well established in the last decades (e.g., as a source of authentic language, as a source of information on current trends, as a source of information about collocation and colligation patterns of lexical units and their semantic prosody, as a correction and assessment aid).¹⁵⁶ As a source of authentic input in the absence of exposure to TL in its naturalistic environment, corpora create the necessary natural and rich context in which to present new lexical items. Torrez-Martinez (2015: 57) states that corpus-driven language learning tasks such as DDL or paper-based DDL may provide enough exposure to form and meaning. In particular, the researcher suggests expanding this needed learner input in the form of communicative activities (Torrez-Martinez: *ibid.*). I can

¹⁵⁶ For the role of LCR in language teaching see part IV in Granger et al. (2015: 443-534).

only agree with this suggestion since reinforcement of what is being learned is essential to establish memory (see Gass et al. 2013: 361). Classroom exercises might be better devoted to potentially problematic multi-word verbs in corpus-extracted contexts. This devotion would be beneficial for all proficiency levels: In lower proficiency levels, the learners might be prevented from committing errors before they get fossilized as learner habits (which would then be hard to eradicate (see Selinker 1972: 215f)) whereas, in higher proficiency levels, learners can explore key attributes that differentiate their language from that of native speakers, i.e., what makes their language unidiomatic.

The strong L1 interference identified in the data points to the need to underline the lexical incongruencies between learners' L1 and L2 since, as stated by Wolter (2006: 745), "when the L1 knowledge of lexical combinations is not sufficient for informing correct lexical choices in the L2, new connections will have to be made." If not, relying on the L1 patterns and translational equivalents, learners produce questionable and deviant word combinations to express their intended meaning. By applying a contrastive approach in vocabulary instruction, teachers who have the same L1 as their learners could make use of this common knowledge to predict potentially problematic multi-word verbs and, very early in the learning process, to draw learners' attention to the difference between the L1 and L2 word combinations. This approach would, as asserted by Yamashita and Jiang (2010: 663), help decrease the interference of L1. The application of this contrastive approach, which has also been suggested in earlier studies (e.g., Bahns 1993;

Granger 1998a; Nesselhauf 2005; Xiao & McEnery 2006; Yamashita & Jiang 2010; Men 2015), has been shown to be more effective compared to vocabulary instruction methods that ignore the crosslinguistic differences (e.g., Laufer & Girsai 2008). Moreover, being immersed in this approach, learners, as suggested by Men (2015: 173), may engage in deep learning during the comparison of L1 and L2 incongruent multi-word verbs.

Error-driven learning, which can promote the learning of multi-word verbs in general, is crucial in the comparison of L1 and L2 with regard to the meanings and patterns of multi-word verbs (see Nesselhauf 2005: 270f; on the benefits of negative evidence in SLA in general see, for instance, Gass et al. 2013: 360ff.; Ellis 1994: 639ff). Early on as a pre-task, learners could be presented with learner corpus-driven use of multi-word verbs including deviations (for instance, in the form of keyword-in-center (KWIC) concordances) to draw their attention to potentially problematic multi-word verbs. Conversely, learners could be exposed to their own errors as a post-task exercise to discover their own problems in the use of multi-word verbs. In the latter case, learners could simultaneously be provided with appropriate uses of problematic multi-word verbs from a native speaker corpus to compare native speakers' use with that of their own. The feasibility and efficiency of combining the learner and native-speaker data in DDL-based instruction has already been shown by, for instance, Cotos (2014).

Given the pedagogical benefits of corpora, I can only agree with Zareva's (2017) suggestion to develop ELT teachers' corpus literacy skills

early on during their training programs. In teaching programs, teachers need hands-on training using software and corpus data in order to learn how to develop instructional materials and classroom activities that enable self-driven and student-specific (in the sense of catering to the L1 background of learners) exploration of lexical patterns or language structures. The abundance of multi-word verbs in language (use), making it improbable to teach all the multi-word verbs learners wish to learn or need for their communicational purposes in the limited classroom time available, adds weight to this suggestion. The ultimate aim of teachers should be to inspire learners to take charge of their own learning, i.e. help them develop learner autonomy. To this end, teachers, as suggested by Nesselhauf (2005: 253), should show learners different strategies to acquire vocabulary outside the classroom. Dictionary skills would be one such strategy. In addition, corpus literacy skills are of paramount importance in gaining knowledge of correct use of multi-word verbs. Indeed, in order to be able to transmit such knowledge and skills to learners, language teachers themselves need to have acquired them.

In the present study, as mentioned in earlier sections, Turkish learners displayed a heavy reliance on a limited number of verbs – mostly high-frequency, ‘easy’ verbs – in the formation of multi-word verbs. However, despite the reliance on these verbs as ‘safe choices’, their use demonstrated a high degree of confusion – either among them or with semantically or formally similar verbs. This finding is of particular pedagogical significance (as pointed out, for instance, by Altenberg and

Granger (2001: 190) and Lennon (1996: 23, 35)) in recognizing the fact that in higher levels of language teaching verbs presented to learners in the initial stages of L2 learning should not be neglected in later stages. Learners apparently need more time and exposure to acquire the different meanings of these polysemous verbs and to derive their collocational possibilities and restrictions as well as their usage patterns. As Gass et al. (2013: 361) commented, acquisition, which appears to be gradual, usually “requires numerous ‘doses’ of evidence” (both positive and negative). Therefore, as Lennon (1996: 35) and Granger (2009: 26) indicated, in the higher level proficiency, the aim of teaching vocabulary should not be just to enlarge the L2 lexicon by teaching low-frequency verbs but to also deepen the knowledge (semantic and collocational) of the early learned, ‘easy’ verbs (by means of various exercises in the classroom).

As for the confusion displayed between semantically related verbs, the pedagogical implication that can be drawn is the necessity of also focusing on a verb’s near-synonyms and of highlighting the subtle semantic distinctions between them in vocabulary instruction. That is, learners should be made aware of what differentiates a group of semantically related verbs from one another – as already suggested by, for instance, Channell (1988: 90; see references therein). Teaching verbs decontextualized and through L1 translation equivalents is rather problematic, especially in the case of ‘divergence’, as the results of the present study clearly revealed. As Men (2015: 172) rightly commented, “with the same translation equivalent, the collocational behavior of semantically

related words is highly likely to be believed as the same by L2 learners.” Therefore, teaching semantically close verbs simultaneously and through their collocates (i.e., presenting the learners the overlapping collocates and collocates exclusive to a particular verb), as suggested by Men (2015: 170f), would allow learning of common meanings and distinguishing features of each verb.

So far, we have focused on what can be done to improve the efficiency of teaching multi-word verbs and of learning on the basis of the results. With the ever-raising awareness of formulaic nature of ordinary language use, the inclusion of multi-word units in teaching materials and L2 pedagogy has been widely accepted. Due to the abundance of multi-word units, the question remains is which multi-word verbs should be presented to L2 learners. In other words, which multi-word verbs are most worthy of examination in limited classroom time available? Various researchers suggested different selection criteria – congruence, semantic opacity and frequency being the common ones. For instance, quite recently, Martinez (2013) presented a model in which the properties of frequency and semantic opacity are considered as ‘complementary continua’. Based on the results of the present study, a few conclusions can be drawn in this regard. First, due to the confusion experienced by learners in multi-word verb use formed with high-frequency verbs (i.e. light verbs in the case of collocations), high-frequency, common verbs warrant great emphasis in language teaching. Therefore, these verbs must be included and dealt with intensively in the ELT materials, including in the materials designed for higher levels

of proficiency. It must be noted that some learner errors in the use of multi-word verbs were more serious, resulting in failures of understanding the intended message by the native speakers consulted. Thus, in the selection process, as suggested by Nesselhauf (2005), the degree of ‘disruption’¹⁵⁷ should be taken into account. The focus in teaching should be on repeatedly used, non-existing multi-word verbs of different learners as well as on the existing multi-word verbs that are repeatedly seen with deviations or are used inappropriately. Most importantly, the strong L1 traces in the data point to the need to focus on multi-word verbs that are dissimilar in L1 and L2 (as suggested by, for instance, Bahns 1993; Marton 1977). However, both congruent and incongruent multi-word verbs need emphasis in teaching. While the latter should be of high priority due to new connections between links in the L2 mental lexicon needing development, the former cannot be ignored since it is hard to determine whether they have really been learned. As claimed by Wolter (2006: 743), learners might bypass the L2 acquisition process entirely by simply relying on their L1 lexical network when there is a marked overlap between the L1 and L2 lexical networks. Therefore, we cannot assume that the correct use of congruent multi-word verbs has been acquired and will be produced as whole units by learners. As seen in the investigated data on learner texts, the frequent inappropriate contextual use of existing multi-word verbs and the numerous occurrences

¹⁵⁷ Nesselhauf (2005: 256) defined ‘disruption’ as “the degree to which an expression, if deviant, draws the listener’s or reader’s attention away from the message or makes the recipient misunderstand or fail to understand the message”.

of correct and incorrect uses of a given verb side-by-side lend credibility to the idea that learners do not always fully acquire lexical knowledge.

The presentation of multi-word verbs in the materials is equally, if not more, important than their selection. Therefore, both writers of learning materials and teachers should systematically present these multi-word units to learners in order to avoid unnecessary confusion (see, for instance, Darwin and Gray (1999) for a systematic approach to phrasal verb classification).

Fortunately, one significant observation in the data carries a positive pedagogical implication worthy of optimism for language teachers. The Turkish learners' use of multi-word verbs – some of which were not only incongruent but also restricted in terms of their collocability and syntactic use – were target-like. This finding can be seen as an indicator that these multi-word units might have been learned and treated as single items by these learners, and in turn, the finding proves that that multi-word verb constructions are learnable after all. This supports the proposal that EFL learners are capable of developing collocational links without relying on their L1 (Wolter & Gyllstad 2011; Men 2015). Now, the important question remains is how we, as educators, contribute to and facilitate the learning of multi-word verbs in an efficient way in the classroom context.

7.4 Limitations of the study and ways forward

This study was born out of the need to fill the gap in the LCR literature on the use of multi-word verbs by Turkish speaking learners of English. It provides a comprehensive insight into Turkish learners' use of multi-word verbs yet it is limited in some ways. The results obtained here are based on the learners' written output and all of the learners in the investigated corpus are at a similar level of proficiency. Therefore, the results are restricted to one learner group with the same L1 background and one mode of language. Some interesting questions are inevitably left unanswered due to these limitations. In what follows, some of these limitations will be pointed out for the benefit of further studies.

Whether some of the errors committed by the learners in their use of multi-word verbs (for instance, the ones claimed to be interference and hypercorrection errors) are developmental or fossilized structures could not be determined in the present study since all the learners investigated, as mentioned earlier, were at a similar level of proficiency. Likewise, no conclusions could be drawn on whether the Turkish learners' high reliance on L1 and on high-frequency verbs (e.g., light verbs) in the use of multi-word verbs (which has been proposed to be a developmental IL feature (e.g., Wang 2016; Men 2015)) would decrease with increasing proficiency. In order to shed light on these questions and gain a better understanding of how multi-word verb knowledge develops in learner language, development studies or studies capturing a wider range of proficiency levels need to be carried out. A longitudinal learner corpus would allow the identification of patterns in L2 learners'

phraseological development but to my knowledge, no such corpus of Turkish learners of English exists as of yet.

One limitation of the study resulted from the learner corpus taken as the database. In TICLE, only four learners had been in an English-speaking country (see Section 4.2). Therefore, the variable *months abroad* (which is highly likely to be influential in the use of phrasal and phrasal-prepositional verbs (see Section 5.2.5)) had to be excluded from the regression analyses. Most of the variability observed in the quantitative use of these two multi-word verb categories could not be explained by (the interplay of) the investigated learner-related and external variables. This finding allows the conclusion that there must be other factors at play that are more relevant to the learners' use of phrasal and/or phrasal-prepositional verbs. If that is the case, what might these factors be? Is time spent in TL environment one of the determining factors for the Turkish-speaking learners' accurate and quantitative use of phrasal and phrasal-prepositional verbs? These questions leave the ground open for follow-up studies and point to the need of the identification and inclusion of more variables in order to explain variability in learner corpus data (see Möller's 2017).

Due to the nature of the data used, it was also not feasible to determine whether the learners were aware of the problems they experienced in multi-word verb production and whether they could correct the deviations in the multi-word verbs they produced if provided the chance. By the same token, as it was not possible to determine the exact meaning of what the learners actually aimed for with the verbs they produced,

the targeted verbs were suggested on the basis of the larger context by the researcher and in questionable and vague cases, native speakers were consulted. These weaknesses of the present study necessitate its complementation with experimental data. A think-aloud study of L2 writing or a post-task interview, for instance, would help determine the difficulties experienced as well as the strategies used by the learners to deal with the lack of lexical knowledge (or lexical retrieval process). Likewise, in the case of underuse attested as regards phrasal verbs, it was not possible to determine whether it resulted from the learners' lack of knowledge or rather due to the learners' predilection for one-word verbs to express the targeted meaning (maybe because of register issues or because they feel safer with simple verbs). Therefore, as suggested by Gilquin (2007), the results attained here should be supplemented by elicitation data for a better understanding of learners' collocational knowledge.

A further limitation of the study is also related to the nature of the data. The deviations attested in the data were accounted for by the reliance on either the 'open choice principle' or L1 patterns. By the same token, in the case of the correctly used multi-word verbs, especially the incongruent ones, we talked about the possibility of these units being produced as wholes (by relying on the idiom principle). However, since the present study is based on L2 learners' output, the validity of these assumptions as well as the inferences drawn about psycholinguistic aspects in the learners' mental lexicon can only be verified by means of

psycholinguistic experiments (see Durrant and Siyanova-Chanturia 2015).

One other limitation of the present study lies in the fact that the classification of multi-word verbs extracted from the data in terms of transparency and idiomaticity were not attempted. Therefore, conclusions as to the influence of the degree of idiomaticity of multi-word verbs on their learnability was not possible on the basis of the present investigation. It is worth conducting future experimental research to compare the performance of learners in learning transparent and idiomatic multi-word verbs in L2 in order to determine whether there is any difference with regard to their learnability.

Another interesting question left unanswered for further research is related to the ELT textbooks. The present study revealed an overall low number of verb types used in constructing multi-word verbs – a high degree of reliance on common verbs. This finding raises the question of to what extent the learners' behavior is a result of the textbooks used in (high) schools. Which multi-word verbs are presented in the books and how (often)? Are the multi-word verbs presented in corpus-driven rich contexts to teach all usage aspects? What are the selection criteria for the inclusion of the multi-word verbs in these books? What kinds of exercises are used to reinforce the learning of new multi-word verbs? Due to time constraints, no textbooks in the teaching of English in Turkey have been investigated to provide answers to these questions. To my knowledge, the only study to date that investigated the phrasal verbs that occur in ELT textbooks which are commonly used in Turkey was

conducted by Kartal (2018), in which he focused on four upper-intermediate textbooks. Kartal's study revealed that most of the phrasal verbs in these textbooks were not frequent in native speaker corpora (namely, COCA and BNC) while some were extremely rare, e.g., *chuck out*, *mull over*, *make off*, etc. Kartal also found that the number of phrasal verbs introduced in the textbooks ranged between 38 and 49 and there was no "lexical agreement" among them, i.e. none of the phrasal verbs attested appeared in all four of the textbooks. Kartal (2018) noted that phrasal verb use was emphasized only in a few chapters or presented in "highlighted" sections in the textbooks, where the aim was to teach vocabulary and grammar. Only a few phrasal verbs appeared somewhere else, e.g. in reading passages. However, Kartal did not report how phrasal verbs were presented and dealt with in the textbooks; whether they were presented only for their most common meaning or in different contexts for different meanings, whether there were any exercises to practice them after they had been presented, etc. Another weakness of Kartal's study is that, as the researcher himself acknowledged, the focus was on one proficiency level. A more extensive study needs to be carried out on a wider number of textbooks and range of proficiency levels. In most state schools, between grades 3 and 8, the ELT books prepared by the Ministry of Education are used. These books, which have been specifically designed for Turkish-speaking learners, in addition to the other ELT books in common use in high schools in Turkey, should be further investigated to determine to which extent the multi-word verbs are integrated into curricula. Answers to

the above-mentioned questions and similar ones would be the first step in the endeavor to improve of teaching materials (in use).

As mentioned earlier, the learner essays investigated were written untimed, allowing the learners to work on their text. The results attained here could also be compared with spoken, i.e. more spontaneous data. This would allow us to gain a fuller understanding of the difficulty these verbs create for the learner group under scrutiny.

Despite the high number of deviations, the Turkish learners under scrutiny displayed correct use of a considerable number of multi-word verbs, some of which are incongruent and restricted in terms of collocability and syntactic use. This fact raises the question as to what accounts for the success in these cases. In other words, what differentiates them from the multi-word verbs with similar features which could not be produced appropriately by the learners. This question needs further research since fixedness and frequency of the multi-word units do not seem to explain the observed discrepancy between the multi-word units in the data. Other potential factors – such as the amount and type of input, the communicational needs of the learner, the learner's proficiency, instruction method (including the exercises used, which may draw the learner's attention to lexical patterns), degree of transparency/idiomaticity, an interplay of these and similar variables) necessitate further research.

As mentioned in earlier sections, half of the learners in TICLE were allowed to make use of reference tools (e.g., dictionaries) during the

writing task but no information was provided on which reference tools the learners consulted nor on which verbs the learners consulted reference tools for, etc. To date, as noted by Manchon et al. (2007), not much attention has been paid to how L2 writers make use of dictionaries while composing their texts. It would be worth expanding the empirical inquiry in dictionary use (both bilingual and monolingual) in connection with the use of multi-word verbs by L2 learners. This would allow us to determine not only the learners' dictionary skills but also the (in)efficiency of different dictionaries available.

Although a possible trace of interlanguage transfer (i.e., the influence of one L2 over another (Gass & Selinker 2008: 519)) was observed in the analysis of the data, this issue was not investigated to the full extent in the present study. Therefore, further investigation may want to establish the role of interlanguage transfer in the use of multi-word verbs by Turkish-speakers of English.

The present study has firmly established the difficulty posed to Turkish speakers of English by multi-word verbs. Clearly this study refers to only one L2 language group, but, as the literature review has shown (see Section 2.3.4), Turkish learners do not differ greatly from other learners in the difficulties they experience in using multi-word verbs in English. Given their importance in language acquisition and use (see Section 2.3.3) and the recurrently made claim in the earlier studies that they are difficult if not impossible to acquire by mere exposure, examination of

how instruction can best assist multi-word verb acquisition by L2 learners is of utmost importance in development of pedagogical excellence in language teaching.

Despite the limitations pointed out above, the present study contributes to our current understanding of multi word verb production in L2 English by Turkish learners. Although the focus was on this specific learner group, the findings have general implications for improving the teaching of such units.

Bibliography

- Aas, Hege Larson. 2011. *Recurrent word-combinations in spoken learner English. A study of corpus data from Swedish and Norwegian advanced learners*. MA Thesis. Oslo. University of Oslo.
- Acklam, Richard. 1992. *Help with phrasal verbs*. Oxford: Heinemann.
- Adolphs, Svenja and Valerie Durow. 2004. 'Social-cultural integration and the development of formulaic sequences'. In Norbert Schmitt (ed.), *Formulaic sequences* (pp. 107–126). Amsterdam: John Benjamins.
- Aitchison, Jean. 1987. 'Reproductive furniture and extinguished professors'. In Ross Steele and Terry Threadgold (eds.), *Language topics. Essays in honour of Michael Halliday*. Vol. II (pp. 3-14). Amsterdam/Philadelphia: John Benjamins.
- Alangari, Manal Abdulrahman. 2019. A corpus-based study of verb-noun collocations and verb complementation clause structures in the writing of advanced Saudi learners of English. Doctoral dissertation. Reading, Berkshire. University of Reading.
- Albrechtsen, Dorte, Birgit Henriksen and Claus Færch. 1980. Native speakers' reactions to learners' spoken interlanguage. *Language Learning* 30(2): 365-396.
- Alexander, Richard. 1984. Fixed expressions in English: Reference books and the teacher. *ELT Journal* 38(2): 127-134.

¹⁵⁸ Note that all internet sources were correct as of November 15, 2020.

-
- Algeo, John. 1995. 'Having a look at the expanded predicate'. In Bas Aarts and Charles F. Meyer (eds.), *The verb in contemporary English* (pp. 203-217). Cambridge: Cambridge University Press.
- Allerton, D. J. 1984. Three (or four) levels of word cooccurrence restriction. *Lingua* 63(1): 17-40.
- Al-Otaibi, Ghuzayyil Mohammed. 2019. A cognitive approach to the instruction of phrasal verbs: Rudzka-Ostyn's Model. *Journal of Language and Education* 5(2): 10-25.
- Altenberg, Bengt. 1991. 'Amplifier collocations in spoken English'. In Stig Johansson and Anna-Brita Stenström (eds.), *English computer corpora. Selected papers and research guide* (pp. 127-147). Berlin: Mouton de Gruyter.
- _____. 1998. 'On the phraseology of spoken English. The evidence of recurrent word-combinations'. In Anthony Paul Cowie (ed.), *Phraseology. Theory, analysis, and applications* (pp. 101-122). Oxford: Oxford University Press.
- Altenberg, Bengt and Mats Eeg-Olofsson. 1990. 'Phraseology in spoken English: Presentation of a project'. In Jan Aarts and Willem Meijs (eds.), *Theory and practice in corpus linguistics* (pp. 1-26). Amsterdam: Rodopi.
- Altenberg, Bengt and Sylviane Granger. 2001. The grammatical and lexical patterning of MAKE in native and non-native student writing. *Applied Linguistics* 22(2): 173-195.
- Anthony, Edward M. Jr. 1954. An exploratory inquiry into lexical clusters. *American Speech* 29(3): 175-180.
- Anthony, Laurence. 2005. AntConc: Design and development of a free-ware corpus analysis toolkit for the technical writing classroom. *2005 IEEE International Professional Communication Conference Proceedings* (pp. 729-737).

-
- Appel, René. 1996. 'The lexicon in second language acquisition'. In Peter Jordens and Josine Lalleman (eds.), *Investigating second language acquisition* (pp. 381-403). Berlin/New Work: Mouton de Gruyter.
- Arabski, Janusz. 1979. *Errors as indications of the development of interlanguage*. Katowice: Uniwersytet Śląski.
- Arik, Beril T. 2020. English in Turkey: A sociolinguistic profile. *World Englishes* 39:514–527. <https://doi.org/10.1111/weng.12488>
- Ard, Josh and Taco Homburg. 1992 [1983]. 'Verification of language transfer'. In Susan M. Gass and Larry Selinker (eds.), *Language transfer in language learning* (pp. 47-70). Amsterdam: John Benjamins.
- Armstrong, Kevin. 2004. Sexing up the dossier: A semantic analysis of phrasal verbs for language teachers. *Language Awareness* 13(4): 213-224.
- Arnold, Jennifer E., Thomas Wasow, Anthony Losongco and Ryan Ginstrom. 2000. Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language* 76(1): 28-55.
- Arnon, Inbal and Eve V. Clark. 2011. Why *brush your teeth* is better than *teeth* – Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development* 7(2): 107–129.
- Arnon, Inbal and Morten H. Christiansen. 2014. 'Chunk-based language acquisition'. In Patricia J. Brooks and Vera Kempe (eds.), *Encyclopedia of language development* (pp. 88-90). Thousand Oaks, CA: Sage Publications.

-
- Arnon, Inbal and Neal Snider. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62(1): 67-82.
- Aslin, Richard N., Julide Z. Woodward, Nicholas P. LaMendola, and Thomas G. Bever. 1996. 'Models of word segmentation in fluent maternal speech to infants'. In James L. Morgan and Katherine Demuth (eds.), *Signal to syntax. Bootstrapping from speech to grammar in early acquisition* (pp. 117-134). Mahwah, NJ: Lawrence Erlbaum Associates.
- Astika, Gusti Gede. 1993. Analytical assessment of foreign students' writing. *RELC Journal* 24(1): 61-72.
- Atkins, Sue, Jeremy Clear and Nicholas Ostler. 1992. Corpus design criteria. *Literary and Linguistic Computing* 7(1): 1-16.
- Atkins, B. T. Sue, Beth Levin and Antonio Zampolli. 1994. 'Computational approaches to the lexicon: An overview'. In B. T. Sue Atkins and Antonio Zampolli (eds.), *Computational approaches to the lexicon* (pp. 17-45). New York: Oxford University Press.
- Babanoğlu, M. Pinar. 2014. A corpus-based study on the use of MAKE by Turkish EFL learners. *International Journal of Education & Literary Studies* 2(2): 43-47.
- Bacchielli, Rolando. 1993. Syntheticity and analyticity in the syntactic make-up of English. In Maurizio Gotti (ed.), *English diachronic syntax* (pp. 55-66). Milano: Guerini Studio.
- Baddeley, Alan. 2000. The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences* 4(11): 417-423.
- Bahns, Jens. 1993. Lexical collocations: A contrastive view. *ELT Journal* 47(1): 56-63.
- Bahns, Jens and Moira Eldaw. 1993. Should we teach EFL collocation? *System* 21(1): 101-114.

-
- Bamgbose, Ayo. 2003. A recurring decimal: English in language policy and planning. *World Englishes* 22(4): 419-431. ^[1]_[SEP]
- Bannard, Colin and Danielle Matthews. 2008. Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19(3): 241-248.
- Bardovi-Harlig, Kathleen. 2002. A new starting point? Investigating formulaic use and input in future expression. *Studies in Second Language Acquisition* 24(2): 189-198.
- Barlow, Michael. 2005. 'Computer-based analyses of learner language'. In Rod Ellis and Gary Barkhuizen (eds.), *Analysing learner language* (pp. 335-357). Oxford: Oxford University Press.
- Barnbrook, Geoff. 1996. *Language and computers: a practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.
- Barton, Michelle E. and Michael Tomasello. 1994. 'The rest of the family: The role of fathers and siblings in early language development'. In Clare Gallaway and Brian J. Richards (eds.), *Input and interaction in language acquisition* (pp. 109-134). Cambridge: Cambridge University Press.
- Bates, Elizabeth, Inge Bretherton and Lynn Snyder. 1988. *From first words to grammar: Individual differences and dissociable mechanisms*. Cambridge: Cambridge University Press.
- Bayyurt, Yasemin. 2012. 'Proposing a model for English language education in the Turkish sociocultural context'. In Yasemin Bayyurt and Yeşim Bektaş-Çetinkaya (eds.), *Research perspectives on teaching and learning English in Turkey: Policies and practices* (pp. 301-312). Frankfurt am Main: Peter Lang.

-
- Bayyurt, Yasemin and Yeşim Bektaş-Çetinkaya (eds.) 2012. *Research perspectives on teaching and learning English in Turkey: Policies and practices*. Frankfurt am Main: Peter Lang.
- Beekhuizen, Barend, Rens Bod and Willem Zuidema. 2013. Three design principles of language: The search for parsimony in redundancy. *Language and Speech* 56(3): 265-290.
- Behrens, Heike. 2009. Usage-based and emergentist approaches to language acquisition. *Linguistics* 47(2): 383-411.
- Bensoussan, Marsha and Batia Laufer. 1984. Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading* 7(1): 15-32.
- Bestgen, Yves and Sylviane Granger. 2014. Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing* 26: 28-41.
- Biber, Douglas, Stig Johannsson, Geoffrey Leech, Susan Conrad and Edward Finegan. 1999. *Longman grammar of spoken and written English*. Harlow: Longman.
- Biskup, Danuta. 1992. 'L1 Influence on learners' renderings of English collocations.' In Pierre J. L. Arnaud and Henri Béjoint (eds.), *Vocabulary and applied linguistics* (pp. 85-93). London: Macmillan.
- Blaas, L. 1982. *Fossilization in the advanced learner's lexicon*. Unpublished manuscript. Department of English. University of Utrecht.
- Bley-Vroman, Robert. 1983. The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33(1): 1-17.
- Bloomfield, Leonard. 1933. *Language*. New York: Holt.
- Bloom, Lois, Patsy Lightbown and Lois Hood. 1975. Structure and variation in child language. *Monographs of the society for research in child development* 40(2): 1-97.

-
- Blum, Shoshana and E. A. Levenston. 1978. Universals of lexical simplification. *Language Learning* 28(2): 399-415.
- Blum, Shoshana and E. A. Levenston. 1980. Lexical simplification in second language acquisition. *Studies in Second Language Acquisition* 2(2): 43-63.
- Bod, Rens. 1998. *Beyond grammar: An experience-based theory of language*. CA: Center for the Study of Language and Information.
- _____ 2009. From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science* 33(5): 752-793.
- Boers, Frank, June Eyckmans, Jenny Kappel, H el ene Stengers and Murielle Demecheleer. 2006. Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research* 10(3): 245-261.
- Bogaards, Paul. 1996. 'Lexicon and grammar in second language learning'. In Peter Jordens and Josine Lalleman (eds.), *Investigating second language acquisition*. (pp. 357-380). Berlin/New York: Mouton de Gruyter.
- Bohn, Ocke-Schwen. 1986. Formulas, frame structures, and stereotypes in early syntactic development: Some new evidence from L2 acquisition. *Linguistics* 24(1): 185-202.
- Bolinger, Dwight. 1961. Syntactic blends and other matters. *Language* 37: 366-381. [SEP]
- _____ 1971. *The phrasal verb in English*. Cambridge, MA: Harvard University Press.
- _____ 1975. *Aspects of language* (2nd ed.) New York: Harcourt Brace Jovanovich.
- _____ 1976. Meaning and memory. *Forum Linguisticum* 1(1): 1-14.

-
- _____ 1979. 'Meaning and memory'. In George G. Haydu (ed.), *Experience forms. Their cultural and individual place and function* (pp. 95-111). Paris/New York: Mouton.
- Bolton, Kingsley, Gerald Nelson and Joseph Hung. 2002. A corpus-based study of connectors in student writing: Research from the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics* 7(2): 165-182.
- Braine, Martin D. S. 1976. Children's first word combinations. With commentary by Melissa Bowerman. *Monographs of the society of research in child development*. 41(1): 1-104.
- Brent, Michael R. and Jeffrey Mark Siskind. 2001. The role of exposure to isolated words in early vocabulary development. *Cognition* 81(2): B33-B44.
- Bretherton, Inge, Sandra McNew, Lynn Synder and Elizabeth Bates. 1983. Individual differences at 20 months: Analytic and holistic strategies in language acquisition. *Journal of Child Language* 10(2): 293-320.
- Brown, Roger. 1973. *A first language: The early stages*. Cambridge, MA: Harvard University Press.
- Brown, Roger and David McNeill. 1966. The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior* 5: 325-337.
- Brown, Roger and Camille Hanlon. 1970. 'Derivational complexity and order of acquisition in child speech'. In John R. Hayes (ed.), *Cognition and the development of language* (pp. 11-53). New York: John Wiley
- Brumfit, Christopher J. and Keith Johnson (eds.) 1979. *The communicative approach to language teaching*. Oxford: Oxford University Press.

Bybee, Joan. 1998. 'The emergent lexicon'. In M. C. Gruber, D. Higgins, K. S. Olson and T. Wysocki (eds.), *Proceedings of the Chicago Linguistic Society 34: The Panels* (pp. 421–435). Chicago, IL: Chicago Linguistic Society.

_____. 2002. Phonological evidence for exemplar storage of multi-word sequences. *Studies in Second Language Acquisition* 24(2): 215–221.

_____. 2010. *Language, usage and cognition*. New York: Cambridge University Press.

Bybee, Joan and Joanne Scheibman. 1999. The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics: An Interdisciplinary Journal of the Language Sciences* 37(4): 575-596.

Bywater, F. V. 1969. *A proficiency course in English*. London: University of London Press.

Canale, Michael. 1983. 'From communicative competence to communicative language pedagogy'. In Jack C. Richards and Richard W. Schmidt (eds.), *Language and communication* (pp. 2-27). New York: Longman.

Canale, Michael and Merrill Swain. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1(1): 1-47.

Can, Cem. 2009. İkinci dil edinimi çalışmalarında bilgisayar destekli bir Türk öğrenici İngilizcesi Derlemi: ICLE'nin bir alt derlemi olarak TICLE [A Turkish learner corpus of English in second language studies: TICLE as a subcorpus of ICLE]. *Dil Dergisi* 144: 16-34.

Candlin, Christopher N. (ed.). 1981. *The communicative teaching of English. Principles and an exercise typology*. London: Longman.

-
- Carlsen, Cecilie. 2012. Proficiency level — A fuzzy variable in computer learner corpora. *Applied Linguistics* 33(2): 161–183.
- Carter, Ronald. 1987. Is there a core vocabulary? Some implications for language teaching. *Applied Linguistics* 8(2): 178-193.
- _____. 2012. *Vocabulary. Applied linguistic perspectives* (2nd ed.). Oxon/New York: Routledge.
- Carter, Ronald and Michael McCarthy. 1988. *Vocabulary and language teaching*. New York: Longman.
- Celce-Murcia, Marianne and Diane Larsen-Freeman. 1999. *The grammar book: An ESL/EFL teacher's course* (2nd ed.). Boston: Heinle & Heinle.
- Chafe, Wallace L. 1980. 'Some reasons for hesitating'. In Hans W. Dechert and Manfred Raupach (eds.), *Temporal variables in speech. Studies in honour of Frieda Golman-Eisler* (pp. 169-180). The Hague: Mouton Publishers.
- Channell, Joanne. 1988. 'Psycholinguistic considerations in the study of L2 vocabulary acquisition'. In Ronald Carter and Michael McCarthy (eds.), *Vocabulary and language teaching* (pp. 83-96). New York: Longman.
- Chen, Meilin. 2013a. Overuse or underuse. A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics* 18(3): 418-442.
- _____. 2013b. 'Phrasal verbs in a longitudinal learner corpus: Quantitative findings'. In Sylviane Granger, Gaëtanelle Gilquin, Fanny Meunier (eds.), *Twenty years of learner corpus research: Looking back, moving ahead. Corpora and Language in Use – Proceedings 1* (pp. 89-101). Louvain-la-Neuve: Presses universitaires de Louvain.

-
- _____ 2017. 'Beyond frequencies: Investigating the semantic and the stylistic features of phrasal verbs in a three-year longitudinal corpus by Chinese university students'. In Vaclav Brezina and Lynne Flowerdew (eds.), *Learner corpus research: New perspectives and applications* (pp. 134-154). London/New York: Bloomsbury.
- Chen, Yu Hua and Paul Baker. 2010. Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology* 14(2): 30-49.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge: MIT Press.
- _____ 1970. 'Remarks on nominalization'. In Roderick A. Jacobs and Peter S. Rosenbaum (eds.), *Readings in English Transformational Grammar* (pp. 184-221). Boston: Ginn.
- Claridge, Claudia. 2000. *Multi-word verbs in Early Modern English: A Corpus-based study*. Amsterdam/Atlanta: Rodopi.
- Clark, Ruth. 1974. Performing without competence. *Journal of Child Language* 1(1): 1-10.
- Conklin, Kathy and Norbert Schmitt. 2008. Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics* 29(1): 72-89.
- Cook, Vivian. 1997. Monolingual bias in second language acquisition research. *Revista Canaria de Estudios Ingleses* 34: 35-50.
- Corder, Stephen Pit. 1967. The significance of learner's errors. *International Review of Applied Linguistics* 5(4): 161-170.
- _____ 1974. 'Error analysis'. In John Patrick Allen and Stephen Pit Corder (eds.), *The Edinburg course in applied linguistics volume 3: Techniques in applied linguistics*. Oxford: Oxford University Press.
- _____ 1981. *Error analysis and interlanguage*. Oxford: Oxford University Press.

-
- Cornell, Alan. 1985. Realistic goals in teaching and learning phrasal verbs. *International Review of Applied Linguistics in Language Teaching* 23(4): 269-280.
- Cortes, Viviana. 2002. 'Lexical bundles in Freshman composition'. In Randi Reppen, Susan M. Fitzmaurice and Douglas Biber (eds.), *Using corpora to explore linguistic variation* (pp. 131-145). Amsterdam/Philadelphia: John Benjamins.
- Cotos, Elena. 2014. Enhancing writing pedagogy with learner corpus data. *ReCALL* 26 (2): 202-224.
- Coulmas, Florian. 1981. 'Introduction: Conversational routine'. In Florian Coulmas (ed.), *Conversational routine. Explorations in standardized communication situations and prepatterned speech* (pp. 1-17). The Hague: Mouton Publishers.
- Cowan, Nelson. 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24(1): 87-114.
- Cowan, Ron. 2008. *The teacher's grammar of English. A course book and reference guide*. New York: Cambridge University Press.
- Cowie, Anthony Paul. 1988. 'Stable and creative aspects of vocabulary use'. In Ronald Carter and Michael McCarthy (eds.), *Vocabulary and language teaching* (p. 126-137). London: Longman.
- _____ 1992. 'Multiword lexical units and communicative language teaching'. In Pierre J. L. Arnaud and Henri Béjoint (eds.), *Vocabulary and applied linguistics* (pp. 1-12). Houndmills: MacMillan.
- _____ 1994. 'Phraseology'. In Ronald E. Asher (ed.), *The encyclopedia of language and linguistics* (pp. 3168-3171). Oxford: Pergamon Press.

-
- _____ 1998a. 'Phraseological dictionaries: Some east-west comparisons.' In Anthony Paul Cowie (ed.), *Phraseology. Theory, analysis, and applications* (pp. 209-228). Oxford: Oxford University Press.
- _____ 1998b. 'Introduction'. In Anthony Paul Cowie (ed.), *Phraseology. Theory, analysis, and applications* (pp. 1-22). Oxford: Oxford University Press.
- _____ 2001. 'Speech formulae in English: Problems of analysis and dictionary treatment'. In Geert van der Meer and Alice G. B. ter Meulen (eds.), *Making sense: From lexeme to discourse. In Honor of Werner Abraham* (p. 1-12). [Groningen Arbeiten zur germanistischen Linguistik 44]. Groningen: Center for Language and Cognition.
- Cowie, Anthony P. and Ronald Mackin. 1975. *Oxford dictionary of current idiomatic English. Volume 1. Verbs with prepositions and particles*. Berlin: Cornelsen/Oxford University Press.
- Cowie, Anthony Paul and Ronald Mackin. 1993. *Oxford dictionary of phrasal verbs*. Oxford: Oxford University Press.
- Cowie, Anthony Paul and Peter Howarth. 1996. 'Phraseological competence and written proficiency'. In George M. Blue and Rosamond Mitchell (eds.), *Language and education. Papers from the annual meeting of the British Association for Applied Linguistics held at the University of Southampton, September 1995* (pp. 80-93). Clevedon: Multilingual Matters.
- Crossley, Scott, Tom Salsbury and Danielle S. McNamara. 2014. Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics* 36(5): 1-22.
- Crystal, David. 1997. *The Cambridge encyclopedia of language* (2nd ed.) Cambridge: Cambridge University Press.

-
- Cumming, Alister. 1990. Metalinguistic and ideational thinking in second language composing. *Written Communication* 7(4): 482–511.
- Dagut, M. B. 1977. Incongruencies in lexical “gridding” – An application of contrastive semantic analysis to language teaching. *International Review of Applied Linguistics in Language Teaching* 15(3): 221-229.
- Dagut, Menachem and Batia Laufer. 1985. Avoidance of phrasal verbs – A case for contrastive analysis. *Studies in Second Language Acquisition* 7(1): 73-79.
- Darus, Saadiyah and Kaladevi Subramaniam. 2009. Error analysis of the written English essays of secondary students in Malaysia: A case study. *European Journal of Social Sciences* 8(3): 483-495.
- Darwin, Clayton M. and Loretta S. Gray. 1999. Going after the phrasal verb: An alternative approach to classification. *TESOL Quarterly* 33(1): 65-83.
- Davies, Alan. 1991. *The native speaker in applied linguistics*. Edinburgh: Edinburgh University Press.
- _____. 2003. *The native speaker: Myth and reality*. Clevedon: Multilingual Matters.
- De Cock, Sylvie. 2000. ‘Repetitive phrasal chunkiness and advanced EFL speech and writing’. In Christian Mair and Marianne Hundt (eds.), *Corpus linguistics and linguistic theory* (pp. 51-68). Amsterdam: Rodopi.
- _____. 2005. ‘Learners and phrasal verbs’. In Michael Rundell (ed. in chief), *Macmillan phrasal verbs plus (LS16-LS20)*. Oxford: Macmillan.
- De Haan, Pieter. 1997. ‘An experiment in English learner data analysis’. In Jan Aarts, Inge de Mönning and Herman Wekker (eds.),

-
- Studies in English language and teaching. In Honour of Flor Aarts* (pp. 215-229). Amsterdam/Atlanta: Rodopi.
- Dechert, Hans W. 1983. 'How a story is done in a second language'. In Claus Færch and Gabriele Kasper (eds.), *Strategies in interlanguage communication* (pp. 175-196). London: Longman.
- Dechert, Hans W. 1984. 'Second language production: Six hypotheses'. In Hans W. Dechert, Dorothea Möhle and Manfred Raupach (eds.), *Second language productions* (pp. 211-230). Tübingen: Gunter Narr Verlag.
- Declerck, Renaat. 1977. Some arguments in favor of a generative semantics analysis of sentences with an adverbial particle or a prepositional phrase of goal. *Orbis* 26: 297-340.
- Demiray Akbulut, Fatma. 2018. The avoidance behaviour of Turkish EFL learners in using multi-word verbs. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi* 46: 74-94.
- Denison, David. 1981. *Aspects of the history of English group-verbs. With particular attention to the syntax of the Ormulum*. Published doctoral dissertation. Oxford: Oxford University.
- Denison, David. 1984. On *get it over with*. *Neophilologus* 68: 271-277.
- Di Pietro, Robert J. 1980. 'Verbal strategies: A neglected dimension in language acquisition studies'. In Hans W. Dechert and Manfred Raupach (eds.), *Temporal variables in speech. Studies in honour of Frieda Goldman-Eisler* (pp. 313-321). The Hague: Mouton.
- Dik, Simon C. 1978. *Functional grammar*. Amsterdam: North-Holland.
- Dixon, R. M. W. 1982. The grammar of English phrasal verbs. *Australian Journal of Linguistics* 2(1): 1-42.
- Dogancay-Aktuna, Seran. 1998. The spread of English in Turkey and its current sociolinguistic profile. *Journal of Multilingual and Multicultural Development* 19(1): 24-39.

-
- Dose-Heidelmayer, Stefanie and Sandra Götz. 2016. The progressive in spoken learner language: A corpus-based analysis of use and misuse. *IRAL* 54(3): 229-256.
- Dörnyei, Zoltán and Judit Kormos. 1998. Problem-solving mechanisms in L2 communication. A psycholinguistic perspective. *Studies in Second Language Acquisition* 20(3): 349-385.
- Dulay, Heidi C. and Marina K. Burt. 1973. Should we teach children syntax? *Language Learning* 23(2): 245-258.
- Dulay, Heidi C. and Bates Marina K. Burt. 1974. Errors and strategies in child second language acquisition. *TESOL Quarterly* 8(2): 129-136.
- Durrant, Philip. 2008. High frequency collocations and second language learning. Doctoral dissertation. University of Nottingham.
- Durrant, Philip and Norbert Schmitt. 2009. To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching* 47(1): 157–177.
- Durrant, Philip and Norbert Schmitt. 2010. Adult learners' retention of collocations from exposure. *Second Language Research* 26(2): 163-188.
- Durrant, Philip and Anna Siyanova-Chanturia. 2015. 'Learner corpora and psycholinguistics'. In Sylviane Granger, Gaëtanelle Gilquin and Fanny Meunier (eds.), *The Cambridge handbook of learner corpus research* (pp. 57-78). Cambridge: Cambridge University Press.
- Eddine, Abeer Naser. 2012. Interlanguage prepositions: An analysis of French learners' productions in L2 English. Doctoral dissertation. Grenoble: Université de Grenoble. <NNT: 2012GRENL014>. <tel-01057881>

-
- Education First English Proficiency Index. 2020. Turkey. Retrieved October 21, 2021 from <https://www.ef.edu/epi/regions/europe/turkey/>
- Edwards, Alison and Rutger-Jan Lange. 2016. *In case of innovation. Academic phraseology in the Three Circles. International Journal of Learner Corpus Research* 2(2): 252-277.
- Edwards, Alison and Samantha Laporte. 2015. Outer and expanding circle Englishes: The competing roles of norm orientation and proficiency levels. *English World-Wide* 36(2): 135-169.
- Elkılıç, Gencer. 2012. Mother tongue traces of Turkish university students on composition papers written in English. *Procedia-Social and Behavioral Sciences* 47: 656–664.
- Ella, Jennibelle R. and Shirley N. Dita. 2017. The phrasal-prepositional verbs in Philippine English: A corpus-based analysis. *Proceedings of the 31st Pasific Asia Conference on Language, Information and Computation* (pp. 34-41). Philippines.
- Ellis, Nick C. 1996. Sequencing in SLA: Phonological memory, chunking and points of order. *Studies in Second Language Acquisition* 18(1): 91-126.
- Ellis, Rod. 1985. *Understanding second language acquisition*. Oxford: Oxford University Press.
- _____. 1994. *The study of second language acquisition*. Oxford: Oxford University Press.
- _____. 2002. 'The place of grammar instruction in the second/foreign language curriculum'. In Eli Hinkel and Sandra Fotos (eds.), *New perspectives on grammar teaching in second language classrooms* (pp. 17-34). Mahwah, NJ: Lawrence Erlbaum.
- _____. 2005. Principles of instructed language learning. *System* 33(2): 209–224.

-
- Ellis, Rod and Gary Patrick Barkhuizen. 2005. *Analysing learner language*. Oxford: Oxford University Press.
- Engber, Cheryl A. 1995. The relationship of lexical proficiency to the quality of ESL compositions. *Journal of Second Language Writing* 4(2): 139-155.
- Erkaya, Odiléa Rocha. 2012. Vocabulary and L1 interference – Error analysis of Turkish students’ English essays. *Mextesol Journal* 36(2): 1–11.
- Erman, Britt and Beatrice Warren. 2000. The idiom principle and the open choice principle. *Text & Talk* 20(1): 29-62.
- Eurobarometer report of European Commission in 2006 titled “Europeans and their languages” (http://ec.europa.eu/public_opinion/archives/ebs/ebs_243_en.pdf).
- Fadanelli, Sabrina Bonqueves. 2012. A corpus study on Brazilian learners’ usage of English phrasal verbs. *Estudos Anglo-Americanos* 37: 158-176.
- Faerch, Claus, Kirsten Haastrup and Robert Phillipson. 1984. *Learner language and language learning*. Clevedon, Avon: Multilingual Matters.
- Farghal, Mohammed and Hussein Obiedat. 1995. Collocations: A neglected variable in EFL. *International Review of Applied Linguistics*, 33(4): 315-331.
- Farsani, Hussein Muhammadi, Ahmad Moinzadeh and Mansoor Tavakoli. 2012. Mnemonic effectiveness of CL-motivated picture-elucidation tasks in foreign learners’ acquisition of English phrasal verbs. *Theory and Practice in Language Studies* 2(3): 498-509.
- Felix, Sascha W. 1981. The effect of formal instruction on second language acquisition. *Language Learning* 31(1): 87-112_[SEP].

-
- Ferguson, Charles A. 1982. 'Foreword'. In Braj B. Kachru (ed.), *The other tongue. English across cultures* (vii-xi). Oxford: Pergamon.
- Field, Yvette and Yip Lee Mee Oi. 1992. A comparison of internal conjunctive cohesion in the English essay writing of Cantonese speakers and native speakers of English. *RELC Journal* 23(1): 15-28.
- Fillmore, Charles J. 1976. The need for a frame semantics within linguistics. *Statistical Methods in Linguistics* 12: 5-29.
- _____ 1979. 'On fluency'. In Charles J. Fillmore, Daniel Kempler and William S. Y. Wang (eds.), *Individual differences in language ability and language behavior* (pp. 85-101). New York: Academic Press.
- Fillmore, Charles J., Paul Kay and Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64(3): 501-538.
- Firth, John R. 1957. *Papers in linguistics: 1934- 1951*. London: Oxford University Press.
- _____ 1968. 'Linguistic analysis as a study of meaning'. In F. R. Palmer (ed.), *Selected papers of J. R. Firth 1952-59* (pp. 12-26). London: Longmans.
- Fletcher, Bryan. 2005. 'Register and phrasal verbs'. In Michael Rundell (ed. in chief), *Macmillan phrasal verbs plus* (LS13-LS15). Oxford: Macmillan.
- Flowerdew, Lynne. 2015. 'Learner corpora and language for academic and specific purposes'. In Sylviane Granger, Gaëtanelle Gilquin and Fanny Meunier (eds.), *The Cambridge handbook of learner corpus research* (pp. 465-484). Cambridge: Cambridge University Press.
- Francis, W. Nelson. 1958. *The structure of American English*. New York: The Ronald

-
- Press Company.
- Fraser, Bruce. 1976. *The verb-particle combination in English*. Tokyo: Taishukan Publishing Company.
- Fries, Charles C. 1945. *Teaching & learning English as a foreign language*. Ann Arbor: The University of Michigan Press.
- Gabryś-Biskup, Danuta. 1990. 'Some remarks on combinability: Lexical collocations'. In Janusz Arabski (ed.), *Foreign language acquisition papers* (pp. 31-45). Katowice: Uniwersytet Slaski.
- Gardner, Dee and Mark Davies. 2007. Pointing out frequent phrasal verbs: A corpus-based analysis. *TESOL Quarterly* 41(2): 339-359.
- Garnier, Mélodie and Norbert Schmitt. 2015. The PHaVE list: A pedagogical list of phrasal verbs and their most frequent meaning senses. *Language Teaching Research* 19(6): 645-666.
- Garnier, Mélodie and Norbert Schmitt. 2016. *Picking up* polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System* 59: 29-44.
- Gass, Susan M. 1988. Second language vocabulary acquisition. *Annual Review of Applied Linguistics* 9: 92-106.
- Gass, Susan M. and Larry Selinker. 2008. *Second language acquisition. An introductory course* (3rd ed.). New York: Routledge.
- Gass, Susan M., Jennifer Behney and Luke Plonsky. 2013. *Second language acquisition. An introductory course* (4th ed.). New York: Routledge.
- Gelderen, Elly van. 2010. *An introduction to the grammar of English*. Amsterdam: John Benjamins.
- Gilquin, Gaëtanelle. 2007. To err is not all. What corpus and elicitation can reveal about the use of collocations by learners. *Zeitschrift für Anglistik und Amerikanistik* 55(3): 273-291.

-
- _____. 2011. 'Corpus linguistics to bridge the gap between World Englishes and Learner Englishes'. In Leonel Ruiz Miyares and María Rosa Álvarez Silva (eds.), *Comunicación social en el siglo XXI, Vol. II* (pp. 638–642). Santiago de Cuba: Centro de Lingüística Aplicada.
- _____. 2015. The use of phrasal verbs by French-speaking EFL learners. A constructional and collostructural corpus-based approach. *Corpus Linguistics and Linguistic Theory* 11(1): 51-88.
- Gilquin, Gaëtanelle and Magali Paquit. 2008. Too chatty. Learner academic writing and register variation. *English Text Construction* 1(1): 41-61.
- Gilquin, Gaëtanelle and Sylviane Granger. 2011. 'From EFL to ESL: Evidence from the *International Corpus of Learner English*'. In Joybrato Mukherjee and Marianne Hundt (eds.), *Exploring second-language varieties of learner Englishes. Bridging a paradigm gap* (pp. 55-78). Amsterdam/Philadelphia: John Benjamins.
- Gilquin, Gaëtanelle, Sylviane Granger and Magali Paquot. 2007. 'Improve your writing skills (Writing sections)'. In Michael Rundell (ed.), *Macmillan English dictionary for advanced learners* (2nd ed.) (IW1–IW29). Oxford: Macmillan Education.
- Gitsaki, Christina. 1999. *Second language lexical acquisition: A study of the development of collocational knowledge*. San Francisco: International Scholars Publications.
- Goad, Heather and Lydia White. 2009. 'Articles in Turkish/English interlanguage revisited. Implications of vowel harmony'. In María del Pilar García Mayo and Roger Hawkins (eds.), *Second language acquisition of articles. Empirical findings and theoretical implications* (pp. 201-232). Amsterdam/Philadelphia: John Benjamins.

-
- Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.
- _____ 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Goldinger, Stephen D. 1996. Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22(5): 1166–1183.
- Goulart, Larissa. 2019. The use of collocations across proficiency levels: A literature review. *Brazilian English Language Teaching Journal* 10(2): 1-15.
- Göksel, Asli and Celia Kerslake. 2005. *Turkish: A comprehensive grammar*. Oxon: Routledge.
- Götz, Sandra. 2015. ‘Tense and aspect errors in spoken learner English. Implications for language testing and assessment’. In Marcus Callies and Sandra Götz (eds.), *Learner corpora in language testing and assessment* (pp. 191-215). Amsterdam: John Benjamins.
- Granger, Sylviane. 1996a. ‘Learner English around the world.’ In Sidney Greenbaum (ed.), *Comparing English world-wide. The International Corpus of English* (pp. 13-24). Oxford: Clarendon.
- _____ 1996b. ‘From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora’. In Karin Aijmer, Bengt Altenberg and Mats Johansson (eds.), *Languages in contrast. Papers from a symposium on text-based cross-linguistic studies* (pp. 37-51). Lund: Lund University Press.
- _____ 1998a. ‘Prefabricated patterns in advanced EFL writing: Collocations and formulae’. In Anthony P. Cowie (ed.), *Phraseology:*

-
- Theory, analysis, and applications* (pp. 145-160). Oxford: Clarendon.
- _____ 1998b. 'The computer learner corpus: A versatile new source of data for SLA research'. In Sylviane Granger (ed.), *Learner English in computer* (pp. 3-18). London/New York: Addison Wesley Longman.
- _____ 2002. 'A bird's-eye view of learner corpus research'. In Sylviane Granger, Joseph Hung and Stephanie Petch-Tyson (eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3-33). Amsterdam/Philadelphia: John Benjamins.
- _____ 2009. 'The contribution of learner corpora and second language acquisition and foreign language teaching: A critical evaluation'. In Karin Aijmer (ed.), *Corpora and language teaching* (pp. 13-32). Amsterdam/Philadelphia: John Benjamins.
- _____ 2012a. 'How to use foreign and second language learner corpora'. In Alison Mackey and Susan M. Gass (eds.), *A guide to research methods in second language acquisition* (pp. 7-29). Malden: Basil Blackwell.
- _____ 2012b. 'Learner corpora'. In Carol A. Chapelle (ed.), *The encyclopedia of applied linguistics* (pp. 3235-3242). Oxford: Wiley-Blackwell.
- _____ 2015. Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research* 1(1): 7-24.
- Granger, Sylviane and Stephanie Tyson. 1996. Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes* 15(1): 17-27.
- Granger, Sylvia and Fanny Meunier (eds.). 2008. *Phraseology: An interdisciplinary perspective*. Amsterdam: John Benjamins.

-
- Granger, Sylviane and Maqali Paquot. 2008. 'Disentangling the phraseological web'. In Sylviane Granger and Fanny Meunier (eds.), *Phraseology: An interdisciplinary perspective* (pp. 27-49). Amsterdam/Philadelphia: Benjamins.
- Granger, Sylviane, Estelle Dagneaux, Fanny Meunier and Magali Paquot (eds.). 2009. *International Corpus of Learner English: Version 2*. Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.
- Granger, Sylviane, Gaëtanelle Gilquin and Fanny Meunier (eds.). 2015. *The Cambridge handbook of learner corpus research*. Cambridge: Cambridge University Press.
- Grauberg, W. 1971. An error analysis in German of first year university students. In G. E. Perren and J. L. M. Trim (eds.), *Applications of linguistics. Selected papers of the Second International Congress of Applied Linguistics, Cambridge 1969* (pp. 257-264). Cambridge: Cambridge University Press.
- Greenbaum, Sidney. 1996. *The Oxford English grammar*. Oxford: Oxford University Press.
- Gries, Stefan Th. 2015. 'Statistics for learner corpus research'. In Sylviane Granger, Gaëtanelle Gilquin and Fanny Meunier (eds.), *The Cambridge handbook of learner corpus research* (pp. 159-182). Cambridge: Cambridge University Press.
- Grobe, Cary. 1981. Syntactic maturity, mechanics, and vocabulary as predictors of quality ratings. *Research in the Teaching of English* 15(1): 75-85.
- Guo, Xiaotian. 2006. Verbs in the written English of Chinese learners: A corpus-based comparison between non-native speakers and native speakers. Doctoral dissertation. University of Birmingham.

-
- Guntermann, Gail. 1978. A study of the frequency and communicative effects of error in Spanish. *The Modern Language Journal* 62(5-6): 249-253.
- Hägglund, Monika. 2001. Do Swedish advanced learners use spoken language when they write in English? *Moderna Språk* 95(1): 2-8.
- Halliday, M. A. K. 1961. Categories of the theory of grammar, *Word* 17(2): 241-292.
- _____. 1966. 'Lexis as a Linguistic Level'. In C. E. Bazell, J. C. Catford, M.A.K. Halliday and R.H. Robins (eds.), *In memory of J.R. Firth* (pp. 148-162). London: Longmans.
- Hanania, Edith A. S. and Harry L. Gradman. 1977. Acquisition of English structures: A case study of an adult native speaker of Arabic in an English-speaking environment. *Language Learning* 27(1): 75-91.
- Harley, Birgit and Mary Lou King. 1989. Verb lexis in the written compositions of young L2 learners. *Studies in Second Language Acquisition* 11(4): 415-439.
- Hasselgård, Hilde and Stig Johansson. 2011. 'Learner corpora and contrastive interlanguage analysis'. In Fanny Meunier, Sylvie De Cock, Gaëtanelle Gilquin and Magali Paquot (eds.), *A Taste of corpora. In honour of Sylviane Granger* (pp. 33-61). Amsterdam: John Benjamins.
- Hasselgren, Angela. 1994. Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics* 4(2): 237-260.
- _____. 2002. 'Learner corpora and language testing. Smallwords as markers of learner fluency'. In Sylviane Granger, Joseph Hung, and

-
- Stephanie Petch-Tyson (eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 143-173). Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Hatch, Evelyn Marcussen. 1983. *Psycholinguistics: A second language perspective*. Rowley, MA: Newbury House.
- Hatch, Evelyn Marcussen (ed.). 1978. *Second language acquisition: A book of readings*. Rowley, Mass.: Newbury House.
- Hemchua, Saengchan and Norbert Schmitt. 2006. An analysis of lexical errors in the English compositions of Thai learners. *Prospect* 21(3): 3-25.
- Hendrickson, James M. 1976. Error analysis and selective correction in the adult ESL classroom: An experiment. *ERIC Document Reproduction Service*, ED 135260: 1-21.
- Hickey, Tina. 1993. Identifying formulas in first language acquisition. *Journal of Child Language* 20(1): 27-41.
- Hoffman, Thomas and Graeme Trousdale (eds.) 2013. *The Oxford handbook of Construction Grammar*. New York: Oxford University Press.
- Hopper, Paul J. 1988. 'Emergent grammar and the a priori grammar postulate'. In Deborah Tannen (ed.), *Linguistics in context: Connecting observation and understanding* (pp. 117-134). Norwood, NJ: Ablex.
- Householder, Fred W. and Sol Saporta (eds.). 1962. *Problems in lexicography*. Bloomington: Indiana University Press.
- Howarth, Peter. 1996. *Phraseology in English academic writing. Some implications for language learning and dictionary making*. Tübingen: Niemeyer.
- _____. 1998a. Phraseology and second language proficiency. *Applied Linguistics* 19(1): 24-44.

-
- _____ 1998b. 'The phraseology of learners' academic writing'. In Anthony Paul Cowie (ed.), *Phraseology: Theory, analysis and applications* (pp. 161-186). Oxford: Oxford University Press.
- Huang, Joseph and Evelyn M. Hatch. 1978. 'A Chinese child's acquisition of English'. In Evelyn Marcussen Hatch (ed.), *Second language acquisition: A book of readings* (pp. 118-131). Rowley, MA: Newbury House.
- Huang, Zeping. 2014. The effects of paper-based DDL on the acquisition of lexico-grammatical patterns in L2 writing. *ReCALL* 26(2):163–183.
- Huddleston, Rodney and Geoffrey K. Pullum. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Hugon, Claire. 2008. 'High-frequency verbs: Starting block or stumbling block for advanced L2 communication? Insights from native and learner corpora'. In Gudrun Rawoens (ed.), *Taal aan den lijve. Het gebruik van corpora in taalkundig onderzoek en taalonderwijs* (Studia Germanica Gandensia (Libri) & Spieghelel Historiae) (pp. 69-98). Academia Press: Gent.
- Hulstijn, Jan H. and Elaine Marchena. 1989. Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition* 11(3): 241-255.
- Hunston, Susan. 2002. *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hunston, Susan and Gill Francis. 2000. *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam/Philadelphia: John Benjamins.

-
- Hyland, Ken and John Milton. 1997. Qualifications and certainty in L1 and L2 students' writing. *Journal of Second Language Writing* 6(2): 183-205.
- Hymes, Dell H. 1968. 'The ethnography of speaking'. In Joshua A. Fisherman (ed.), *Readings in the sociology of language* (pp. 99-138). The Hague/Paris: Mouton. [Repr. from T. Gladwin and W. C. Sturtevant (eds.), 1962. *Anthropology and human behavior* (pp. 13-53). Washington: The Anthropology Society of Washington.]
- Inal, Dilek and Esra Özdemir. 2015. 'Re/considering the English language teacher education programs in Turkey from an ELF standpoint: What do the academia, pre-service and in-service teachers think?'. In Yasemin Bayyurt and Sumru Akcan (eds.), *Current perspectives on pedagogy for English as a lingua franca* (pp. 135-152). De Gruyter Mouton: Berlin
- Işık, Ali. 2008. Yabancı dil eğitimimizdeki yanlışlar neden kaynaklanıyor? *Journal of Language and Linguistic Studies* 4(2): 15-26.
- Jackendoff, Ray. 1995. 'The boundaries of the lexicon'. In Martin Everaert, Erik-Jan van der Linden, André Schenk and Rob Schreuder (eds.), *Idioms: Structural and psychological perspectives* (pp. 133-165). Hillsdale, NJ: Lawrence Erlbaum.
- Jackson, Howard and Etienne Zé Amvela. 2007. *Words, meaning and vocabulary. An introduction to modern English lexicology* (2nd ed.). London: Continuum.
- Jarvis, Scott and Terence Odlin. 2000. Morphological type, spatial reference, and language transfer. *Studies in Second Language Acquisition* 22(4): 535-556.
- Janssen, Niels and Horacio A. Barber. 2012. Phrase frequency effects in language production. *Plos One* 7(3): e33202. doi:10.1371/journal.pone.0033202

-
- Jiang, Nan. 2002. Form-meaning mapping in vocabulary acquisition in a second language. *Studies in Second Language Acquisition* 24(4): 617-637.
- Jiang, Nan and Tatiana M. Nekrasova. 2007. The processing of formulaic sequences by second language speakers. *The Modern Language Journal* 91(3): 433-445.
- Johansson, Stig. 1978. *Studies in error gravity. Native reactions to errors produced by Swedish learners of English*. Göteborg, Sweden: Acta Universitatis Gothoburgensis.
- Johansson, Stig and Knut Hofland. 1989. *Frequency analysis of English vocabulary and grammar. Based on the LOB corpus. Volume 1: Tag frequencies and word frequencies*. Oxford: Clarendon.
- Johnson, Kyle. 1991. Object positions. *Natural Language & Linguistic Theory* 9(4): 577-636.
- Judd, Elliott L. 1978. Vocabulary teaching and TESOL: A need for reevaluation of existing assumptions. *TESOL Quarterly* 12(1): 71-76.
- Juknevičienė, Rita. 2008. Collocations with high frequency verbs in learner English: Lithuanian learners vs native speakers. *Kalbotyra* 59: 119-127. doi: 10.15388/Klbt.2008.7599.
- Jusczyk, Peter W. and Richard N. Aslin. 1995. Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology* 29(1): 1-23.
- Källkvist, Marie. 1999. *Form-class and task-type effects in learner English: A study of advanced Swedish learners*. Lund: Lund University Press.
- Kachru, Braj B. 1985. 'Standards, codification and sociolinguistic realism: The English language in the outer circle'. In Randolph Quirk and H.G. Widdowson (eds.), *English in the world: Teaching and*

-
- learning the language and literatures* (pp. 11-30). Cambridge: Cambridge University Press.
- _____. 1992. *The other tongue: English across cultures*. Urbana/Chicago: University of Illinois Press.
- Kaluza, Henryk. 1990. 'English verbs with prepositions and particles'. In Dietrich Nehls (ed.), *Grammatical studies in the English language* (pp. 19-23). Heidelberg: Groos.
- Kamarudin, Rafidah. 2013. A study on the use of phrasal verbs by Malaysian learners of English. Doctoral Dissertation. Birmingham: University of Birmingham.
- Kantor, Hadassa. 1978. An analysis of lexical errors in the interlanguage of Hebrew Learners. MA Thesis. Ramat Gan: Bar-Ilan University.
- Kao, Rong-Rong. 2001. Where have the prepositions gone? A study of English prepositional verbs and input enhancement in instructed SLA. *IRAL* 39(3): 195-215.
- Karahan, Pinar. 2015. The effect of conceptual metaphors on Turkish EFL learners' comprehension and production of phrasal verbs. *International Journal of Linguistics and Communication* 3(1): 76-88.
- Karakuş, Esra. 2017. The preferences of Turkish EFL learners in using phrasal verbs or synonymous one-word verbs. *International Journal of Language Academy* 5(5): 216-225.
- Karniol, Rachel. 1990. Second language acquisition via immersion in daycare. *Journal of Child Language* 17(1): 147-170.
- Kartal, Galip. 2018. Phrasal verbs in ELT coursebooks used in Turkey: A corpus-based analysis. *Cumhuriyet International Journal of Education* 7(4): 534-550.

-
- Kayael, Recep. 2007. Do Turkish teacher trainees avoid English phrasal verbs? A study with the students of ELT Department. Anadolu University. Unpublished MA Thesis. Anadolu University: Eskişehir.
- Kellerman, Eric. 1976. Elicitation, lateralization and error analysis. *Interlanguage Studies Bulletin* 1(1): 79-114.
- _____ 1977. Towards a characterization of the strategy of transfer in second language learning. *Interlanguage Studies Bulletin* 2: 58-145.
- _____ 1979. Transfer and non-transfer: where are we now? *Studies in Second Language Acquisition* 2(1): 37-57.
- _____ 1991. 'Compensatory strategies in second language research: A critic, a revision, and some (non-)implications for the classroom'. In Robert Phillipson, Eric Kellerman, Larry Selinker, Michael Sharwood Smith and Merrill Swain (eds.), *Foreign/second language pedagogy research. A commemorative volume for Claus Færch* (pp. 142-161). Clevedon, UK: Multilingual Matters.
- Kennedy, Arthur Garfield. 1920. *The Modern English verb-adverb combination*. Stanford: Stanford University Press. [Reprint 1967 New York: AMS Press.]
- Kennedy, Graeme. 1991. 'Between and through: The company they keep and the functions they serve'. In Karin Aijmer and Bengt Altenberg (eds.), *English corpus linguistics. Studies in honour of Jan Svartvik*. (pp. 95-110). New York: Longman.
- Khalil, Aziz. 1985. Communicative error evaluation: Native speakers' evaluation and interpretation of written errors of Arab EFL learners. *TESOL Quarterly* 19(2): 335-351.
- Kıray, Gülay and Seval Kömürçü. 2014. 'The use of verbs in written productions of Turkish learners of English'. *EJER Congress 2014*. Istanbul, Turkey 24-26 April (p. 704). Ankara: Anı Yayıncılık.

-
- Kilimci, Abdurahman and Cem Can. 2009. 'The status of English in Turkey'. In Sylviane Granger, Estelle Dagneaux, Fanny Meunier and Magali Paquot (eds.), *International Corpus of Learner English: Version 2* (pp. 205-210). Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.
- Kirkgöz, Yasemin. 2006. 'Teaching EFL at the primary level in Turkey'. In Mary Lou McCloskey, Janet Orr and Marlene Dolitsky (eds.), *Teaching English as a foreign language in primary school* (pp. 85-99). Alexandria, VA: TESOL.^[1]_[SEP]
- _____. 2007. English language teaching in Turkey: Policy changes and their implementations. *RELC Journal* 38(2): 216-228.^[1]_[SEP]
- _____. 2009. Globalization and English language policy in Turkey. *Educational Policy* 23(5): 663-684.
- _____. 2010. An analysis of written errors of Turkish adult learners of English. *Procedia Social and Behavioral Sciences*, 2: 4352-4358.
- Kirchner, Gustav. 1952. *Die zehn Hauptverben des Englischen im Britischen und Amerikanischen*. Halle: Max Niemeyer Verlag.
- Kirjavainen, Minna, Anna Theakston and Elena Lieven. 2009. Can input explain children's me-for-I errors? *Journal of Child Language* 36(5): 1091-1114.
- Kjellmer, Göran. 1991. 'A mint of phrases'. In Karin Aijmer and Bengt Altenberg (eds.), *English corpus linguistics. Studies in Honour of Jan Svartvik* (pp. 111-127). London: Longman.
- Kleinmann, Howard H. 1977. Avoidance in adult second language acquisition. *Language Learning* 27(1): 93-107.
- Kornfilt, Jaklin. 1997. *Turkish*. London: Routledge.
- Kövecses, Zoltán and Péter Szabó. 1996. Idioms: A view from cognitive semantics. *Applied Linguistics* 17(3): 326-355.

-
- Krashen, Stephen and Robin Scarcella. 1978. On routines and patterns in language acquisition and performance. *Language Learning* 28(2): 283-300.
- Kreidler, Charles W. 1966. English prepositions. *ELT Journal* 20(2): 119-122.
- Krishnamurthy, Ramesh. 2003. 'Language as chunks, not words'. In Malcolm Swanson and Kent Hill (eds.), *On JALT2002: Conference proceedings: waves of the future* (pp. 288-294). Tokyo: JALT.
<http://jalt-publications.org/archive/proceedings/2002/288.pdf>
- Kroll, Judith F. and Erika Stewart. 1994. Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language* 33: 149-174.
- Kuiper, Koenraad. 1992. The oral tradition in auction speech. *American Speech* 67(3): 279-289.
- _____ 1996. *Smooth talkers: The linguistic performance of auctioneers and sportscasters*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- _____ 2004. 'Formulaic performance in conventionalised varieties of speech'. In Norbert Schmitt (ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 37-54). Amsterdam: John Benjamins.
- Kuiper, Koenraad and Douglas Haggio. 1984. Livestock auctions, oral poetry, and ordinary language. *Language in Society* 13(2): 205-234.
- Kurtyka, Andrzej. 2001. 'Teaching English phrasal verbs: A cognitive approach'. In Martin Pütz, Susanne Niemeier and René Dirven (eds.), *Applied cognitive linguistics II: Language pedagogy* (pp. 29-54). Berlin/New York: Mouton de Gruyter.

-
- Lado, Robert. 1957. *Linguistics across cultures. Applied linguistics for language teachers*. Ann Arbor, MI: University of Michigan Press.
- _____. 1972. 'Patterns of difficulty in vocabulary'. In Harold B. Allen and Russell N. Campbell (eds.), *Teaching English as a second language. A book of readings* (pp. 275-288). New York: McGraw-Hill.
- Langacker, Ronald W. 1987. *Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites*. Stanford, CA: Stanford University Press.
- Laufer, Batia. 1990a. Why are some words more difficult than others? *International Review of Applied Linguistics in Language Teaching* 28(4): 293-307.
- _____. 1990b. 'Words you know: How they affect the words you learn'. In Jacek Fisiak (ed.), *Further insights into contrastive linguistics* (pp. 573-593). Amsterdam/Philadelphia: John Benjamins.
- _____. 1997. 'What is in a word that makes it hard or easy: Some intralexical factors that affect the learning of words'. In Norbert Schmitt and Michael McCarthy (eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 140-155). Cambridge: Cambridge University Press.
- _____. 2002. 'The lexical plight in second language reading'. In James Coady and Thomas Huckin (eds.), *Second language vocabulary acquisition* (pp. 20-34). Cambridge: Cambridge University Press.
- Laufer, Batia and Stig Eliasson. 1993. What causes avoidance in L2 learning: L1-L2 difference, L1-L2 similarity, or complexity? *Studies in Second Language Acquisition* 15(1): 35-48.
- Laufer, Batia and Nany Girsai. 2008. Form-focused instruction in second language vocabulary learning: A case for contrastive analysis and translation. *Applied Linguistics* 29(4): 694-716.

-
- Laufer, Batia and Tina Waldman. 2011. Verb-noun collocations in second language writing: A corpus analysis of learners' English. *Language Learning* 61(2): 647-672.
- Laufer-Dvorkin, Batia. 1991. *Similar lexical forms in interlanguage*. Tübingen. Gunter Narr Verlag.
- Leech, Geoffrey. 2001. 'The role of frequency in ELT: New corpus evidence brings a re-appraisal'. In W. Hu (ed.), *ELT in China 2001: Papers presented at the 3rd International Symposium on ELT in China* (pp. 1-23). Beijing: Foreign Language Teaching and Research Press.
- Leech, Geoffrey and Michael Short. 1981. *Style in fiction. A linguistic introduction to English fictional prose*. New York: Longman Group Limited.
- Leki, Ilona and Joan G. Carson. 1994. Students' perception of EAP writing instruction and writing needs across the discipline. *TESOL Quarterly* 28(1): 81-101.
- Lennon, Paul. 1991a. Error and the very advanced learner. *International Review of Applied Linguistics in Language Teaching* 29(1): 31-44.
- _____ 1991b. Error: Some problems of definition, identification, and distinction. *Applied Linguistics* 12(2): 180-196.
- _____ 1996. Getting 'easy' verbs wrong at the advanced level. *International Review of Applied Linguistics in Language Teaching* 34(1): 23-36.
- Levelt, Willem J. M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- _____ 1993. 'Language use in normal speakers and its disorders'. In Gerhard Blanken, Jürgen Dittmann, Hannelore Grimm, John C. Marshall and Claus-W. Wallesch (eds.), *Linguistic disorders and*

-
- pathologies. An international handbook* (pp. 1-15). Berlin: Walter de Gruyter.
- _____. 1995. The ability to speak: From intentions to spoken words. *European Review* 3(1): 13-23.
- Levenston, E. A. 1979. Second language acquisition: Issues and problems. *Interlanguage Studies Bulletin* 4(2): 147-160.
- Levenston, E. A. and S. Blum. 1977. 'Aspects of lexical simplification in the speech and writing of advanced adult learners'. In Stephen Pit Corder and E. Roulet (eds.), *The notions of simplification, interlanguages and pidgins and their relation to second language pedagogy* (pp. 51-71). Neuchâtel: Université de Neuchâtel, Faculté des Lettres.
- Lewis, Michael (ed.) 2000. *Teaching collocation: Further developments in the Lexical Approach*. Howe: LTP.
- Lewis, Michael. 2001. *Implementing the Lexical Approach: Putting theory into practice*. Hove, England: Language Teaching Publications.
- _____. 2002 [1993]. *The Lexical Approach. The state of ELT and a way forward*. Boston: Thomson.
- Li, Jie and Norbert Schmitt. 2009. The acquisition of lexical phrases in academic writing: A longitudinal case study. *Journal of Second Language Writing* 18(2): 85-102.
- Li, Wei, Xiuhong Zhang, Cheng Niu, Yuankai Jiang and Rohini Srihari. 2003. 'An expert lexicon approach to identifying English phrasal verbs'. In *Proceedings of the 41st Annual Meeting of the ACL* (pp. 513-520). Sapporo, Japan.
- Liao, Yan D. and Yoshinori J. Fukuya. 2004. Avoidance of phrasal verbs: The case of Chinese learners of English. *Second Language Studies* 20(2): 71-106.

-
- Lieven, Elena V. M., Julian M. Pine and Helen Dresner Barnes. 1992. Individual differences in early vocabulary development: Redefining the referential-expressive distinction. *Journal of Child Language* 19(2): 287-310.
- Lieven, Elena V. M., Julian M. Pine and Gillian Baldwin. 1997. Lexically-based learning and early grammatical development. *Journal of Child Language* 24(1): 187-220.
- Lieven, Elena V. M., Dorothé Salome and Michael Tomasello. 2009. Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics* 20(3): 481-507.
- Lightbown, Patsy M. 1983. 'Exploring relationships between developmental and instructional sequences in L2 acquisition'. In Herbert W. Seliger and Michael H. Long (eds.), *Classroom oriented research in second language acquisition* (pp. 217- 245). Rowley, MA: Newbury House.
- Lin, Phoebe Ming Sum. 2010. 'The phonology of formulaic sequences: A review'. In David Wood (ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 174-193). London: Continuum.
- Lindner, Christoph. 1994. Unnaturalness in advanced learners' English: A corpus-based feasibility study. Unpublished MA Thesis. Freiburg: University of Freiburg.
- Lindstromberg, Seth. 2010. *English prepositions explained*. Amsterdam: John Benjamins.
- Linnarud, Moira. 1986. *Lexis in composition: A performance analysis of Swedish learners' written English*. Malmö, Sweden: Liber Förlag Malmö.

-
- Lipka, Leonhard. 1972. *Semantic structure and word-formation. Verb-particle constructions in contemporary English*. München: Wilhelm Fink Verlag.
- Liu, Dilin. 2011. The most-frequently used English phrasal verbs in American and British English: A multi-corpus examination. *TESOL Quarterly* 45(4): 661-688.
- Live, Anna H. 1965. The discontinuous verb in English. *Word* 21(3): 428-451.
- _____. 1973. The take-have phrasal in English. *Linguistics* 95(11): 31-50.
- Lorenz, Gunter R. 1999. *Adjective intensification – Learners versus native speakers. A corpus study of argumentative writing*. Amsterdam/Atlanta: Rodopi.
- Lounsbury, Floyd G. 1963. 'Linguistics and psychology'. In Sigmund Koch (ed.), *Psychology: A study of a science. Study II. Empirical substructure and relations with other sciences. Vol. 6. Investigations of man as socius: Their place in psychology and the social sciences* (pp. 552–582). New York: McGraw-Hill.
- Lovelace, Eugene. 1987. Attributes that come to mind in the TOT state. *Bulletin of the Psychonomic Society* 25: 370–372.
- Lu, Zhi and Juan Sun. 2017. Presenting English polysemous phrasal verbs with two metaphor-based cognitive methods to Chinese EFL learners. *System* 69: 153-161.
- Mair, Christian. 1991. 'Quantitative or qualitative corpus analysis? Infinitival complement clauses in the Survey of English Usage corpus'. In Stig Johansson and Anna-Brita Stenström (eds.), *English computer corpora. Selected papers and research guide* (pp. 67-80). Berlin: Mouton de Gruyter.

-
- _____ 2013. The World System of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide* 34(3): 253-278.
- Manchón, Rosa M., Liz Murphy and Julia Roca. 2007. Lexical retrieval processes and strategies in second language writing: A synthesis of empirical research. *International Journal of English Studies* 7(2): 149–174.
- Martin, Marilyn. 1984. Advanced vocabulary teaching: The problem of synonyms. *The Modern Language Journal* 68(2): 130-137.
- Martinez, Ron. 2013. A framework for the inclusion of multi-word expressions in ELT. *ELT Journal* 67(2): 184-198.
- Marion, Waldemar. 1977. Foreign vocabulary learning as problem No. 1 of language teaching at the advanced level. *Interlanguage Studies Bulletin* 2(1): 33-57.
- Masangya, Elaine M. and Louella Lozada. 2009. An investigation on the relationship between the language exposures and errors in English essays of high school students. *Philippine ESL Journal* 2: 31-50.
- Mazaherylaghab, Hamzeh. 2013. *Iranian learner English: A corpus-based study of phrasal verb usage*. Published doctoral dissertation. Freiburg: University of Freiburg.
- McArthur, Tom. 1989. The long-neglected phrasal verb. *English Today* 5(2): 38-44.
- McArthur, Tom and Beryl Atkins. 1974. *Dictionary of English phrasal verbs and their idioms*. Stuttgart: Klett.
- McCarthy, Michael. 2004. *From corpus to course book*. New York: Cambridge University Press.
- _____ 2010. Spoken fluency revisited. *English Profile Journal* 1(1): 1-15.

-
- McCarthy, Michael and Felicity O'Dell. 2005. *English collocations in use*. Cambridge: Cambridge University Press.
- McClelland, James L. 2010. Emergence in cognitive science. *Topics in Cognitive Science* 2(4): 751-770.
- McClelland, James L., Matthew M. Botvinick, David C. Noelle, David C. Plaut, Timothy T. Rogers, Mark S. Seidenberg and Linda B. Smith. 2010. Letting structure emerge: Connectionist and dynamical systems approaches to understanding cognition. *Trends in Cognitive Sciences* 14(8): 348-356.
- McCrostie, James. 2008. Writer visibility in EFL learner academic writing: A corpus-based study. *ICAME Journal* 32: 97-114.
- Meara, Paul. 1980. Vocabulary acquisition: A neglected aspect of language learning. *Language Teaching* 13(3-4): 221-246.
- _____. 1984. 'The study of lexis in interlanguage'. In Alan Davies, Clive Criper and Anthony P. R. Howatt (eds.), *Interlanguage* (pp. 225-235). Edinburgh: Edinburgh University Press.
- _____. 1996. 'The dimensions of lexical competence'. In Gillian Brown, Kirsten Malmkjær and John Williams (eds.), *Performance and competence in second language acquisition* (pp. 35-53). Cambridge: Cambridge University Press.
- Mel'čuk, Igor. 1998. 'Collocations and lexical functions.' In Anthony P. Cowie (ed.), *Phraseology. Theory, analysis, and applications* (pp. 23-53). Oxford: Clarendon.
- Melka-Teichroew, Francine J. 1982. Receptive versus productive vocabulary: A survey. *Interlanguage Studies Bulletin* 6(2): 5-33.
- Medgyes, Péter. 1992. Native or non-native: Who's worth more? *ELT Journal* 46(4): 340-349.

-
- Men, Haiyan. 2015. *Vocabulary increase and collocation learning: A corpus-based cross-sectional study of Chinese EFL learners*. Published doctoral dissertation. Birmingham: Birmingham City University.
- Meunier, Fanny. 2002. 'The pedagogical value of native and learner corpora in EFL grammar teaching'. In Sylviane Granger, Joseph Hung and Stephanie Petch-Tyson (eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 120-141). Amsterdam/Philadelphia: John Benjamins.
- _____. 2012. Formulaic language and language teaching. *Annual Review of Applied Linguistics* 32: 111-129.
- Meunier, Fanny and Sylviane Granger (eds.). 2008. *Phraseology in foreign language learning and teaching*. Amsterdam/Philadelphia: John Benjamins.
- Meyer, George A. 1975. *The two-word verb. A dictionary of the verb-preposition phrases in American English*. The Hague/Paris: Mouton.
- Millar, Neil. 2011. The processing of malformed formulaic language. *Applied Linguistics* 32(2): 129-148.
- Miller, George A. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63: 81-97.
- Mitchell, T. F. 1958. 'Syntagmatic relations in linguistic analysis'. In *Transactions of the Philological Society* (pp. 101-118). Oxford: Basil Blackwell.
- Mondor, Monika. 2008. *Figuring it out*. A corpus-based comparison of the verb-particle construction in argumentative writing by Swedish advanced learners and native speakers of English. Doctoral dissertation. Gothenburg: University of Gothenburg.

-
- Moon, Rosamund. 1998. *Fixed expressions and idioms in English*. Oxford: Clarendon Press.
- Möller, Verena. 2017. 'A statistical analysis of learner corpus data, experimental data and individual differences: Monofactorial vs. multifactorial approaches'. In Peter de Haan, Rina de Vries and Sanne van Vuuren (eds.), *Language, learners and levels: Progression and variation. Corpora and language in use – Proceedings 3* (pp. 409-439). Louvain-la-Neuve: Presses universitaires de Louvain.
- Mukherjee, Joybrato. 2009. 'The lexicogrammar of present-day Indian English: Corpus-based perspectives on structural nativisation'. In Ute Römer and Rainer Schulze (eds.), *Exploring the lexis-grammar interface* (pp. 117-135). Amsterdam/ Philadelphia: John Benjamins.
- _____. 2010. 'Corpus-based insights into verb-complementational innovations in Indian English: Cases of nativised semantico-structural analogy'. In Alexandra N. Lenz and Albrecht Plewnia (eds.), *Grammar between norm and variation* (pp. 219-241). Frankfurt am Main: Peter Lang.
- Munby, John. 1978. *Communicative syllabus design. A sociolinguistic model for defining the content of purpose-specific language programmes*. Cambridge: Cambridge University Press.
- Myles, Florence. 2004. From data to theory: The over-representation of linguistic knowledge in SLA. *Transactions of the Philological Society* 102(2): 139–168.
- Myles, Florence, Janet Hooper and Rosamond Mitchell. 1998. Rote or rule? Exploring the role of formulaic language in classroom foreign language learning. *Language Learning* 48(3): 323–363.
- Myles, Florence, Rosamond Mitchell and Janet Hooper. 1999. Interrogative chunks in French L2. *Studies in Second Language Acquisition* 21(1): 49–80.

-
- Mwangi, Serah. 2004. Prepositions vanishing in Kenya. A case of syntactic simplification in a variety of East African English. *English Today* 20(1): 27-32.
- Namwar, Firooz. 2012. The relationship between language proficiency and use of collocation by Iranian EFL students. *3L: The Southeast Asian Journal of English Language Studies* 18(1): 41-52.
- Nation, Paul. 2006. 'Vocabulary: Second language'. In Keith Brown (ed.), *Encyclopaedia of language and linguistics* (2nd ed.). (pp. 448-454). Oxford: Elsevier.
- Nattinger, James R. 1988. 'Some current trends in vocabulary teaching'. In Ronald Carter and Michael McCarthy (eds.), *Vocabulary and language teaching* (pp. 62-82). London/New York: Longman.
- Nattinger, James R. and Jeanette S. DeCarrico. 1992. *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Nelson, Katherine. 1973. Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development* 38(1/2): 1-135.
- Nesselhauf, Nadja. 2003. The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics* 24(2): 223-242.
- _____. 2005. *Collocations in a learner corpus*. Amsterdam: John Benjamins.
- _____. 2009. Co-selection phenomena across New Englishes: Parallels (and differences) to foreign learner varieties. *English World-Wide* 30(1): 1-25.
- Newell, Allen. 1990. *Unified theories of cognition*. Cambridge, Mass.: Harvard University Press.

-
- Nickel, Gerhard 1978. 'Complex verbal structures in English'. In Dietrich Nehls (ed.), *Studies in descriptive English grammar* (pp. 63-83). Heidelberg: Julius Groos.
- Nickerson, Raymond S. and Marilyn J. Adams. 1979. Long-term memory for a common object. *Cognitive Psychology* 11(3): 287–307.
- Oakes, Michael P. 1998. *Statistics for corpus linguistics*. Edinburgh: Edinburgh University Press.
- Obukadeta, Peter. 2019. Collocations in a learner English corpus: Analysis of Yoruba-speaking Nigerian English learners' use of collocations. Doctoral dissertation. School of Arts, Culture and Communication, Kingston University London.
- O'Dell, Felicity and Michael McCarthy. 2008. *English collocations in use. Advanced*. Cambridge: Cambridge University Press.
- O'Dowd, Elizabeth. 1998. *Prepositions and particles in English*. New York/Oxford: Oxford University Press.
- O'Keeffe, Anne, Michael McCarthy and Ronald Carter. 2007. *From corpus to classroom. Language use and language teaching*. Cambridge: Cambridge University Press.
- Olsson, Margareta. 1972. *Intelligibility. A study of errors and their importance*. Department of Educational Research, Gothenburg School of Education. Gothenburg, Sweden: Gothenburg University. (ERIC Document Reproduction Service ED 072681).
- _____. 1973. 'The effects of different types of errors in the communication situation'. In Jan Startvik (ed.), *Errata. Papers in error analysis*. (pp. 153-160). Lund: CWK Gleerup.

-
- Olson, Gary M. 1973. 'Developmental changes in memory and the acquisition of language'. In Timothy E. Moore (ed.), *Cognitive development and the acquisition of language* (pp.145-157). New York: Academic Press.
- Oros, Martin. 2006. A comparison of phrasal verbs used in British and American English. BA Thesis. Masaryk University. Department of English and American Studies.
- Özışık, Cem. 2014. Identifying preposition errors of Turkish EFL students. *ELT Research Journal* 3(2): 59-69.
- Özsoy, A. Sumru (ed.). 2019. *Word order in Turkish*. Cham: Springer International Publishing.
- Paikeday, Thomas M. 1985. *The native speaker is dead!* Toronto/New York: Paikeday Publishing Inc.
- Page, Brian. 1990. 'Why do we have to get it right anyway?'. In Brian Page (ed.), *What do you mean... it's wrong?* (pp. 102-106). London: Centre for Information on Language Teaching and Research.
- Palmer, Harold E. ¹⁴1961[1938]. *A grammar of English words*. London: Longmans.
- Palmer, Frank R. 1974. *The English verb*. London: Longman Group Limited.
- _____ 1981. *Semantics*. (2nd ed.) Cambridge: Cambridge University Press.
- Paquot, Magali and Sylviane Granger. 2012. Formulaic language in learner corpora. *Annual Review of Applied Linguistics* 32: 130-149.
- Pawley Andrew and Frances Hodgetts Syder. 1983. 'Two puzzles for linguistic theory: Nativelike selection and nativelike fluency.' In Jack C. Richards and Richard W. Schmidt (eds.), *Language and communication* (pp. 191-226). New York: Longman.

-
- Pietilä, Päivi, Katalin Doró and Renata Pípalová (eds.). 2015. *Lexical issues in L2 writing*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Percillier, Michael. 2016. *World Englishes and second language acquisition. Insights from Southeast Asian Englishes*. Amsterdam/Philadelphia: John Benjamins.
- Perera, Natsuko Shibata. 2001. The role of prefabricated language in young children's second language acquisition. *Bilingual Research Journal* 25(3): 327–356.
- Peters, Ann M. 1977. Language learning strategies: Does the whole equal the sum of the parts? *Language* 53(3): 560-573.
- _____. 1983. *The units of language acquisition*. Cambridge: Cambridge University Press.
- Pierrehumbert, Janet B. 2001. 'Exemplar dynamics: Word frequency, lenition, and contrast'. In Joan Bybee and Paul Hopper (eds.), *Frequency effects and the emergence of lexical structure* (pp. 137–157). Amsterdam: John Benjamins.
- Pine, Julian M. 1990. *Individual differences in early language development and their relationship to maternal style*. Unpublished doctoral dissertation. University of Manchester.
- Pine, Julian M. and Elena V. M. Lieven. 1993. Reanalysing rote-learned phrases: Individual differences in the transition to multi-word speech. *Journal of Child Language* 20(3): 551-571.
- Plunkett, Kim. 1993. Lexical segmentation and vocabulary growth in early language acquisition. *Journal of Child Language* 20(1): 43-60.
- Politzer, Robert L. 1978. Errors of English speakers of German as perceived and evaluated by German natives. *Modern Language Journal* 62(5/6): 253:261.

-
- Porte, Graeme. 1997. The etiology of poor second language writing: The influence of perceived teacher preferences on second language revision strategies. *Journal of Second Language Writing* 6(1): 61-78.
- Poulisse, Nanda. 1993. 'A theoretical account of lexical communication strategies'. In Robert Schreuder and Bert Weltens (eds.), *The bilingual lexicon* (pp. 157-189). Amsterdam: Benjamins.
- Poulisse, Nanda and Erik Schils. 1989. The influence of task- and proficiency-related factors on the use of compensatory strategies: A quantitative analysis. *Language Learning* 39(1): 15-48.
- Poulisse, Nanda and Theo Bongaerts. 1994. First language use in second language production. *Applied Linguistics* 15(1): 36-57.
- Prodromou, Luke. 2005. 'You see, it's sort of tricky for the L2 user': The puzzle of idiomaticity in English as a Lingua Franca. Doctoral dissertation. University of Nottingham.
- Quirk, Randolp, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A comprehensive grammar of the English language*. Harlow: Longman.
- Rahman, Tariq. 1990. *Pakistani English*. Islamabad: National Institute of Pakistan Studies Quaid-i-Azam University.
- Rajagopalan, Kanavillil. 1997. Linguistics and the myth of nativity: Comments on the controversy over 'new/non-native Englishes'. *Journal of Pragmatics* 27(2): 225-231.
- Rakıcıoğlu Söylemez, Anıl. 2016. 'Türkiye'de yabancı dil eğitiminde yaşanan sorunlar'. In Erdiñ Yücel, Hasan Yılmaz and M. Serkan Öztürk (eds.), *Yabancı dil öğretimine genel bir bakış* (pp. 119-140). Konya: Çizgi Kitabevi.
- Rampton, M.B.H. 1990. Displacing the 'native speaker': Expertise, affiliation, and inheritance. *ELT Journal* 44(2): 97-101.

-
- Ramsey, Nick F., Johan Martijn Jansma, Gerry Jager, Tamar R. van Raalten and R. S. Kahn. 2004. Neurophysiological factors in human information processing capacity. *Brain* 127(3): 517-525.
- Rastall, Paul. 1994. The prepositional flux. *International Review of Applied Linguistics in Language Teaching* 32(3): 229-231.
- Read, John and Paul Nation. 2004. 'Measurement of formulaic sequences'. In Norbert Schmitt (ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 23-55). Amsterdam: John Benjamins.
- Reisberg, Daniel. 2013. *Cognition. Exploring the science of the mind*. (5th ed.). New York: W. W. Norton Company.^[1]_{SEP}
- Renouf, Antoinette and John Sinclair. 1991. 'Collocational frameworks in English'. In Karin Aijmer and Bengt Altenberg (eds.), *English corpus linguistics. Studies in honour of Jan Svartvik* (pp. 128-143). London & New York: Longman.
- Richards, Jack C. and Theodore S. Rodgers. 2014. *Approaches and methods in language teaching* (3rd ed.). Cambridge: Cambridge University Press.
- Rifkin, Benjamin and Felicia D. Roberts. 1995. Error gravity: A critical review of research design. *Language Learning* 45(3): 511-537.
- Riguel, Emilie. 2014. 'Phrasal verbs, "the scourge of the learner"'. In Carolina Pérez-Arredondo, Margarita Calderón-López, Hilda Hidalgo-Avilés and David Pask-Hughes (eds.), *Papers from the 9th Lancaster University Postgraduate Conference in Linguistics & Language Teaching* (pp. 1- 19). Lancaster University, Department of Linguistics and English Language.
- Ringbom, Håkan. 1987. *The role of the first language in foreign language learning*. Clevedon: Multilingual Matters.

-
- _____ 1998a. 'Vocabulary frequencies in advanced learner English: a cross-linguistic approach'. In Sylviane Granger (ed.), *Learner English on computer* (pp. 41-52). London: Longman.
- _____ 1998b. 'High-frequency verbs in the ICLE Corpus'. In Antoinette Renouf (ed.), *Explorations in corpus linguistics* (pp. 191-200). Amsterdam & Atlanta, GA: Rodopi.
- Roberts, Murat H. 1936. The antiquity of the Germanic verb-adverb locution. *The Journal of English and Germanic Philology* 35(4): 466-481.
- Römer, Ute. 2009. The inseparability of lexis and grammar. Corpus linguistic perspectives. *Annual Review of Cognitive Linguistics* 7: 141-163.
- Römer, Ute and Rainer Schulze (eds.). 2009. *Exploring the lexis-grammar interface*. Amsterdam: John Benjamins.
- Römer, Ute and Rainer Schulze (eds.). 2010. *Patterns, meaningful units and specialized discourses*. Amsterdam: John Benjamins.
- Rowland, Caroline F. 2007. Explaining errors in children's questions. *Cognition* 104(1): 106-134.
- Rowland, Caroline and Julian Pine. 2000. Subject-auxiliary inversion errors and wh-question acquisition: 'What children do know?' *Journal of Child Language* 27(1): 157-181.
- Rudzka-Ostyn, Brygida. 2003. *Word power: Phrasal verbs and compounds. A cognitive approach*. Berlin/New York: Mouton de Gruyter.
- Rumelhart, David E. and James L. McClelland. 1986. 'On learning the past tenses of English verbs'. In David E. Rumelhart and James L. McClelland (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 216-271). Cambridge: MIT Press.

-
- Saffran, Jenny R., Richard N. Aslin and Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274(5294): 1926-1928.
- Santos, Terry. 1988. Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly* 22(1): 69-90.
- Sarıçoban, Gülay and Arif Sarıçoban. 2012. Atatürk and the history of foreign language education in Turkey. *The Journal of Language and Linguistic Studies* 8(1): 24-49.
- Saussure, Ferdinand De. 1966[1916]. *Course in general linguistics*. New York: McGraw-Hill.
- Savignon, Sandra J. 1972. *Communicative competence: An experiment in foreign language teaching*. Philadelphia: Center for Curriculum Development.
- _____. 1983. *Communicative competence: Theory and classroom practice*. Reading, Massachusetts: Addison-Wesley.
- Saville-Troike, Muriel. 1984. What really matters in second language learning for academic achievement? *TESOL Quarterly* 18(2): 199-219.
- Savran, Hülya. 2008. About teaching compound verbs in Turkish grammar education. *Sosyal Bilimler Dergisi* 107-113.
<http://sbe.balikesir.edu.tr/dergi/edergi/c5s7/makale/c5s7m7.pdf>
- Saxton, Matthew. 2010. *Child language. Acquisition and development*. London: Sage.
- Schachter, Jacquelyn. 1974. An error in error analysis. *Language Learning* 24(2): 205-214.
- Schachter, Jacquelyn and Marianne Celce-Murcia. 1977. Some reservations concerning error analysis. *TESOL Quarterly* 11(4): 441-451.

-
- Schmidt, Richard W. 1990. The role of consciousness in second language learning. *Applied Linguistics* 11(2): 129-150.
- Schmitt, Norbert (ed.). 2004. *Formulaic sequences. Acquisition, processing and use*. Amsterdam & Philadelphia: John Benjamins.
- Schmitt, Norbert and Ronald Carter. 2004. 'Formulaic sequences in action. An introduction'. In Norbert Schmitt (ed.), *Formulaic sequences. Acquisition, processing and use* (pp. 1-22). Amsterdam & Philadelphia: John Benjamins.
- Schmitt, Norbert and Stephen Redwood. 2011. 'Learner knowledge of phrasal verbs: A corpus-informed study'. In Fanny Meunier, Sylvie De Cock, Gaëtanelle Gilquin and Magali Paquot (eds.), *A Taste for corpora. In honour of Sylviane Granger* (pp. 173-207). Amsterdam/Philadelphia: John Benjamins.
- Schmidt, Richard. 1983. 'Interaction, acculturation, and the acquisition of communicative competence: A case study of an adult'. In Nessa Wolfson and Elliot Judd (eds.), *Sociolinguistics and language acquisition* (pp. 137-174). Rowley, MA: Newbury House.
- _____ 1992. Psychological mechanisms underlying second language fluency. *Studies in Second Language Acquisition* 14(4): 357-385.
- Schneider, Gerold and Gaëtanelle Gilquin. 2018. 'Detecting innovations in a parsed corpus of learner English'. In Sandra C. Deshors, Sandra Götz and Samanatha Laporte (eds.), *Rethinking linguistic creativity in non-native Englishes* (pp. 47-74). Amsterdam: John Benjamins Publishing.
- Schneider, Gerold and Lena Zipp. 2013. Discovering new verb-preposition combinations in New Englishes. *Studies in Variation, Contacts and Change in English*, 13:online. <http://doi.org/10.5167/uzh-82328>

-
- Schumann, John H. 1978. *The pidginization process: A model for second language acquisition*. Rowley, MA: Newbury House.
- Seidl, Jennifer. 1990. *English idioms: Exercises on phrasal verbs*. Oxford: Oxford University Press.
- Seidenberg, Mark. S. 1994. Language and connectionism: The developing interface. *Cognition* 50(1-3): 385–401.
- Selinker, Larry. 1972. Interlanguage. *IRAL* 10: 209-232.
- Shapira, R. 1978. 'The non-learning of English: A case study of an adult'. In Evelyn Marcussen Hatch (ed.), *Second language acquisition: A book of readings* (pp. 246-255). Rowley, MA: Newbury House.
- Side, Richard. 1990. Phrasal verbs: Sorting them out. *English Language Teaching Journal* 44(2): 144-152.
- Silva, Tony. 1993. Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly* 27(4): 657–677.
- Sinclair, John McHardy. 1966. 'Beginning the study of lexis'. In Charles E. Bazell, John C. Catford, Michael A.K. Halliday and Robert H. Robins (eds.), *In memory of J.R. Firth* (pp. 410-430). London: Longman.
- _____ 1987. 'Collocation: A progress report'. In Ross Steele and Terry Threadgold (eds.), *Language topics. Essays in honour of Michael Halliday* (pp. 319-331). Amsterdam/Philadelphia: John Benjamins.
- _____ 1990 (ed. in chief). *Collins COBUILD English grammar*. London: Collins.
- _____ 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

-
- _____ 1996. The search for units of meaning. *TEXTUS* 9(1): 75-106.
- _____ 1998. 'The lexical item.' In Edda Weigand (ed.), *Contrastive lexical semantics* (pp. 1-24). Amsterdam: John Benjamins.
- _____ 2004. *Trust the text. Language, corpus and discourse*. London: Routledge.
- Sinclair, John McHardy and Antoinette Renouf. 1988. 'A lexical syllabus for language learning'. In Ronald Carter and Michael McCarthy (eds.), *Vocabulary and language teaching* (pp. 140-160). New York: Longman.
- Siyanova, Anna and Norbert Schmitt. 2007. Native and nonnative use of multi-word vs. one-word verbs. *IRAL* 45(2): 119-139.
- Siyanova, Anna and Norbert Schmitt. 2008. L2 learner production and processing of collocation: A multi-study perspective. *The Canadian Modern Language Review* 64(3): 429-458.
- Siyanova-Chanturia, Anna and Ron Martinez. 2015. The idiom principle revisited. *Applied Linguistics* 36(5): 549-569.
- Siyanova-Chanturia, Anna and Ana Pellicer-Sánchez. 2019. *Understanding formulaic language. A second language acquisition perspective*. New York: Routledge.
- Sjöholm, Kaj. 1995. *The influence of crosslinguistic, semantic, and input factors on the acquisition of English phrasal verbs: A comparison between Finnish and Swedish learners at an intermediate and advanced level*. Published doctoral dissertation. Abo Akademi University Press.
- Skehan, Peter. 1998. *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Smith, Logan Pearsall. 1925. *Words and idioms: Studies in the English language* (2nd ed.). London: Constable & Company Ltd.

-
- Snider, Neal and Inbal Arnon. 2012. 'A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon'. In Dagmar Divjak and Stefan Th. Gries (eds.), *Frequency effects in language representation* (pp. 127-164). Berlin: Mouton de Gruyter.
- Snow, Catherine E. 1986. 'Conversations with children'. In Paul Fletcher and Michael Garman (eds.), *Language acquisition* (2nd ed.) (pp. 69-89). Cambridge: Cambridge University Press.
- Sonaiya, Remi. 1991. Vocabulary acquisition as a process of continuous lexical discrimination. *IRAL* 29 (4): 273-284.
- Spring, Ryan. 2018. Teaching phrasal verbs more efficiently: Using corpus studies and cognitive linguistics to create a particle list. *Advances in Language and Literary Studies* 9(5): 121-135.
- Sroka, Kazimierz A. 1972. *The syntax of English phrasal verbs*. The Hague: Mouton.
- Stein, Gabriela. 1991. The phrasal verb type 'to have a look' in Modern English. *IRAL* 29(1): 1-29.
- Stella, Massimo, Nicole M. Beckage, Marcus Brede and Manlio De Domenico. 2018. Multiplex model of mental lexicon reveals explosive learning in humans. *Scientific Reports* 8: 2259. DOI:10.1038/s41598-018-20730-5.
- Stevick, Earl W. 1950. The 'deferred preposition'. *American Speech* 25(3): 211-214.
- Stubbs, Michael. 1997. 'Eine Sprache idiomatisch sprechen: Computer, Korpora, kommunikative Kompetenz und Kultur'. In Klaus J. Mattheier (ed.), *Norm und Variation*. (pp. 151-167). Frankfurt am Main: Peter Lang.
- Stubbs, Michael. 2001. *Words and phrases. Corpus studies of lexical semantics*. Oxford: Blackwell.

-
- Sung, Min-Chang. 2017. Underuse of English verb-particle constructions in an L2 learner corpus: Focus on structural patterns and one-word preference. *Corpus Linguistics and Linguistic Theory* 16(1): 1-26.
- Swan, Michael. 2009. 'German speakers'. In Michael Swan and Bernard Smith (eds.), *Learner English. A teacher's guide to interference and other problems* (pp. 37-51). Cambridge: Cambridge University Press.
- Taha, Abdul Karim. 1960. The structure of two-word verbs in English. *Language Learning* 10(3-4): 115-122.
- Tan, Melinda. 2005. Authentic language or language errors? Lessons from a learner corpus. *ELT Journal* 59(2): 126-134.
- Tannen, Deborah. 2007. *Talking voices. Repetition, dialogue, and imagery in conversational discourse* (2nd ed.). New York: Cambridge University Press.
- Taylor, Gordon. 1986. Errors and explanations. *Applied Linguistics* 7(2): 144-166.
- Terkourafi, Marina. 2002. Politeness and formulaicity: Evidence from Cypriot Greek. *Journal of Greek Linguistics* 3: 179–201.
- Thewissen, Jennifer. 2013. Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus. *Modern Language Journal* 97 (Supplement): 77–101 [L_{SEP}].
- Thiessen, Erik D. and Jenny R. Saffran. 2003. When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology* 39(4): 706-716.
- Thim, Stefan. 2012. *Phrasal verbs. The English verb-particle construction and its history*. Berlin: De Gruyter Mouton.
- Thomas, Margaret. 1994. Assessment of L2 proficiency in second language acquisition research. *Language Learning* 44(2): 307-336.

-
- Thompson, Ian. 2009. 'Turkish speakers'. In Michael Swan and Bernard Smith (eds.), *Learner English. A teacher's guide to interference and other problems* (pp. 214-226). Cambridge: Cambridge University Press.
- Tomasello, Michael. 1992. *First verbs: A case study in early grammatical development*. Cambridge: Cambridge University Press.
- _____. 2000. Do young children have adult syntactic competence? *Cognition* 74(3): 209-253.
- _____. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge: Harvard University Press.
- _____. 2009. 'The usage-based theory of language acquisition'. In Edith L. Bavin (ed.), *The Cambridge handbook of child language* (pp. 69-87). Cambridge: Cambridge University Press.
- Tomasello, Michael and Patricia J. Brooks. 1999. 'Early syntactic development: A construction grammar approach'. In Martyn Barrett (ed.), *The development of language* (pp. 161-190). Hove: Psychology Press.
- Tono, Yukio. 2003. 'Learner corpora: design, development and applications'. In Paul Rayson, Andrew Wilson, Tony McEnery, Andrew Hardie and Shereen Khoja (eds.), *Proceedings of the Corpus Linguistics 2001 Conference* (pp. 801-809). University Centre for Computer Corpus Research on Language Technical Papers, Number 13. Lancaster University.
- Torres-Martínez, Sergio. 2015. A constructionist approach to the teaching of phrasal verbs. *English Today* 123, 31(3): 46-58.
- Towell, Richard, Roger D. Hawkins and N. Bazergui. 1996. The development of fluency in advanced learners of French. *Applied Linguistics* 17(1): 84-119.

-
- Tremblay, Antoine, Bruce Derwing, Gary Libben and Chris Westbury. 2011. Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning* 61(2): 569-613.
- Trenkic, Danijela. 2009. 'Accounting for patterns of article omissions and substitutions in second language production'. In María del Pilar García Mayo and Roger Hawkins (eds.), *Second language acquisition of articles. Empirical findings and theoretical implications* (pp. 115-143). Amsterdam & Philadelphia: John Benjamins.
- Turton, Nigel D. and John B. Heaton. 1996. *Longman dictionary of common errors*. (2nd ed.). Essex: Pearson.
- Twaddell, Freeman. 1973. Vocabulary expansion in the TESOL classroom. *TESOL Quarterly* 7(1): 61-78.
- Uchida, Tomio. 2012. Use of multiword verbs by non-advanced EFL learners: Focusing on common verb + particle combinations. Retrieved August 18, 2017 from http://cblle.tufs.ac.jp/assets/files/publications/working_papers_08/section/303-323.pdf
- Underwood, Geoffrey, Norbert Schmitt and Adam Galpin. 2004. 'The eyes have it: An eye-movement study into the processing of formulaic sequences'. In Norbert Schmitt (ed.), *Formulaic sequences* (pp. 153-172). Amsterdam: John Benjamins.
- Van Lancker-Sidtis, Diana. 2001. Is your syntactic component really necessary? *Aphasiology* 15(4): 343-406.
- Van Lancker-Sidtis, Diana and Gail Rallon. 2004. Tracking the incidence of formulaic expressions in everyday speech: Methods for classification and verification. *Language & Communication* 24(3): 207-240.

-
- VanPatten, Bill. 1990. Attending to form and content in the input: An experiment in consciousness. *Studies in Second Language Acquisition* 12(3): 287-301.
- Vestergaard, Torben. 1977. *Prepositional phrases and prepositional verbs. A study in grammatical function*. The Hague & Paris: Mouton.
- Vihman, Marilyn May. 1982. 'Formulas in first and second language acquisition.' In Loraine K. Obler and Lise Menn (eds.), *Exceptional language and linguistics* (pp. 261-284). London: Academic Press.
- Vyatkina, Nina. 2016. Data-driven learning of collocations: Learner performance, proficiency, and perceptions. *Language Learning and Technology* 20(3): 159-179.
- Waibel, Birgit. 2005. Phrasal verbs and the foreign language learner: Results from a pilot study based on the *International Corpus of Learner English*. *Ranam* 38: 65- 74.
- _____. 2007. *Phrasal verbs in learner English: A corpus-based study of German and Italian students*. Published doctoral dissertation. Freiburg: Albert-Ludwigs-Universität Freiburg.
- Wang, Ying. 2016. *The idiom principle and L1 influence: A contrastive learner-corpus study of delexical verb-noun collocations*. Amsterdam & Philadelphia: John Benjamins.
- Wang, Ying and Philip Shaw. 2008. Transfer and universality: Collocation use in advanced Chinese and Swedish learner English. *ICAME Journal* 32: 201-232.
- Warren, Beatrice. 2001. 'Accounting for compositionality'. In Karin Aijmer (ed.), *A wealth of English - Studies in honour of Göran Kjellmer* (pp. 103-114). Göteborg: Acta Universitatis Gothoburgensis.

-
- Webb, Stuart and Eve Kagimoto. 2011. Learning collocations: Do the number of collocates, position of the node word, and synonymy affect learning? *Applied Linguistics* 32(2): 259-276.
- Webb, Stuart, Jonathan Newton and Anna Chang. 2013. Incidental learning of collocation. *Language Learning* 63(1): 91-120.
- Weinert, Regina. 1995. The role of formulaic language in second language acquisition: A review. *Applied Linguistics* 16(2): 180-205.
- White, Benjamin J. 2012. A conceptual approach to the instruction of phrasal verbs. *The Modern Language Journal* 96(3): 419-438.
- Widdowson, H. G. 1978. *Teaching language as communication*. Oxford: Oxford University Press.
- Wierszycka, Joanna. 2013. Phrasal verbs in learner English: A semantic approach. A study based on a POS tagged spoken corpus of learner English. *Research in Corpus Linguistics* 1: 81-93.
- Wiktorsson, Maria. 2003. *Learning idiomaticity. A corpus-based study of idiomatic expressions in learners' written production*. Doctoral dissertation. Lund: Lund University.
- Willis, Dave. 1990. *The lexical syllabus: A new approach to language teaching*. London: Collins COBUILD.
- _____ 2003. *Rules, patterns and words: Grammar and lexis in English language teaching*. Cambridge: Cambridge University Press.
- Wilkins, D. A. 1972. *Linguistics in language teaching*. Cambridge: The MIT Press.
- Wolter, Brent. 2001. Comparing the L1 and L2 mental lexicon. A depth of individual word knowledge model. *Studies in Second Language Acquisition* 23(1): 41-69.

-
- _____ 2006. Lexical network structures and L2 vocabulary acquisition: The role of L1 lexical/conceptual knowledge. *Applied Linguistics* 27(4): 741-747.
- Wolter, Brent and Henrik Gyllstad. 2011. Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics* 32(4): 430-449.
- Wong Fillmore, Lily. 1976. *The second time around: Cognitive and social strategies in language acquisition*. Unpublished doctoral dissertation. Palo Alto, CA: Stanford University.
- Wood, David. 2006. Uses and functions of formulaic sequences in second language speech: An exploration of the foundations of fluency. *Canadian Modern Language Review* 63(1): 13–33.
- Wray, Alison. 1999. Formulaic language in learners and native speakers. *Language Teaching* 32(4): 213-231.
- _____ 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- _____ 2004. ‘Here’s one I prepared earlier’: Formulaic language learning on television. In Norbert Schmitt (ed.), *Formulaic sequences: Acquisition, processing and use* (pp. 249–268). Amsterdam: John Benjamins.
- _____ 2008. *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.
- _____ 2012. What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics* 32: 231-254.
- Wray, Alison and Michael P. Perkins. 2000. The functions of formulaic language: An integrated model. *Language & Communication* 20(1): 1-28.

-
- Xiao, Richard and Tony McEnery. 2006. Collocation, semantic prosody, and near synonymy: A crosslinguistic perspective. *Applied Linguistics* 27(1): 103-129
- Yamashita, Junko and Nan Jiang. 2010. L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly* 44(4): 647-668.
- Yasuda, Sachiko. 2010. Learning phrasal verbs through conceptual metaphors: A case of Japanese EFL learners. *TESOL* 44(2): 250-273.
- Yazıcı, İlkey Çelik and Başak Tan Irtes. 2014. A real dilemma: Countable or uncountable? *Asian Journal of Education and e-Learning* 2(2): 159-164.
- Yıldız, Mustafa. 2016a. A cross-linguistic inquiry into the potential reasons for the avoidance of English phrasal verbs: The case of Turkish and Norwegian EFL learners. *The Linguistics Journal* 10(1): 124-140.
- _____ 2016b. Contrastive analysis of Turkish and English in Turkish EFL learners' spoken discourse. *International Journal of English Studies* 16(1): 57-74.
- Yoon, Hyung-Jo. 2016. Association strength of verb-noun combinations in experienced NS and less experienced NNS writing: Longitudinal and cross-sectional findings. *Journal of Second Language Writing* 34: 42-57.
- Yorio, Carlos A. 1989. 'Idiomaticity as an indicator of second language proficiency.' In Kenneth Hyldenstam and Loraine K. Obler (eds.), *Bilingualism across the lifespan* (pp. 55-72). Cambridge: Cambridge University Press.
- Yule, George. 1997. *Referential communication tasks*. Mahwah, NJ: Erlbaum.

-
- Zareva, Alla. 2016. Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal* 71(1): 69-79.
- Zhou, Xiujuan. 2016. A corpus-based study on high frequency verb collocations in the case of "HAVE". *International Forum of Teaching and Studies* 12(1): 42-50.
- Zok, David. 2010. Turkey's language revolution and the status of English today. *The English Languages: History, Diaspora, Culture* 1(1): 1-14. Retrieved August 27, 2017 from <http://jps.library.utoronto.ca/index.php/elhdc/article/view/14300/11349>
- Zughoul, Muhammad Raji. 1991. Lexical choice: Towards writing problematic word lists. *International Review of Applied Linguistics* 29(1): 45-60.

Corpora, dictionaries, and software

- Anthony, Laurence. 2014. AntConc (version 3.4.3) [computer software]. Tokyo, Japan: Waseda University. Available at <http://www.antlab.sci.waseda.ac.jp/>
- Cambridge English Dictionary Online. 2020.
Accessible at <https://dictionary.cambridge.org/>
- Cowie, Anthony P. and Ronald Mackin. 1975. *Oxford dictionary of current idiomatic English. Volume 1. Verbs with prepositions and particles*. Berlin: Cornelsen/Oxford University Press.
- Crowther, Jonathan, Sheila Dignen and Diana Lea (eds.) 2002. *Oxford collocations dictionary for students of English*. Oxford: Oxford University Press.
- Cullen, Kay and Howard Sargeant. 1996. *Chambers dictionary of phrasal verbs*. Edinburg: Chambers Harrap.

Davies, Mark. (2008-) *The Corpus of Contemporary American English: 520 million words, 1990-present*. Available online at <http://corpus.byu.edu/coca/>.

Davies, Mark. (2013) *Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries*. Available online at <http://corpus.byu.edu/glowbe/>.

Granger, Sylviane, Estelle Dagneaux, Fanny Meunier and Magali Paquot (eds.) 2009. *International Corpus of Learner English: Version 2*. Louvain-la-Neuve, Belgium: Presses universitaires de Louvain.

Hill, Jimmie and Michael Lewis (eds) 1997. *LTP dictionary of selected collocations*. Hove: LTP Language.

Louvain Corpus of Native English Essays.

Accessible at <https://uclouvain.be/en/research-institutes/ilc/cecl/locness.html>

Rundell, Michael (ed. in chief). 2005. *Macmillan phrasal verbs plus*. Oxford: Macmillan.

Sinclair, John McCardy (ed. in chief). 1989. *Collins COBUILD dictionary of phrasal verbs*. London/Glasgow: Collins.

Wehmeier, Sally (ed.). 2000. *Oxford advanced learner's dictionary of current English* (6th ed.). Oxford: Oxford University Press.

Appendices

Appendix 1. Frequency list of phrasal verb types per million words (raw)

TICLE		LOCNESS ¹⁵⁹	
Phrasal verb	Frequency	Phrasal verb	Frequency
<i>go on</i>	284 (57)	<i>go on</i>	201 (53)
<i>sum up</i>	199 (40)	<i>carry out</i>	182 (48)
<i>grow up</i>	174 (35)	<i>point out</i>	155 (41)
<i>go out</i>	144 (29)	<i>take away</i>	117 (31)
<i>bring up</i>	114 (23)	<i>bring up</i>	110 (29)
<i>carry out</i>	104 (21)	<i>take on</i>	102 (27)
<i>give up</i>	85 (17)	<i>end up</i>	98 (26)
<i>come together</i>	35 (7)	<i>grow up</i>	98 (26)
<i>go back</i>	35 (7)	<i>give up</i>	95 (25)
<i>find out</i>	30 (6)	<i>bring about</i>	87 (23)
<i>get on</i>	30 (6)	<i>find out</i>	72 (19)
<i>point out</i>	30 (6)	<i>make up</i>	68 (18)
<i>take away</i>	30 (6)	<i>set up</i>	64 (17)
<i>wake up</i>	30 (6)	<i>go back</i>	61 (16)
<i>keep on</i>	25 (5)	<i>break down</i>	53 (14)
<i>turn out</i>	25 (5)	<i>get away</i>	53 (14)
<i>build up</i>	20 (4)	<i>cut off</i>	45 (12)
<i>come back</i>	20 (4)	<i>be out</i>	45 (12)
<i>come over</i>	20 (4)	<i>bring in</i>	42 (11)
<i>lock up</i>	20 (4)	<i>carry on</i>	42 (11)
<i>break out</i>	15 (3)	<i>go out</i>	42 (11)
<i>break up</i>	15 (3)	<i>run up</i>	42 (11)
<i>bring about</i>	15 (3)	<i>turn out</i>	42 (11)
<i>bring back</i>	15 (3)	<i>fit in</i>	38 (10)
<i>come out</i>	15 (3)	<i>get out</i>	38 (10)
<i>end up</i>	15 (3)	<i>set out</i>	38 (10)
<i>go away</i>	15 (3)	<i>be around</i>	38 (10)
<i>leak out</i>	15 (3)	<i>back up</i>	34 (9)

¹⁵⁹ The figures for LOCNESS were taken from Waibel (2007: 188ff).

TICLE (cont.)		LOCNESS (cont.)	
Phrasal verb	Frequency	Phrasal Verb	Frequency
<i>make up</i>	15 (3)	<i>bring back</i>	34 (9)
<i>pass over</i>	15 (3)	<i>come out</i>	34 (9)
<i>raise up</i>	15 (3)	<i>put forward</i>	34 (9)
<i>turn back</i>	15 (3)	<i>start out</i>	34 (9)
<i>turn on</i>	15 (3)	<i>build up</i>	30 (8)
<i>wash up</i>	15 (3)	<i>come up</i>	30 (8)
<i>work out</i>	15 (3)	<i>get back</i>	30 (8)
<i>bring on</i>	10 (2)	<i>give back</i>	30 (8)
<i>bring together</i>	10 (2)	<i>hold up</i>	30 (8)
<i>calm down</i>	10 (2)	<i>look down</i>	30 (8)
<i>even up</i>	10 (2)	<i>take over</i>	30 (8)
<i>fall apart</i>	10 (2)	<i>bring out</i>	27 (7)
<i>fall down</i>	10 (2)	<i>come about</i>	27 (7)
<i>get up</i>	10 (2)	<i>come back</i>	27 (7)
<i>hold up</i>	10 (2)	<i>draw up</i>	27 (7)
<i>lay out</i>	10 (2)	<i>pay off</i>	27 (7)
<i>move up</i>	10 (2)	<i>pick up</i>	27 (7)
<i>put down</i>	10 (2)	<i>speed up</i>	27 (7)
<i>run away</i>	10 (2)	<i>stand up</i>	27 (7)
<i>sit down</i>	10 (2)	<i>take up</i>	27 (7)
<i>spread out</i>	10 (2)	<i>break away</i>	23 (6)
<i>stay behind</i>	10 (2)	<i>bring down</i>	23 (6)
<i>stick out</i>	10 (2)	<i>bring together</i>	23 (6)
<i>take on</i>	10 (2)	<i>go down</i>	23 (6)
<i>take out</i>	10 (2)	<i>open up</i>	23 (6)
<i>take over</i>	10 (2)	<i>slow down</i>	23 (6)
<i>take up</i>	10 (2)	<i>start off</i>	23 (6)
<i>wash out</i>	10 (2)	<i>take out</i>	23 (6)
<i>act out</i>	5 (1)	<i>throw out</i>	23 (6)
<i>add up</i>	5 (1)	<i>turn off</i>	23 (6)
<i>back up</i>	5 (1)	<i>come together</i>	19 (5)
<i>be apart</i>	5 (1)	<i>cut down</i>	19 (5)
<i>be off</i>	5 (1)	<i>face up</i>	19 (5)
<i>be out</i>	5 (1)	<i>fall back</i>	19 (5)
<i>be over</i>	5 (1)	<i>hand over</i>	19 (5)
<i>breathe in</i>	5 (1)	<i>lay off</i>	19 (5)
<i>carry on</i>	5 (1)	<i>look back</i>	19 (5)
<i>collapse down</i>	5 (1)	<i>pass on</i>	19 (5)

TICLE (cont.)		LOCNESS (cont.)	
Phrasal verb	Frequency	Phrasal Verb	Frequency
<i>come along</i>	5 (1)	<i>put up</i>	19 (5)
<i>come up</i>	5 (1)	<i>turn on</i>	19 (5)
<i>cover up</i>	5 (1)	<i>wake up</i>	19 (5)
<i>cut down</i>	5 (1)	<i>act out</i>	15 (4)
<i>dig up</i>	5 (1)	<i>bring on</i>	15 (4)
<i>divide up</i>	5 (1)	<i>get on</i>	15 (4)
<i>drive away</i>	5 (1)	<i>hold on</i>	15 (4)
<i>fall away</i>	5 (1)	<i>keep out</i>	15 (4)
<i>fall out</i>	5 (1)	<i>move forward</i>	15 (4)
<i>figure out</i>	5 (1)	<i>move on</i>	15 (4)
<i>fill up</i>	5 (1)	<i>put forth</i>	15 (4)
<i>fit in</i>	5 (1)	<i>slip back</i>	15 (4)
<i>fly away</i>	5 (1)	<i>sort out</i>	15 (4)
<i>get down</i>	5 (1)	<i>turn away</i>	15 (4)
<i>get together</i>	5 (1)	<i>break apart</i>	11 (3)
<i>give back</i>	5 (1)	<i>carry through</i>	11 (3)
<i>give off</i>	5 (1)	<i>catch up</i>	11 (3)
<i>give out</i>	5 (1)	<i>come along</i>	11 (3)
<i>give over</i>	5 (1)	<i>come off</i>	11 (3)
<i>go by</i>	5 (1)	<i>drop out</i>	11 (3)
<i>go down</i>	5 (1)	<i>explain away</i>	11 (3)
<i>go forward</i>	5 (1)	<i>get up</i>	11 (3)
<i>go round</i>	5 (1)	<i>get down</i>	11 (3)
<i>go up</i>	5 (1)	<i>get ahead</i>	11 (3)
<i>hang on</i>	5 (1)	<i>give away</i>	11 (3)
<i>heap up</i>	5 (1)	<i>give in</i>	11 (3)
<i>*highten up</i>	5 (1)	<i>go away</i>	11 (3)
<i>hike up</i>	5 (1)	<i>go through</i>	11 (3)
<i>keep together</i>	5 (1)	<i>go by</i>	11 (3)
<i>knit up</i>	5 (1)	<i>kick out</i>	11 (3)
<i>leave aside</i>	5 (1)	<i>lead away</i>	11 (3)
<i>leave behind</i>	5 (1)	<i>leave behind</i>	11 (3)
<i>leave out</i>	5 (1)	<i>line up</i>	11 (3)
<i>let off</i>	5 (1)	<i>look up</i>	11 (3)
<i>let out</i>	5 (1)	<i>lose out</i>	11 (3)
<i>live on</i>	5 (1)	<i>make out</i>	11 (3)
<i>lock down</i>	5 (1)	<i>miss out</i>	11 (3)
<i>look back</i>	5 (1)	<i>move away</i>	11 (3)

TICLE (cont.)		LOCNESS (cont.)	
Phrasal verb	Frequency	Phrasal Verb	Frequency
<i>marry off</i>	5 (1)	<i>play out</i>	11 (3)
<i>move away</i>	5 (1)	<i>pull out</i>	11 (3)
<i>pass by</i>	5 (1)	<i>put on</i>	11 (3)
<i>print out</i>	5 (1)	<i>put off</i>	11 (3)
<i>put forward</i>	5 (1)	<i>rip away</i>	11 (3)
<i>put on</i>	5 (1)	<i>run away</i>	11 (3)
<i>put out</i>	5 (1)	<i>run out</i>	11 (3)
<i>put up</i>	5 (1)	<i>seek out</i>	11 (3)
<i>return back</i>	5 (1)	<i>set apart</i>	11 (3)
<i>rise up</i>	5 (1)	<i>sign away</i>	11 (3)
<i>rule out</i>	5 (1)	<i>single out</i>	11 (3)
<i>run around</i>	5 (1)	<i>stand out</i>	11 (3)
<i>seek out</i>	5 (1)	<i>take off</i>	11 (3)
<i>set forth</i>	5 (1)	<i>think out</i>	11 (3)
<i>set up</i>	5 (1)	<i>throw away</i>	11 (3)
<i>settle down</i>	5 (1)	<i>turn around</i>	11 (3)
<i>sort out</i>	5 (1)	<i>turn back</i>	11 (3)
<i>stand by</i>	5 (1)	<i>walk in</i>	11 (3)
<i>stand on</i>	5 (1)	<i>watch out</i>	11 (3)
<i>stand together</i>	5 (1)	<i>write down</i>	11 (3)
<i>stand up</i>	5 (1)	<i>back out</i>	8 (2)
<i>start over</i>	5 (1)	<i>band together</i>	8 (2)
<i>stick through</i>	5 (1)	<i>blow away</i>	8 (2)
<i>take back</i>	5 (1)	<i>bog down</i>	8 (2)
<i>take in</i>	5 (1)	<i>break out</i>	8 (2)
<i>take off</i>	5 (1)	<i>clean up</i>	8 (2)
<i>tell off</i>	5 (1)	<i>clear up</i>	8 (2)
<i>throw out</i>	5 (1)	<i>cling on</i>	8 (2)
<i>try out</i>	5 (1)	<i>come across</i>	8 (2)
<i>wither away</i>	5 (1)	<i>come around</i>	8 (2)
<i>word off</i>	5 (1)	<i>come forth</i>	8 (2)
		<i>come in</i>	8 (2)
		<i>cry out</i>	8 (2)
		<i>cut back</i>	8 (2)
		<i>die down</i>	8 (2)
		<i>fight back</i>	8 (2)
		<i>figure out</i>	8 (2)
		<i>fill out</i>	8 (2)

		LOCNESS (cont.)	
		Phrasal Verb	Frequency
		<i>follow up</i>	8 (2)
		<i>go along</i>	8 (2)
		<i>go off</i>	8 (2)
		<i>hit out</i>	8 (2)
		<i>hold down</i>	8 (2)
		<i>join together</i>	8 (2)
		<i>keep up</i>	8 (2)
		<i>lay down</i>	8 (2)
		<i>lay out</i>	8 (2)
		<i>leave out</i>	8 (2)
		<i>let down</i>	8 (2)
		<i>link together</i>	8 (2)
		<i>live out</i>	8 (2)
		<i>look forward</i>	8 (2)
		<i>look over</i>	8 (2)
		<i>move out</i>	8 (2)
		<i>pass down</i>	8 (2)
		<i>pay back</i>	8 (2)
		<i>pick out</i>	8 (2)
		<i>put together</i>	8 (2)
		<i>put aside</i>	8 (2)
		<i>revert back</i>	8 (2)
		<i>roll down</i>	8 (2)
		<i>run back</i>	8 (2)
		<i>rush around</i>	8 (2)
		<i>seize back</i>	8 (2)
		<i>sell out</i>	8 (2)
		<i>shake off</i>	8 (2)
		<i>show up</i>	8 (2)
		<i>sit down</i>	8 (2)
		<i>spring up</i>	8 (2)
		<i>stay away</i>	8 (2)
		<i>strike back</i>	8 (2)
		<i>take back</i>	8 (2)
		<i>tear apart</i>	8 (2)
		<i>tear away</i>	8 (2)
		<i>tie together</i>	8 (2)
		<i>trace back</i>	8 (2)

		LOCNESS (cont.)	
		Phrasal Verb	Frequency
		<i>tune in</i>	8 (2)
		<i>turn in</i>	8 (2)
		<i>turn over</i>	8 (2)
		<i>walk out</i>	8 (2)
		<i>want back</i>	8 (2)
		<i>whip up</i>	8 (2)
		<i>win over</i>	8 (2)
		<i>wipe out</i>	8 (2)
		<i>work out</i>	8 (2)
		<i>be away</i>	8 (2)
		<i>be back</i>	8 (2)
		<i>be off</i>	8 (2)
		<i>be on</i>	8 (2)
		<i>be over</i>	8 (2)
		<i>be up</i>	8 (2)
		<i>have over</i>	8 (2)
		<i>allow in</i>	4 (1)
		<i>allow back</i>	4 (1)
		<i>ash back</i>	4 (1)
		<i>back off</i>	4 (1)
		<i>be down</i>	4 (1)
		<i>bear out</i>	4 (1)
		<i>beat up</i>	4 (1)
		<i>bind together</i>	4 (1)
		<i>block out</i>	4 (1)
		<i>blow out</i>	4 (1)
		<i>boil up</i>	4 (1)
		<i>bottle up</i>	4 (1)
		<i>bounce back</i>	4 (1)
		<i>buckle up</i>	4 (1)
		<i>break up</i>	4 (1)
		<i>bring forth</i>	4 (1)
		<i>bring over</i>	4 (1)
		<i>call back</i>	4 (1)
		<i>call forth</i>	4 (1)
		<i>call in</i>	4 (1)
		<i>call out</i>	4 (1)
		<i>carry away</i>	4 (1)

		LOCNESS (cont.)	
		Phrasal Verb	Frequency
		<i>carry over</i>	4 (1)
		<i>check out</i>	4 (1)
		<i>check up</i>	4 (1)
		<i>chill out</i>	4 (1)
		<i>chip in</i>	4 (1)
		<i>churn out</i>	4 (1)
		<i>combine to- gether</i>	4 (1)
		<i>come by</i>	4 (1)
		<i>contract out</i>	4 (1)
		<i>cover up</i>	4 (1)
		<i>crack down</i>	4 (1)
		<i>cram in</i>	4 (1)
		<i>cut up</i>	4 (1)
		<i>date back</i>	4 (1)
		<i>divide down</i>	4 (1)
		<i>divvy up</i>	4 (1)
		<i>drive around</i>	4 (1)
		<i>drive out</i>	4 (1)
		<i>dig up</i>	4 (1)
		<i>do back</i>	4 (1)
		<i>drag up</i>	4 (1)
		<i>erode away</i>	4 (1)
		<i>fade away</i>	4 (1)
		<i>fall down</i>	4 (1)
		<i>feed back</i>	4 (1)
		<i>fill up</i>	4 (1)
		<i>filter over</i>	4 (1)
		<i>flame on</i>	4 (1)
		<i>flare up</i>	4 (1)
		<i>flow in</i>	4 (1)
		<i>focus in</i>	4 (1)
		<i>follow back</i>	4 (1)
		<i>follow through</i>	4 (1)
		<i>free up</i>	4 (1)
		<i>gather together</i>	4 (1)
		<i>get across</i>	4 (1)
		<i>get over</i>	4 (1)

		LOCNESS (cont.)	
		Phrasal Verb	Frequency
		<i>get through</i>	4 (1)
		<i>get together</i>	4 (1)
		<i>go ahead</i>	4 (1)
		<i>go around</i>	4 (1)
		<i>go forward</i>	4 (1)
		<i>go in</i>	4 (1)
		<i>go under</i>	4 (1)
		<i>go up</i>	4 (1)
		<i>gouge out</i>	4 (1)
		<i>group together</i>	4 (1)
		<i>hand down</i>	4 (1)
		<i>hand out</i>	4 (1)
		<i>hang on</i>	4 (1)
		<i>heat up</i>	4 (1)
		<i>help out</i>	4 (1)
		<i>hold back</i>	4 (1)
		<i>hold in</i>	4 (1)
		<i>hold out</i>	4 (1)
		<i>hold together</i>	4 (1)
		<i>hook up</i>	4 (1)
		<i>join up</i>	4 (1)
		<i>jump in</i>	4 (1)
		<i>keep apart</i>	4 (1)
		<i>keep down</i>	4 (1)
		<i>keep together</i>	4 (1)
		<i>knock down</i>	4 (1)
		<i>lag behind</i>	4 (1)
		<i>lash out</i>	4 (1)
		<i>lead up</i>	4 (1)
		<i>let off</i>	4 (1)
		<i>lie ahead</i>	4 (1)
		<i>lie down</i>	4 (1)
		<i>lie around</i>	4 (1)
		<i>light up</i>	4 (1)
		<i>linger on</i>	4 (1)
		<i>live on</i>	4 (1)
		<i>look on</i>	4 (1)
		<i>lop off</i>	4 (1)

		LOCNESS (cont.)	
		Phrasal Verb	Frequency
		<i>lure in</i>	4 (1)
		<i>march in</i>	4 (1)
		<i>mark up</i>	4 (1)
		<i>measure up</i>	4 (1)
		<i>meet up</i>	4 (1)
		<i>melt together</i>	4 (1)
		<i>merge together</i>	4 (1)
		<i>mix in</i>	4 (1)
		<i>mix together</i>	4 (1)
		<i>move ahead</i>	4 (1)
		<i>move along</i>	4 (1)
		<i>own up</i>	4 (1)
		<i>pair up</i>	4 (1)
		<i>pass by</i>	4 (1)
		<i>pass out</i>	4 (1)
		<i>pass up</i>	4 (1)
		<i>pay out</i>	4 (1)
		<i>persuade away</i>	4 (1)
		<i>pile on</i>	4 (1)
		<i>plan out</i>	4 (1)
		<i>pop up</i>	4 (1)
		<i>press ahead</i>	4 (1)
		<i>price out</i>	4 (1)
		<i>print out</i>	4 (1)
		<i>push aside</i>	4 (1)
		<i>push away</i>	4 (1)
		<i>push back</i>	4 (1)
		<i>push forward</i>	4 (1)
		<i>push off</i>	4 (1)
		<i>push out</i>	4 (1)
		<i>put away</i>	4 (1)
		<i>put back</i>	4 (1)
		<i>put in</i>	4 (1)
		<i>rain in</i>	4 (1)
		<i>reach out</i>	4 (1)
		<i>read out</i>	4 (1)
		<i>refer back</i>	4 (1)
		<i>rip apart</i>	4 (1)

		LOCNESS (cont.)	
		Phrasal Verb	Frequency
		<i>rip off</i>	4 (1)
		<i>roam around</i>	4 (1)
		<i>roll around</i>	4 (1)
		<i>roll away</i>	4 (1)
		<i>round up</i>	4 (1)
		<i>rule out</i>	4 (1)
		<i>run around</i>	4 (1)
		<i>run about</i>	4 (1)
		<i>run off</i>	4 (1)
		<i>rush out</i>	4 (1)
		<i>scout out</i>	4 (1)
		<i>scrape by</i>	4 (1)
		<i>scream out</i>	4 (1)
		<i>screw up</i>	4 (1)
		<i>send away</i>	4 (1)
		<i>send back</i>	4 (1)
		<i>send out</i>	4 (1)
		<i>set aside</i>	4 (1)
		<i>set down</i>	4 (1)
		<i>set in</i>	4 (1)
		<i>shine through</i>	4 (1)
		<i>ship out</i>	4 (1)
		<i>shoot back</i>	4 (1)
		<i>shoot down</i>	4 (1)
		<i>shove off</i>	4 (1)
		<i>show off</i>	4 (1)
		<i>shrug off</i>	4 (1)
		<i>shut down</i>	4 (1)
		<i>shy away</i>	4 (1)
		<i>sign in</i>	4 (1)
		<i>sign on</i>	4 (1)
		<i>sit around</i>	4 (1)
		<i>sit back</i>	4 (1)
		<i>sit in</i>	4 (1)
		<i>slave away</i>	4 (1)
		<i>slide with</i>	4 (1)
		<i>snap out</i>	4 (1)
		<i>sneak around</i>	4 (1)

		LOCNESS (cont.)	
		Phrasal Verb	Frequency
		<i>snuff out</i>	4 (1)
		<i>spark off</i>	4 (1)
		<i>spark up</i>	4 (1)
		<i>spread out</i>	4 (1)
		<i>spur on</i>	4 (1)
		<i>stamp out</i>	4 (1)
		<i>stand around</i>	4 (1)
		<i>stand back</i>	4 (1)
		<i>stand down</i>	4 (1)
		<i>stay on</i>	4 (1)
		<i>steer away</i>	4 (1)
		<i>stem back</i>	4 (1)
		<i>step back</i>	4 (1)
		<i>step down</i>	4 (1)
		<i>step forward</i>	4 (1)
		<i>step in</i>	4 (1)
		<i>step out</i>	4 (1)
		<i>step up</i>	4 (1)
		<i>stress out</i>	4 (1)
		<i>strewn about</i>	4 (1)
		<i>suck up</i>	4 (1)
		<i>sum up</i>	4 (1)
		<i>swoop down</i>	4 (1)
		<i>table up</i>	4 (1)
		<i>take down</i>	4 (1)
		<i>take in</i>	4 (1)
		<i>talk back</i>	4 (1)
		<i>throw back</i>	4 (1)
		<i>tie up</i>	4 (1)
		<i>tip off</i>	4 (1)
		<i>tone down</i>	4 (1)
		<i>toss out</i>	4 (1)
		<i>train out</i>	4 (1)
		<i>travel back</i>	4 (1)
		<i>tune out</i>	4 (1)
		<i>turn up</i>	4 (1)
		<i>veer away</i>	4 (1)
		<i>walk away</i>	4 (1)

LOCNESS (cont.)			
		Phrasal Verb	Frequency
		<i>walk by</i>	4 (1)
		<i>walk on</i>	4 (1)
		<i>ward off</i>	4 (1)
		<i>weed out</i>	4 (1)
		<i>weigh down</i>	4 (1)
		<i>wind up</i>	4 (1)
		<i>work off</i>	4 (1)
		<i>wrap up</i>	4 (1)
		<i>wear away</i>	4 (1)

Appendix 2. Frequency list of phrasal-prepositional verb types per million words (raw)

TICLE		LOCNESS	
Phrasal-prepositional verb	Frequency	Phrasal-prepositional verb	Frequency
<i>*come across with</i>	45 (9)	<i>come down to</i>	15 (4)
<i>get on with</i>	30 (6)	<i>do away with</i>	11 (3)
<i>get along with</i>	25 (5)		
<i>come up with</i>	20 (4)		
<i>keep away from</i>	20 (4)		
<i>*give up from</i>	15 (3)		
<i>get back to</i>	10 (2)		
<i>put up with</i>	10 (2)		
<i>lead up to</i>	10 (2)		
<i>*face up with</i>	10 (2)		
<i>be up to</i>	5 (1)		
<i>*bring about with</i>	5 (1)		
<i>*come up across</i>	5 (1)		
<i>*come with up</i>	5 (1)		
<i>get away from</i>	5 (1)		
<i>get away with</i>	5 (1)		
<i>get down to</i>	5 (1)		
<i>give back to</i>	5 (1)		
<i>give up on</i>	5 (1)		
<i>go back to</i>	5 (1)		

TICLE		LOCNESS	
Phrasal-prepositional verb	Frequency	Phrasal-prepositional verb	Frequency
<i>go down to</i>	5 (1)		
<i>go on with</i>	5 (1)		
<i>go together with</i>	5 (1)		
<i>live apart from</i>	5 (1)		
<i>pass on to</i>	5 (1)		
<i>*run across with</i>	5 (1)		
<i>run away from</i>	5 (1)		
<i>stand up for</i>	5 (1)		
<i>strap down to</i>	5 (1)		
<i>take over from</i>	5 (1)		
<i>catch up with</i>	5 (1)		
<i>cling on to</i>	5 (1)		
<i>do away with</i>	5 (1)		
<i>look forward to</i>	5 (1)		
<i>pair up with</i>	5 (1)		

Appendix 3. Productive¹⁶⁰ verbs that combine with particles to form phrasal and phrasal-prepositional verbs

TICLE		LOCNESS	
verb	number of particles it combines with	verb	number of particles it combines with
<i>go</i>	10	<i>go</i>	15
<i>come</i>	8	<i>come</i>	13
<i>take</i>	8	<i>get</i>	11
<i>get</i>	7	<i>bring</i>	10
<i>bring</i>	5	<i>put</i>	10
<i>give</i>	5	<i>take</i>	9
<i>put</i>	5	<i>turn</i>	9
<i>fall</i>	4	<i>be</i>	9
<i>stand</i>	4	<i>hold</i>	7
<i>be</i>	3	<i>run</i>	7

¹⁶⁰ Verbs that combine with at least three different particles to form phrasal and phrasal-prepositional verbs.

TICLE		LOCNESS	
verb	number of particles it combines with	verb	number of particles it combines with
<i>keep</i>	3	<i>look</i>	6
<i>leave</i>	3	<i>move</i>	6
<i>pass</i>	3	<i>push</i>	6
<i>run</i>	3	<i>set</i>	6
<i>turn</i>	3	<i>step</i>	6
		<i>break</i>	5
		<i>carry</i>	5
		<i>keep</i>	5
		<i>pass</i>	5
		<i>stand</i>	5
		<i>walk</i>	5
		<i>call</i>	4
		<i>cut</i>	4
		<i>give</i>	4
		<i>sit</i>	4
		<i>back</i>	3
		<i>follow</i>	3
		<i>hand</i>	3
		<i>lay</i>	3
		<i>lie</i>	3
		<i>pay</i>	3
		<i>rip</i>	3
		<i>roll</i>	3
		<i>send</i>	3
		<i>sign</i>	3
		<i>throw</i>	3

Appendix 4. Particles

Appendix 4a. Particle productivity (combination with X verb types)

TICLE		LOCNESS	
particle	productivity	particle	productivity
<i>up</i>	39	<i>out</i>	84
<i>out</i>	28	<i>up</i>	70
<i>on</i>	13	<i>back</i>	41
<i>down</i>	11	<i>down</i>	34
<i>away</i>	12	<i>away</i>	32
<i>back</i>	9	<i>in</i>	26
<i>off</i>	7	<i>off</i>	25
<i>together</i>	6	<i>on</i>	23
<i>over</i>	5	<i>together</i>	17
<i>apart</i>	3	<i>around</i>	13
<i>by</i>	3	<i>over</i>	11
<i>in</i>	3	<i>forward</i>	6
<i>forward</i>	3	<i>ahead</i>	5
<i>across</i>	2	<i>apart</i>	5
<i>along</i>	2	<i>through</i>	5
<i>behind</i>	2	<i>by</i>	5
<i>about</i>	1	<i>about</i>	4
<i>around</i>	1	<i>forth</i>	4
<i>aside</i>	1	<i>aside</i>	3
<i>forth</i>	1	<i>along</i>	3
<i>round</i>	1	<i>behind</i>	2
<i>through</i>	1	<i>across</i>	2
<i>with</i>	1	<i>with</i>	1
		<i>under</i>	1

Appendix 4b. Frequency of particles per million words

TICLE		LOCNESS	
particle	frequency (pmw)	particle	frequency (pmw)
<i>up</i>	935	<i>out</i>	1201
<i>out</i>	512	<i>up</i>	1163
<i>on</i>	443	<i>on</i>	508
<i>away</i>	124	<i>back</i>	409
<i>back</i>	124	<i>away</i>	405
<i>down</i>	85	<i>down</i>	333
<i>together</i>	65	<i>off</i>	246
<i>over</i>	60	<i>in</i>	197
<i>across</i>	50	<i>about</i>	121
<i>off</i>	35	<i>together</i>	117
<i>along</i>	30	<i>over</i>	102
<i>about</i>	20	<i>around</i>	98
<i>apart</i>	20	<i>forward</i>	68
<i>behind</i>	15	<i>apart</i>	38
<i>by</i>	15	<i>through</i>	34
<i>in</i>	15	<i>forth</i>	30
<i>forward</i>	15	<i>ahead</i>	27
<i>around</i>	5	<i>by</i>	27
<i>aside</i>	5	<i>along</i>	23
<i>forth</i>	5	<i>aside</i>	15
<i>round</i>	5	<i>behind</i>	15
<i>through</i>	5	<i>with</i>	4
<i>with</i>	5	<i>under</i>	4
		<i>across</i>	4

Appendix 5. Frequency list of prepositional verb types in TICLE (raw)

TICLE			
Prepositional verb	Frequency	Prepositional verb	Frequency
<i>go to</i>	115	<i>push into</i>	2
<i>prepare for</i>	90	<i>*react for</i>	2
<i>look at</i>	72	<i>release from</i>	2
<i>see as</i>	62	<i>replace by</i>	2
<i>think of</i>	56	<i>result with</i>	2
<i>think about</i>	55	<i>*serve to</i>	2
<i>depend on</i>	54	<i>source from</i>	2
<i>look after</i>	50	<i>speak of</i>	2
<i>graduate from</i>	48	<i>specialize in</i>	2
<i>talk about</i>	44	<i>stand against</i>	2
<i>deal with</i>	40	<i>to start with</i>	2
<i>suffer from</i>	32	<i>stay with</i>	2
<i>*prepare to</i>	30	<i>*support to</i>	2
<i>apply to</i>	26	<i>sympathize with</i>	2
<i>base on</i>	26	<i>thank for</i>	2
<i>look for</i>	22	<i>*torture to</i>	2
<i>mention about</i>	21	<i>touch on</i>	2
<i>speak about</i>	20	<i>trust to</i>	2
<i>be against</i>	19	<i>*trust on</i>	2
<i>benefit from</i>	19	<i>try for</i>	2
<i>agree with</i>	17	<i>turn around</i>	2
<i>learn about</i>	17	<i>warn against</i>	2
<i>wait for</i>	17	<i>work on</i>	2
<i>come from</i>	16	<i>abstain from</i>	1
<i>expose to</i>	16	<i>act as</i>	1
<i>believe in</i>	15	<i>accommodate to</i>	1
<i>lead to</i>	15	<i>account for</i>	1
<i>send to</i>	15	<i>*achieve on</i>	1
<i>come to</i>	14	<i>act on</i>	1
<i>focus on</i>	14	<i>*add into</i>	1
<i>pay for</i>	14	<i>adhere to</i>	1
<i>belong to</i>	13	<i>agree upon</i>	1
<i>cope with</i>	13	<i>appoint to</i>	1
<i>*face with</i>	13	<i>*appreciate with</i>	1

TICLE			
Prepositional verb	Frequency	Prepositional verb	Frequency
<i>prevent from</i>	13	<i>*arrest of</i>	1
<i>spend for</i>	13	<i>arrive in</i>	1
<i>care about</i>	12	<i>ask about</i>	1
<i>compare with</i>	12	<i>*assign for</i>	1
<i>consist of</i>	12	<i>attach to</i>	1
<i>decide on</i>	12	<i>*attend in</i>	1
<i>escape from</i>	12	<i>attract to</i>	1
<i>*face to</i>	12	<i>be above</i>	1
<i>study for</i>	12	<i>be after</i>	1
<i>talk to</i>	12	<i>be for</i>	1
<i>struggle for</i>	11	<i>become of</i>	1
<i>regard as</i>	10	<i>*begin to</i>	1
<i>result in</i>	10	<i>*benefit for</i>	1
<i>save from</i>	10	<i>*blame of</i>	1
<i>struggle with</i>	10	<i>bombard with</i>	1
<i>compare to</i>	9	<i>build on</i>	1
<i>differ from</i>	9	<i>*call with</i>	1
<i>expect from</i>	9	<i>*care on</i>	1
<i>fight for</i>	9	<i>*care of</i>	1
<i>sentence to</i>	9	<i>*care with</i>	1
<i>supply with</i>	9	<i>*challenge with</i>	1
<i>adopt to</i>	8	<i>*challenge to</i>	1
<i>call as</i>	8	<i>charge for</i>	1
<i>come across</i>	8	<i>clash with</i>	1
<i>consider as</i>	8	<i>coincide with</i>	1
<i>define as</i>	8	<i>collide with</i>	1
<i>rely upon</i>	8	<i>combine with</i>	1
<i>search for</i>	8	<i>*come into</i>	1
<i>subject to</i>	8	<i>*compare to</i>	1
<i>*ask to</i>	7	<i>*compete to</i>	1
<i>ask for</i>	7	<i>comply with</i>	1
<i>break into</i>	7	<i>*compose with</i>	1
<i>compete with</i>	7	<i>concentrate on</i>	1
<i>concern with</i>	7	<i>conduct on</i>	1
<i>direct to</i>	7	<i>confine to</i>	1
<i>discuss about</i>	7	<i>*confine with</i>	1
<i>divide into</i>	7	<i>connect to</i>	1

TICLE			
Prepositional verb	Frequency	Prepositional verb	Frequency
<i>interfere in</i>	7	<i>*consider of</i>	1
<i>live in</i>	7	<i>*consist with</i>	1
<i>*respect to</i>	7	<i>consort to</i>	1
<i>talk with</i>	7	<i>*constitute of</i>	1
<i>turn into</i>	7	<i>constitute from</i>	1
<i>adapt to</i>	6	<i>construct on</i>	1
<i>apply for</i>	6	<i>*consult to</i>	1
<i>complain about</i>	6	<i>content with</i>	1
<i>confront with</i>	6	<i>convict of</i>	1
<i>end with</i>	6	<i>correspond with</i>	1
<i>involve in</i>	6	<i>*cost to</i>	1
<i>object to</i>	6	<i>cover with</i>	1
<i>protect from</i>	6	<i>cry for</i>	1
<i>provide with</i>	6	<i>*date with</i>	1
<i>reach to</i>	6	<i>deduce from</i>	1
<i>resort to</i>	6	<i>*defend toward</i>	1
<i>result from</i>	6	<i>depart from</i>	1
<i>spend on</i>	6	<i>depend upon</i>	1
<i>spend with</i>	6	<i>despair of</i>	1
<i>supply to</i>	6	<i>detect in</i>	1
<i>blame for</i>	5	<i>deter from</i>	1
<i>deprive of</i>	5	<i>develop into</i>	1
<i>describe as</i>	5	<i>deviate from</i>	1
<i>disagree on</i>	5	<i>diagnose as</i>	1
<i>fight with</i>	5	<i>die for</i>	1
<i>hear from</i>	5	<i>*die through</i>	1
<i>help to</i>	5	<i>differentiate from</i>	1
<i>interact with</i>	5	<i>direct at</i>	1
<i>lie behind</i>	5	<i>direct towards</i>	1
<i>lie under</i>	5	<i>disappear from</i>	1
<i>live under</i>	5	<i>disapprove of</i>	1
<i>meet with</i>	5	<i>discriminate from</i>	1
<i>remove from</i>	5	<i>*disgrace of</i>	1
<i>return to</i>	5	<i>divert into</i>	1
<i>separate from</i>	5	<i>divert to</i>	1
<i>teach about</i>	5	<i>divide among</i>	1
<i>transfer (from) to</i>	5	<i>divide between</i>	1

TICLE			
Prepositional verb	Frequency	Prepositional verb	Frequency
<i>worry about</i>	5	<i>*divorce from</i>	1
<i>add to</i>	4	<i>*dominate over</i>	1
<i>agree on</i>	4	<i>donate with</i>	1
<i>appeal to</i>	4	<i>dream about</i>	1
<i>appoint as</i>	4	<i>drive into</i>	1
<i>argue about</i>	4	<i>emerge from</i>	1
<i>arise from</i>	4	<i>*encounter with</i>	1
<i>contribute to</i>	4	<i>*enroll to</i>	1
<i>equip with</i>	4	<i>enter for</i>	1
<i>get into</i>	4	<i>entitle to</i>	1
<i>go through</i>	4	<i>*envy of</i>	1
<i>graduate with</i>	4	<i>equate with</i>	1
<i>know about</i>	4	<i>examine as</i>	1
<i>live on</i>	4	<i>excile to</i>	1
<i>*marry with</i>	4	<i>experiment on</i>	1
<i>*need to</i>	4	<i>*explore with</i>	1
<i>read about</i>	4	<i>extricate from</i>	1
<i>say about</i>	4	<i>*face of</i>	1
<i>speak to</i>	4	<i>fall for</i>	1
<i>stem from</i>	4	<i>fight against</i>	1
<i>tell about</i>	4	<i>final in</i>	1
<i>turn to</i>	4	<i>fire from</i>	1
<i>*aim to</i>	3	<i>*focus to</i>	1
<i>assist in</i>	3	<i>forget about</i>	1
<i>*attack to</i>	3	<i>glide into</i>	1
<i>beg for</i>	3	<i>go in</i>	1
<i>care for</i>	3	<i>go with</i>	1
<i>change to</i>	3	<i>go after</i>	1
<i>change into</i>	3	<i>go around</i>	1
<i>charge with</i>	3	<i>go for</i>	1
<i>classify as</i>	3	<i>go by</i>	1
<i>comment on</i>	3	<i>gossip about</i>	1
<i>concern about</i>	3	<i>grant to</i>	1
<i>connect with</i>	3	<i>grapple with</i>	1
<i>*continue to</i>	3	<i>help with</i>	1
<i>cooperate with</i>	3	<i>hesitate about</i>	1
<i>decide for</i>	3	<i>inject with</i>	1

TICLE			
Prepositional verb	Frequency	Prepositional verb	Frequency
<i>decide about</i>	3	<i>*interest with</i>	1
<i>devote to</i>	3	<i>interfere with</i>	1
<i>disagree with</i>	3	<i>introduce to</i>	1
<i>discriminate against</i>	3	<i>*involve with</i>	1
<i>distinguish from</i>	3	<i>*join in</i>	1
<i>*enter to</i>	3	<i>keep to</i>	1
<i>enter in</i>	3	<i>*lead into</i>	1
<i>get to</i>	3	<i>leave to</i>	1
<i>go beyond</i>	3	<i>liken to</i>	1
<i>hear about</i>	3	<i>limit to</i>	1
<i>impose on</i>	3	<i>live through</i>	1
<i>insist on</i>	3	<i>look through</i>	1
<i>isolate from</i>	3	<i>*look forward</i>	1
<i>lie in</i>	3	<i>look over</i>	1
<i>live for</i>	3	<i>*look to</i>	1
<i>*oppose to</i>	3	<i>*marry to</i>	1
<i>participate in</i>	3	<i>operate on</i>	1
<i>prey on</i>	3	<i>pelt with</i>	1
<i>range from</i>	3	<i>perceive as</i>	1
<i>*resemble to</i>	3	<i>*perpetrate on</i>	1
<i>respond to</i>	3	<i>*persist on</i>	1
<i>say on</i>	3	<i>plot against</i>	1
<i>*seem as</i>	3	<i>plunge into</i>	1
<i>stand for</i>	3	<i>pray for</i>	1
<i>start from</i>	3	<i>put through</i>	1
<i>struggle against</i>	3	<i>quarrel with</i>	1
<i>*tell to</i>	3	<i>react to</i>	1
<i>trust in</i>	3	<i>*realize for</i>	1
<i>*utilize from</i>	3	<i>record from</i>	1
<i>adjust to</i>	2	<i>recover from</i>	1
<i>agree to</i>	2	<i>refer to</i>	1
<i>aim at</i>	2	<i>*regret of</i>	1
<i>approve of</i>	2	<i>relate to</i>	1
<i>argue on</i>	2	<i>relate with</i>	1
<i>argue for</i>	2	<i>rely on</i>	1
<i>arrive at</i>	2	<i>remind of</i>	1

TICLE			
Prepositional verb	Frequency	Prepositional verb	Frequency
<i>attend to</i>	2	<i>replace with</i>	1
<i>attribute to</i>	2	<i>rescue from</i>	1
<i>base upon</i>	2	<i>rest on</i>	1
<i>to begin with</i>	2	<i>rest upon</i>	1
<i>call for</i>	2	<i>rethink of</i>	1
<i>classify into</i>	2	<i>search on</i>	1
<i>conform to</i>	2	<i>seek for</i>	1
<i>*consider about</i>	2	<i>sell for</i>	1
<i>cut into</i>	2	<i>serve for</i>	1
<i>die from</i>	2	<i>serve with</i>	1
<i>disconnect from</i>	2	<i>speak on</i>	1
<i>discourage from</i>	2	<i>speculate over</i>	1
<i>discuss with</i>	2	<i>*spend to</i>	1
<i>doom to</i>	2	<i>spread to</i>	1
<i>drag into</i>	2	<i>step into</i>	1
<i>*drag to</i>	2	<i>strip of</i>	1
<i>dream of</i>	2	<i>strive for</i>	1
<i>elect as</i>	2	<i>subscribe to</i>	1
<i>end in</i>	2	<i>sue for</i>	1
<i>engage in</i>	2	<i>suffer through</i>	1
<i>enter into</i>	2	<i>suffer for</i>	1
<i>*esteem to</i>	2	<i>surrender to</i>	1
<i>explain to</i>	2	<i>talk for</i>	1
<i>fall into</i>	2	<i>temper with</i>	1
<i>function as</i>	2	<i>testify to</i>	1
<i>get over</i>	2	<i>*tip to</i>	1
<i>go into</i>	2	<i>*tolerate for</i>	1
<i>handle with</i>	2	<i>*touch to</i>	1
<i>identify with</i>	2	<i>train for</i>	1
<i>*impose to</i>	2	<i>transfer into</i>	1
<i>incorporate into</i>	2	<i>translate into</i>	1
<i>irradiate with</i>	2	<i>transmit to</i>	1
<i>*join to</i>	2	<i>*trigger for</i>	1
<i>keep from</i>	2	<i>unite with</i>	1
<i>long for</i>	2	<i>warn about</i>	1
<i>look around</i>	2	<i>warn of</i>	1
<i>press on</i>	2	<i>weigh on</i>	1

TICLE			
Prepositional verb	Frequency	Prepositional verb	Frequency
<i>profit from</i>	2	<i>weigh upon</i>	1
<i>protect against</i>	2	<i>withhold from</i>	1
		<i>wonder about</i>	1
		<i>work towards</i>	1
		<i>write about</i>	1

Appendix 6. Productive¹⁶¹ verbs that combine with prepositions to form prepositional verbs

TICLE			
verb	number of prepositions it combines with	verb	number of prepositions it combines with
<i>go</i>	10	<i>decide</i>	3
<i>look</i>	8	<i>result</i>	3
<i>live</i>	5	<i>ask</i>	3
<i>care</i>	5	<i>fight</i>	3
<i>talk</i>	4	<i>turn</i>	3
<i>come</i>	4	<i>call</i>	3
<i>speak</i>	4	<i>direct</i>	3
<i>spend</i>	4	<i>divide</i>	3
<i>agree</i>	4	<i>get</i>	3
<i>be</i>	4	<i>argue</i>	3
<i>enter</i>	4	<i>trust</i>	3
<i>suffer</i>	3	<i>transfer</i>	3
<i>face</i>	3	<i>die</i>	3
<i>struggle</i>	3	<i>warn</i>	3

¹⁶¹ Verbs that combine with at least three different prepositions to form prepositional verbs.

Appendix 7. Productivity (combination with X verb types) and frequency (raw) of prepositions

Preposition	Prepositional verb types	Overall frequency
<i>to</i>	95	493
<i>with</i>	68	251
<i>from</i>	46	249
<i>for</i>	45	263
<i>on</i>	34	166
<i>about</i>	27	231
<i>into</i>	24	55
<i>of</i>	23	96
<i>in</i>	17	70
<i>as</i>	15	119
<i>against</i>	8	33
<i>upon</i>	6	14
<i>through</i>	6	9
<i>at</i>	4	77
<i>over</i>	4	5
<i>after</i>	3	52
<i>around</i>	3	5
<i>towards</i>	3	3
<i>under</i>	2	10
<i>by</i>	2	3
<i>across</i>	1	8
<i>behind</i>	1	5
<i>beyond</i>	1	3
<i>above</i>	1	1
<i>among</i>	1	1
<i>forward</i>	1	1
<i>between</i>	1	1

Appendix 8. Frequency lists of verb-noun collocations in TICLE

Appendix 8a. Group I verb-noun collocations (verb + (determiner) + noun) occurring in the essays of 3 or more learners¹⁶²

Collocation	Number of occurrences
<i>acquire knowledge</i>	5
<i>affect (one's) life</i>	9
<i>affect (one's) health</i>	3
<i>affect sb.'s psychology</i> ¹⁶³	4
<i>affect (the) society</i>	3
<i>answer a question</i>	14
<i>apply a law</i>	3
<i>apply euthanasia</i>	5
<i>apply knowledge</i>	5
<i>approve euthanasia</i>	3
<i>ask a question</i>	27
<i>attend (*to) a class</i>	6
<i>attend (the) lessons</i>	3
<i>attend (*to) school</i>	4
<i>attend (a) university</i>	5
<i>attract one's attention</i>	3
<i>bring happiness</i>	6
<i>cause (*to) (sb.'s) death</i>	15
<i>cause divorce</i>	7
<i>cause harm</i>	3

¹⁶² Differences in determiners in verb-noun collocations are disregarded. In the case of compounds, only the last unit is taken into consideration. That is, for instance, the instances of *protect + the/human/woman + rights* are given under *protect (one's) rights*. Although the nouns occasionally occur with a plural marker in the data, singular and plural occurrences have been merged and all the instances are given in singular form in this table unless all instances of the noun are in plural or the use of plural marker is erroneous, i.e. additional.

¹⁶³ One learner spelled the noun erroneously as *pshology* twice.

Collocation	Number of occurrences
<i>cause a problem</i>	23
<i>cause suffer(ing)</i>	3
<i>change (sb. 's) mind</i>	5
<i>come home</i>	9
<i>commit (a) crime</i>	32
<i>commit (*a) suicide</i>	37
<i>consider a fact</i>	4
<i>create atmosphere</i>	5
<i>create opportunities</i>	3
<i>cure an illness</i>	8
<i>decrease the value (of)</i>	6
<i>defend an idea</i>	4
<i>defend sb. 's rights</i>	6
<i>develop a skill</i>	4
<i>dial the number</i>	3
<i>do abortion</i>	6
<i>do an experiment</i>	9
<i>do one's best</i>	16
<i>do (the) cooking</i>	3
<i>do a/one's duty</i>	4
<i>do euthanasia</i>	8
<i>do homework</i>	7
<i>do housework</i>	27
<i>do a/one's job</i>	37
<i>do work</i>	38
<i>draw sb. 's attention</i>	6
<i>drink alcohol</i>	4
<i>earn (sb. 's own) living</i>	4
<i>earn income</i>	3
<i>end a/one's pregnancy</i>	8
<i>end a marriage</i>	8
<i>end one's life</i>	36
<i>enter (one's) life</i>	5
<i>enter university</i>	6
<i>express one's feelings</i>	3
<i>express one's opinion</i>	3
<i>face (*to) difficulties</i>	10
<i>face (*to) a problem</i>	11
<i>feel pain</i>	8

Collocation	Number of occurrences
<i>feel the need</i>	4
<i>find a job</i>	57
<i>find an answer</i>	6
<i>find a cure</i>	7
<i>find information</i>	4
<i>find a method</i>	6
<i>find a reason</i>	5
<i>find a solution</i>	31
<i>find time</i>	3
<i>find a way</i>	11
<i>forbid abortion</i>	3
<i>gain (one's) rights</i>	3
<i>gain value</i>	3
<i>get a chance</i>	3
<i>get a degree</i>	11
<i>get a job</i>	13
<i>get benefit (from)</i>	4
<i>*get divorce</i>	8
<i>get education</i>	4
<i>get freedom</i>	4
<i>get help</i>	6
<i>get information</i>	11
<i>get knowledge</i>	9
<i>get (a) mark</i>	35
<i>get (one's) power</i>	5
<i>get one's right</i>	4
<i>get permission (from)</i>	3
<i>get (the) pleasure (of)</i>	3
<i>give an answer</i>	6
<i>give birth (to)</i>	20
<i>give (N) a chance</i>	25
<i>give a decision¹⁶⁴</i>	18
<i>give (an) education</i>	15
<i>give effort</i>	4
<i>give an example</i>	29

¹⁶⁴ In one instance the noun collocate was written erroneously as 'decesion' (TRKE2019).

Collocation	Number of occurrences
<i>give experience</i>	3
<i>give feedback</i>	3
<i>give (sb.) a feeling</i>	3
<i>give harm (to N) / give (sb.) harm</i>	29
<i>give importance</i>	46
<i>give information</i> ¹⁶⁵	15
<i>give knowledge(*s)</i>	20
<i>give (sb.) a lesson</i>	15
<i>give (a) life</i>	13
<i>give love</i>	5
<i>give (sb.) an opportunity</i>	19
<i>give pain</i>	6
<i>give permission</i>	11
<i>give power</i>	3
<i>give punishment</i>	5
<i>give responsibilities</i>	6
<i>give (sb.) (one's) right</i>	32
<i>give (sb.) a role</i>	3
<i>give salary</i>	3
<i>give time</i>	3
<i>*grow a child</i>	3
<i>have a (dis)advantage</i>	21
<i>have a (dis)agreement</i>	3
<i>have an (in)ability</i>	14
<i>have a career</i>	5
<i>have (an) abortion</i>	27
<i>have an accident</i>	3
<i>have an aim</i>	4
<i>have an alternative</i>	6
<i>have an argument</i>	5
<i>have benefits</i>	4
<i>have (the) capacity (of)</i>	4
<i>have a chance (of)</i>	46
<i>have a character</i>	3
<i>have a choice</i>	7

¹⁶⁵ In one instance the noun collocate was written erroneously as 'infomation' (TRCU1130).

Collocation	Number of occurrences
<i>have a condition</i>	4
<i>have (self-)confidence</i>	7
<i>have control (on/of)</i>	4
<i>have a course</i>	11
<i>have deficiency</i>	3
<i>have a degree</i>	21
<i>have differences</i>	5
<i>have (the) difficulty (in/of)</i>	31
<i>have a disease</i>	9
<i>have a duty (of)</i>	4
<i>have education</i>	16
<i>have an effect</i>¹⁶⁶	28
<i>have equality</i>	10
<i>have an exam</i>	3
<i>have experience(*s)</i>	17
<i>have a feature (of)</i>	18
<i>have a feeling (of)</i>	10
<i>have (the/one's) freedom (of)</i>	12
<i>have a function</i>	3
<i>have future</i>	4
<i>have a habit</i>	3
<i>have (no) hope</i>	4
<i>have an idea (of)</i>	8
<i>have illness</i>	5
<i>have an impact</i>	5
<i>have (an) importance</i>	12
<i>have an influence (on)</i>	5
<i>have information</i>	8
<i>have a job</i>	38
<i>have (*a) knowledge (about/of)</i>	28
<i>have law</i>	6
<i>have (a) lesson</i>	13
<i>have a life</i>	13
<i>have a look (at)</i>	4
<i>have a mark</i>	3

¹⁶⁶ In three instances the noun collocate was written erroneously as 'affect' by the learner 'TRCU1008'.

Collocation	Number of occurrences
<i>have a meaning</i>	5
<i>have an operation</i>	7
<i>have (an) opportunity</i>	33
<i>have pain</i>	12
<i>have a part</i>	4
<i>have penalty</i>	3
<i>have a period</i>	4
<i>have a place</i>	6
<i>have power</i>	23
<i>have priority</i>	3
<i>have a problem (with)</i>	72
<i>have quality</i>	5
<i>have a reason</i>	12
<i>have a relationship (with)</i>	9
<i>have respect (*in/*on)</i>	4
<i>have a responsibility</i>	21
<i>have (a) right</i>	172
<i>have a role</i>	20
<i>have rules</i>	5
<i>have situation</i>	4
<i>have skills</i>	4
<i>have a structure</i>	3
<i>have style (of)</i>	5
<i>have a system</i>	6
<i>have a talent</i>	4
<i>have a tendency</i>	4
<i>have a thought (about)</i>	5
<i>have time</i>	13
<i>have (a) trouble</i>	6
<i>have (a) value</i> ¹⁶⁷	28
<i>have (no) way (of)</i>	5
<i>hear news</i>	5
<i>improve one's skills</i>	6
<i>know the/one's answer</i>	4
<i>know the meaning of</i>	4

¹⁶⁷ In two instances the noun collocate was written erroneously as 'valve' by the learner TRME3024.

Collocation	Number of occurrences
<i>know the value of</i>	3
<i>learn information</i>	6
<i>learn knowledge(*s)</i>	7
<i>learn a language</i>	6
<i>learn a lesson</i>	6
<i>learn life</i>	4
<i>learn methods</i>	3
<i>learn a subject</i>	4
<i>leave school</i>	3
<i>live a/one's life</i>	12
<i>lose one's conscious(ness)</i>	6
<i>lose one's hope</i>	3
<i>lose importance</i>	3
<i>lose one's job</i>	3
<i>lose one's life</i>	8
<i>lose one's value</i>	6
<i>make a decision</i>	34
<i>make a mistake</i>	10
<i>make an operation</i>	3
<i>make a choice</i>	5
<i>make a contribution (to)</i>	3
<i>make an effort</i>	5
<i>make experiment</i>	3
<i>make invention</i>	6
<i>make laws</i>	5
<i>make a/one's living</i>	5
<i>make a plan</i>	3
<i>make practice/*practise</i>	6
<i>make projects</i>	3
<i>make research</i>	9
<i>meet the/one's need</i>	6
<i>meet a problem</i>	4
<i>memorize information</i>	5
<i>minimize opportunity</i>	3
<i>need help</i>	4
<i>obey a law</i>	5
<i>obey a rule</i>	12
<i>overcome a difficulty</i>	7
<i>overcome a problem</i>	6

Collocation	Number of occurrences
<i>pass a class</i>	5
<i>pass a law</i>	3
<i>pass an exam/examination</i>	31
<i>pay attention (to)</i>	25
<i>pay taxes</i>	5
<i>perform a test</i>	3
<i>play a game</i>	11
<i>play a role (in/*on/*to)</i>	28
<i>play football</i>	3
<i>practice/practise knowledge(*s)</i>	6
<i>present a lesson</i>	3
<i>prevent a problem</i>	5
<i>prevent death</i>	8
<i>prevent suffering</i>	4
<i>produce a solution</i>	3
<i>protect (one's) rights</i>	3
<i>provide a condition</i>	6
<i>provide easiness</i>	3
<i>provide (sb.) an environment</i>	3
<i>provide equality</i>	3
<i>provide (sb.) a life</i>	4
<i>provide opportunities</i>	7
<i>reach a conclusion</i>	4
<i>reach a goal</i>	3
<i>reach a level</i>	4
<i>reach a point</i>	3
<i>reach an/one's aim</i>	8
<i>reach (*a) knowledge</i>	3
<i>read news</i>	4
<i>repeat an experiment</i>	4
<i>require strength</i>	4
<i>save sb.'s life</i>	13
<i>save time</i>	6
<i>show (*a) respect (to/for)</i>	7
<i>solve a problem</i>	70
<i>spend (one's) time</i>	22
<i>spend a life</i>	5
<i>spend effort</i>	9
<i>spread an idea (to)</i>	5

Collocation	Number of occurrences
<i>study a lesson</i>	9
<i>suffer pain</i>	7
<i>supply needs</i>	3
<i>support an idea</i>	6
<i>support one's/sb.'s need</i>	3
<i>survive a life</i>	3
<i>take a course</i>	12
<i>take a decision</i>	16
<i>take a degree</i>	7
<i>take a lesson</i>	5
<i>take (a) mark</i>	9
<i>take education</i>	9
<i>take knowledge(*s)</i>	8
<i>take measures</i>	4
<i>take notes</i>	4
<i>take one's life</i>	19
<i>take one's place</i>	8
<i>take (*a) place(*s)</i>	24
<i>take pleasure (in)</i>	3
<i>take precaution (against)</i>	12
<i>take responsibility (of/for)</i>	11
<i>take revenge</i>	3
<i>take rights (from)</i>	9
<i>take side</i>	3
<i>take steps (towards)</i>	4
<i>take time</i>	9
<i>teach a language</i>	4
<i>teach a lesson</i>	12
<i>teach a subject</i>	6
<i>teach information</i>	7
<i>teach knowledge</i>	4
<i>transfer information</i>	6
<i>understand a lesson</i>	7
<i>understand a subject</i>	3
<i>use a method</i>	17
<i>use the/one's right (of)</i>	12
<i>use one's abilities</i>	4
<i>use information</i>	10
<i>use (one's) knowledge</i>	31

Collocation	Number of occurrences
<i>use means</i>	5
<i>use a technique</i>	3
<i>want death</i>	3
<i>want euthanasia</i>	9
<i>waste one's time (with)</i>	4
<i>watch (the) news</i>	7
<i>watch a film</i>	8
<i>*win exam/examination</i>	5

Appendix 8b. Group II verb-noun collocations (verb + noun + preposition)

Collocation	Number of occurrences
<i>become (a) part of</i>	7
<i>become a matter of</i>	2
<i>bring an end to</i>	1
<i>bring a use to</i>	1
<i>give an end to</i>	3
<i>give way to</i>	1
<i>have lack of</i>	5
<i>make a stand against</i>	2
<i>make contact with</i>	1
<i>make use of</i>	13
<i>put a distance between</i>	1
<i>put a stop to</i>	1
<i>put an end(*ing) to</i>	5
<i>put emphasis on</i>	1
<i>put the blame on</i>	1
<i>take (a) part in</i>	6
<i>take (the) place of</i>	3
<i>take advantage of</i>	3
<i>take care of (*about/*for)</i>	18

Appendix 8c. Group III verb-noun collocations (verb + prepositional phrase)

Collocation	Number of occurrences
<i>bring to one's attention</i>	1
<i>*bring N to lawsuit</i>	1
<i>come in/(in)to (the/one's) mind</i>	16
<i>come to the point / come *in a point</i>	2
<i>come to a conclusion</i>	4
<i>come to a solution</i>	1
<i>come into use</i>	1
<i>*come into daylight</i>	1
<i>come to earth</i>	1
<i>come (in)to (the) world</i>	2
<i>come to an agreement</i>	1
<i>come to life</i>	2
<i>*come to fore</i>	1
<i>come to an end</i>	3
<i>come *on the verge of N</i>	1
<i>come to the surface</i>	1
<i>fall in love with</i>	2
<i>give as an example</i>	4
<i>*go to great distance</i>	1
<i>go to (a) court</i>	2
<i>go for a walk</i>	1
<i>go on holiday</i>	1
<i>keep to a minimum (level)</i>	2
<i>keep in touch with</i>	2
<i>keep in mind</i>	4
<i>keep under control</i>	3
<i>keep in step with</i>	1
<i>learn by heart</i>	4
<i>put into trouble</i>	1
<i>put in(to) practice/*practical</i>	11
<i>put N in charge of</i>	1
<i>put N in danger</i>	1
<i>put N to death</i>	2
<i>put N to sleep</i>	1
<i>run in a marathon</i>	1
<i>sink into one's subconscious</i>	1

Collocation	Number of occurrences
<i>speak in a language</i>	1
<i>starve to death</i>	1
<i>stay in (one's) mind</i>	1
<i>*stay as a secret</i>	1
<i>stone to death</i>	4
<i>suffer to death</i>	1
<i>take into account</i>	1
<i>take under (*the) control</i>	2
<i>*take under guarantee</i>	1
<i>take under one's hegemony</i>	1
<i>take (in)to consideration</i>	14
<i>writhe in pain</i>	1

Appendix 9. Frequent noun collocates used in delexical verb-noun collocations in TICLE

Table A9.1. Noun collocates of *have* used by ≥ 3 learners and their raw frequencies

Noun	Frequency	Noun	Frequency
<i>right</i>	172	<i>trouble</i>	6
<i>problem</i>	72	<i>career</i>	5
<i>chance</i>	46	<i>argument</i>	5
<i>job</i>	38	<i>difference</i>	5
<i>opportunity</i>	33	<i>illness</i>	5
<i>difficulty</i>	31	<i>impact</i>	5
<i>effect</i>	28	<i>influence</i>	5
<i>knowledge</i>	28	<i>meaning</i>	5
<i>value</i>	28	<i>quality</i>	5
<i>abortion</i>	27	<i>rule</i>	5
<i>power</i>	23	<i>style</i>	5
<i>(dis)advantage</i>	21	<i>thought</i>	5
<i>degree</i>	21	<i>way</i>	5
<i>responsibility</i>	21	<i>look</i>	4
<i>role</i>	20	<i>aim</i>	4
<i>feature</i>	18	<i>benefit</i>	4
<i>experience(*s)</i>	17	<i>capacity</i>	4
<i>education</i>	16	<i>condition</i>	4
<i>(in)ability</i>	14	<i>control</i>	4
<i>lesson</i>	13	<i>deficiency</i>	4
<i>life</i>	13	<i>duty</i>	4
<i>time</i>	13	<i>future</i>	4
<i>freedom</i>	12	<i>hope</i>	4
<i>importance</i>	12	<i>part</i>	4
<i>pain</i>	12	<i>period</i>	4
<i>reason</i>	12	<i>respect</i>	4
<i>course</i>	11	<i>situation</i>	4
<i>feeling</i>	10	<i>skill</i>	4
<i>equality</i>	10	<i>talent</i>	4
<i>disease</i>	9	<i>tendency</i>	4
<i>relationship</i>	9	<i>(dis)agreement</i>	3
<i>idea</i>	8	<i>accident</i>	3
<i>information</i>	8	<i>character</i>	3

Noun	Frequency	Noun	Frequency
<i>choice</i>	7	<i>exam</i>	3
<i>(self-)confidence</i>	7	<i>function</i>	3
<i>operation</i>	7	<i>habit</i>	3
<i>alternative</i>	6	<i>mark</i>	3
<i>law</i>	6	<i>penalty</i>	3
<i>place</i>	6	<i>priority</i>	3
<i>system</i>	6	<i>structure</i>	3

Table A9.2. Noun collocates of *give* used by ≥ 3 learners and their raw frequencies

Noun	Frequency	Noun	Frequency
<i>importance</i>	46	<i>answer</i>	6
<i>example</i>	33	<i>pain</i>	6
<i>right</i>	32	<i>responsibility</i>	6
<i>harm</i>	29	<i>love</i>	5
<i>chance</i>	25	<i>punishment</i>	5
<i>birth</i>	20	<i>effort</i>	4
<i>knowledge(*s)</i>	20	<i>experience</i>	3
<i>opportunity</i>	19	<i>feedback</i>	3
<i>decision</i>	18	<i>feeling</i>	3
<i>education</i>	15	<i>power</i>	3
<i>information</i>	15	<i>role</i>	3
<i>lesson</i>	15	<i>salary</i>	3
<i>life</i>	13	<i>time</i>	3
<i>permission</i>	11	<i>end</i>	3

Table A9.3. Noun collocates of *take* used by ≥ 3 learners and their raw frequencies

Noun	Frequency	Noun	Frequency
<i>place</i>	35	<i>knowledge</i>	8
<i>life</i>	19	<i>degree</i>	7
<i>care</i>	18	<i>part</i>	6
<i>decision</i>	16	<i>lesson</i>	5
<i>consideration</i>	14	<i>measures</i>	4
<i>course</i>	12	<i>notes</i>	4
<i>precaution</i>	12	<i>step</i>	4
<i>responsibility</i>	11	<i>advantage</i>	3
<i>mark</i>	9	<i>pleasure</i>	3
<i>education</i>	9	<i>revenge</i>	3
<i>right</i>	9	<i>side</i>	3
<i>time</i>	9		

Table A9.4. Noun collocates of *make* used by ≥ 3 learners and their raw frequencies

Noun	Frequency	Noun	Frequency
<i>decision</i>	34	<i>laws</i>	5
<i>use</i>	13	<i>living</i>	5
<i>mistake</i>	10	<i>contribution</i>	3
<i>research</i>	9	<i>experiment</i>	3
<i>invention</i>	6	<i>operation</i>	3
<i>practice/*practise</i>	6	<i>plan</i>	3
<i>choice</i>	5	<i>project</i>	3
<i>effort</i>	5		

Table A9.5. Noun collocates of *get* used by ≥ 3 learners and their raw frequencies

Noun	Frequency	Noun	Frequency
<i>mark</i>	35	<i>benefit</i>	4
<i>job</i>	13	<i>education</i>	4
<i>degree</i>	11	<i>freedom</i>	4
<i>information</i>	11	<i>one's right</i>	4
<i>knowledge</i>	9	<i>pleasure</i>	3
<i>divorce</i>	8	<i>chance</i>	3
<i>help</i>	6	<i>permission</i>	3
<i>power</i>	5		

Table A9.6. Noun collocates of *do* used by ≥ 3 learners and their raw frequencies

Noun	Frequency	Noun	Frequency
<i>work</i>	38	<i>euthanasia</i>	8
<i>job</i>	35	<i>homework</i>	7
<i>housework</i>	27	<i>abortion</i>	6
<i>one's best</i>	16	<i>duty</i>	4
<i>experiment</i>	9	<i>cooking</i>	3

Zusammenfassung in der deutscher Sprache

Die vorliegende Arbeit beschäftigt sich mit sprachlichen Abweichungen türkischsprachiger Englischlerner/innen bei den im Englischen häufigen Mehrwortverben. Diese Verben haben sich wegen ihrer strukturellen, semantischen und kontrastiven Komplexität sowohl beim Erwerb als auch beim aktiven Gebrauch als problematisch erwiesen. Es wird also untersucht, wie türkische Englischlerner/innen, deren Sprachniveau im mittleren bis fortgeschrittenen Bereich liegt, Mehrwortverben in ihrer schriftlichen Sprachproduktion verwenden und inwieweit sie dabei von der Zielnorm abweichen, inwiefern sich ihr Gebrauch also von dem von Muttersprachler/innen unterscheidet. Auch die Frage, wie die Lerner/innen ihre mangelnden Kenntnisse kompensieren und inwieweit ihre Muttersprache den Gebrauch englischer Mehrwortverben beeinflusst, wird analysiert. Die Basis der Untersuchung bildet das türkische Subkorpus (TICLE) des *International Corpus of Learner English* (ICLE). Die Lernerdaten wurden sowohl in quantitativer als auch in qualitativer Hinsicht untersucht, wobei vier Kategorien von Mehrwortverben in den Blick genommen wurden, Phrasalverben wie *find out* oder *bring about*, Präpositionalverben wie *refer to* oder *rely on*, Phrasal-Präpositionalverben wie *come up with* oder *look forward to* und Verb-Nomen-Kollokationen wie *solve a problem*, *give way to* oder *take into consideration*. Aus dem untersuchten Lernerkorpus wurden insgesamt

6.129 Token extrahiert. Für phrasale und phrasal-präpositionale Verben, von denen bekannt ist, dass sie von einigen Lernergruppen aufgrund eines L1-Einflusses vermieden oder nur selten verwendet werden, wird das *Louvain Corpus of Native English Essays* (LOCNESS) als Kontrollkorpus herangezogen. Damit können die Gemeinsamkeiten und Unterschiede bei der Verwendung dieser Verbkategorien durch türkischsprachige Englischler/innen und englische Muttersprachler/innen identifiziert werden.

Nach dieser allgemeinen Übersicht werden im Folgenden die Hauptergebnisse der Studie präsentiert. Ein erstes wichtiges Ergebnis ist, dass die Verwendung englischer Mehrwortverben ein generelles Problem für die untersuchte Lernergruppe darstellt, unabhängig davon, ob eine ähnliche Mehrwortverbstruktur in der L1 der Lerner/innen existiert oder nicht. Dabei konnte eine Fülle eindeutig von der muttersprachlichen Norm abweichender Verwendungen von Mehrwortverben (für alle untersuchten Kategorien) festgestellt werden. Diese Abweichungen zusammen mit vielen unkonventionellen, also nicht idiomatischen Verwendungen führen zu einem generellen Mangel an Idiomatizität in der schriftlichen Produktion der Lerner/innen. Nennenswerte Abweichungen traten dabei in allen Komponenten der Mehrwortverben auf, also bei den Verben selbst, aber auch bei Substantivkollokaten, Präpositionen, Partikeln und Determinatoren. Daneben wurde eine beträchtliche Zahl der vorhandenen Mehrwortverben in Kontexten verwendet, in denen sie nicht mit den in Wörterbüchern gefundenen Bedeutungen übereinstimmen. Dies betrifft

sowohl die unangemessene Wahl eines Verbs oder eines Substantivkollokats (auch dessen Numerus), aber auch Abweichungen bei Präpositionen, Partikeln (Hinzufügung, Auslassung oder Ersetzung) und Determinatoren (Hinzufügung oder Auslassung). Einige dieser Abweichungen von den muttersprachlichen Normen lassen sich direkt auf eine L1-Interferenz zurückführen, das heißt, die Lerner/innen übertrugen ihre allgemeinen L1-Muster bei der Verwendung von Mehrwortverben in ihre L2 bzw. übersetzten Mehrwortverben direkt Wort für Wort. Dies führt meist zu zusätzlichen Präpositionen oder zu Substitutionsfehlern. Insgesamt ist in den Daten deutlich zu erkennen, dass die Lerner/innen bei der Verwendung von Mehrwortverben meist nach Sinclairs (1987, 1991) “*open-choice-Prinzip*” vorgehen, das besagt, dass die Lerner/innen Einheiten kombinieren, indem sie zwar den grammatikalischen Regeln der Zielsprache folgen, die Idiomatizität aber außer Acht lassen.

In Bezug auf phrasale Verben bestätigen die Ergebnisse dieser Studie die bisherige Literatur. Diese Verbkategorie wird deutlich seltener verwendet als bei Muttersprachler/innen. Stattdessen wählen die Lerner/innen in ihrer schriftlichen Produktion oftmals das semantisch ähnliche oder äquivalente Ein-Wort-Verb, also zum Beispiel *continue* statt *go on* oder *keep on*. Außerdem verwenden die Lerner/innen diese Verbkategorie semantisch eingeschränkter, das heißt, sie wenden sie nicht in einer so großen Bandbreite von Bedeutungen an wie Muttersprachler/innen.

Die durchgeführten Regressionsanalysen, die sowohl den Einfluss lernerbezogener Variablen wie *Geschlecht*, *Alter*, *Anzahl der Jahre Englisch in der Schule/Universität* oder *andere Sprachkenntnisse* als auch externer Variablen wie *Verwendung von Referenzwerkzeugen*, *Textlänge* oder *Essaythemen* auf die Verwendung von phrasalen und phrasal-präpositionalen Verben durch türkische Lerner/innen untersuchen, erklären nur 9 Prozent der Gesamtvariation. Dies zeigt, dass noch viele andere Faktoren im Spiel sein müssen, die entweder individueller Natur sind oder hier nicht identifiziert werden konnten. In der vorliegenden Studie erwies sich nur der Einfluss der abhängigen Variable *Textlänge* als signifikant. Sie zeigt eine positive Korrelation mit der Gebrauchshäufigkeit von phrasalen und phrasal-präpositionalen Verben, das heißt, wer längere Texte verfasst, gebraucht absolut und relativ mehr Mehrwortverben als jemand, der kürzere Texte verfasst.

Bei allen vier untersuchten Mehrwortverbkategorien zeigen die Lerner/innen eine starke Fokussierung auf eine begrenzte Anzahl von Verbtypen, meist solche, die ein hochfrequentes Verb enthalten. Dies deutet darauf hin, dass es den Lerner/innen an Vielfalt in ihrem Mehrwortverbrepertoire mangelt. Es finden sich also nur sehr wenige Verbtypen in TICLE, die dann aber besonders häufig und produktiv verwendet werden. Einige Mehrwortverben wie *grow up* oder *commit suicide*, die durch ein bestimmtes Essaythema getriggert wurden, wurden zum Beispiel von vielen Lernenden immer wieder verwendet. Daneben kommen Token wie *wake up* oder *leak out* auch in den von

TICLE vorgegebenen formellen Kontexten vor, in denen Muttersprachler/innen sie eher nicht verwenden würde.

Zusammenfassend lässt sich feststellen, dass der Gebrauch von englischen Mehrwortverben die untersuchten türkischsprachigen Lerner/innen vor große Probleme stellt, unabhängig davon, ob eine ähnliche Mehrwortverbstruktur im Türkischen existiert oder nicht. Den Lerner/innen fehlt es, selbst in Bezug auf häufige Mehrwortverben, an Wissen über Kollokationsmöglichkeiten und Gebrauchsrestriktionen. Die vorliegende Studie ist die erste groß angelegte korpusbasierte Studie, die zur *Interlanguage performance* von türkischen Englischlerner/innen in Bezug auf den Mehrwortgebrauch im Englischen durchgeführt wurde. Die Ergebnisse, die auch im Hinblick auf ein klareres Verständnis des Lernprozesses bei Mehrwortverben in der L2 und auf pädagogische Implikationen hin diskutiert werden, stimmen gut mit früheren Untersuchungen überein, die den Gebrauch von Mehrwortverben in der *Interlanguage* von Lernergruppen mit verschiedenen L1-Sprachen untersucht haben.

Die vorgelegte Arbeit umfasst sieben Kapitel. Nach einem Einführungskapitel beschäftigt sich das zweite Kapitel zuerst mit dem Lexikon im Allgemeinen, wobei die Rolle des Lexikons in der Linguistik und im Fremdsprachenunterricht behandelt wird, und dann mit Mehrworteinheiten im Besonderen. Kapitel zwei gibt auch einen Überblick über die Rolle der Mehrworteinheiten beim Sprachenlernen und beim Fremdsprachengebrauch und erörtert die vermuteten

Schwierigkeiten sowie die Bedeutung dieser Einheiten für Fremd- und Zweitsprachenlerner/innen. Ein Literaturüberblick über lernerbezogene Forschungen in diesem Bereich rundet dieses Kapitel ab. Die Probleme, die englische Mehrwortverben für Fremdsprachenlerner/innen im Allgemeinen und für türkische Lerner/innen im Besonderen darstellen, werden dabei kontrastiv angesprochen. Es ist dieser Kontrast, der die Wahl von Mehrwortverben als Untersuchungsgegenstand rechtfertigt. Das dritte Kapitel beschäftigt sich näher mit dem Begriff der Mehrwortverben und stellt die verschiedenen in der Studie untersuchten Kategorien von Mehrwortverben vor. Kapitel vier ist den methodischen Aspekten der vorliegenden Studie gewidmet. Neben der für die Analyse relevanten Software für Retrieval und Analyse, werden die Referenzwerkzeuge (Wörterbücher und Korpora, die zur Überprüfung der Angemessenheit der attestierten Mehrwortverben herangezogen wurden) und die untersuchten Korpora detailliert vorgestellt. Das Kapitel endet mit einer kurzen Besprechung zentraler Begriffe wie Unterverwendung, Überverwendung, Fehler und Muttersprachler. In den Kapiteln 5 und 6 werden die Ergebnisse der quantitativen und qualitativen Analysen vorgestellt. Das letzte Kapitel fasst dann die Ergebnisse zusammen und stellt ihre Implikationen für den lexikalischen Lernprozess in der L2 im Allgemeinen und in Bezug auf den Wortschatzunterricht im Besonderen dar. Dabei werden auch die vorhandenen Beschränkungen der vorliegenden Studie benannt. Abschließend bietet dieses letzte Kapitel einen Ausblick auf mögliche weiterführende Forschungen.

In the context of foreign and second language learning, multi-word units constitute a particularly interesting phenomenon since they are known to cause problems for learners. One group of multi-word units that causes great difficulty even for advanced learners of English is multi-word verbs. Their acquisition and active usage is a challenge since they are complex both in terms of their grammatical form and their lexical meaning. This study provides a detailed, descriptive investigation of four different categories of multi-word verbs – namely phrasal, phrasal-prepositional, prepositional verbs and verb-noun collocations – in the essays written by intermediate to advanced level of Turkish learners of English.

The Turkish sub-corpus (TICLE) of the International Corpus of Learner English (ICLE) has been the basis for the investigation. In order to thoroughly capture difficulties the learners experience in the use multi-word verbs and gain a better understanding of their phraseological competence, both the qualitative and quantitative aspects of the learners' performance are investigated. An important aim of the study is to determine whether, and if so, to what extent, the learner's first language (L1) influences their use of multi-word verbs in English. In addition to the learner's L1, possible effects of other factors (both learner-related and external variables) are also investigated in the context of two categories of multi-word verbs, namely phrasal and phrasal-prepositional verbs – the two verb categories reported to be avoided and/or underused by many learner groups.

Die Publikationsreihe NIHIN – New Ideas in Human Interaction – entstand 2010 und ist ein Kooperationsprojekt zwischen der Hermann Paul School of Linguistics (HPSL) und der Universitätsbibliothek Freiburg (UB).

NIHIN bietet eine moderne, frei zugängliche Plattform für wissenschaftliche Essays erfahrener WissenschaftlerInnen sowie Prädikatsdissertationen, Textsammlungen zum Thema Sprache in der Interaktion und multimodale Sprachkorpora.



Universität
Basel

UNI
FREIBURG