Perception of Learner Errors and Non-standard Features in the Native and Non-native Language: Evaluation vs. Processing Cost

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

vorgelegt von
Natalia Ilin
aus Kursk, Russland

SS 2017

Erstgutachter: Prof. Dr. Dr. h.c. Bernd Kortmann

Zweitgutachter: Prof. Dr. Lars Konieczny

Drittgutachterin: Prof. Dr. Brigitte Halford

Vorsitzender des Promotionsausschusses
der Gemeinsamen Kommission
der Philologischen und
der Philosophischen Fakultät: Prof. Dr. Joachim Grage

Datum der Disputation: 26.01.2018

Table of Contents

Acknowledgements

The present research was funded by the Hermann Paul School of Linguistics Basel-Freiburg (HPSL) to which I am deeply grateful. I am highly indebted to Stefan Pfänder for his generous support.

I want to thank my supervisor in English linguistics, Bernd Kortmann, for his incredible heart and commitment to doctoral students. He is the only person I know who is dedicating so much time and care to discussing the topic during his office hours and colloquia, to reading and correcting our drafts, giving his feedback and encouragement, sharing his suggestions and explanations, supporting us in every way in and outside the university. I believe that such investment in students is the key to the high quality of our theses. I admire not only his high professional expertise but also patience and understanding.

Special thanks to my supervisor in psycholinguistics, Lars Konieczny, for sitting long hours with me, explaining statistics, helping with the R codes, answering my numerous questions and for just being so kind to me! I appreciate that he allowed me to use his eye-tracking lab and advised me in all details from experimental design to data analysis and the interpretation of the results. I am deeply thankful to the former team of the Center for Cognitive Science, Daniel Müller-Feldmeth and, particularly(!), Sascha Wolfer. I have greatly benefitted from their assistance.

I would also like to acknowledge the help with the visualization of my statistical models provided by Christoph Wolk, good advice on Construction Grammar and usage-based theories by Alice Blumenthal-Dramé as well as practical guidance in the multitude of non-standard features in English by my friend Kerstin Lunkenheimer.

The present thesis has been constantly improved thanks to the insightful comments and valuable suggestions of the members of the English Department and the Center for Cognitive Science. My traveling to the international conferences was financially supported by the International Graduate Academy of the Freiburg Research Services, and my web-based surveys were distributed via e-mail by the International Office. I am thankful to all who helped me find the participants, especially native speakers of English, which has been a challenge. I owe my deepest gratitude to the University of Freiburg that in the last dozen of years has become my second home.

Freiburg, March 2018

1. Introduction

First, and most obvious, errors stand out.
They attract our attention like brightly
colored flags, flapping above the parapet.
(Saxton 2011: 154)

1.1 Research questions and methods

What makes some written errors irritating, while others are quite acceptable or even go completely unnoticed? Deviations from the norm were closely studied within Error Analysis, a field inspired by the theoretical inquiries about the causes of errors and practical needs to improve language teaching (see an overview in Chapter 2). To make cross-cultural communication more successful, researchers in Second Language Acquisition conducted numerous questionnaire-based evaluation studies. They investigated the assessments of errors by native and non-native speakers, by people of different professions, especially teachers, and explored other sociolinguistic factors. The evaluation studies elicited the informants' judgments about the well-formedness, acceptability, intelligibility or 'foreignness' of erroneous utterances, and the gradable effect of errors – their gravity.

Sociolinguistic methods of collecting data have failed to identify the cognitive mechanisms responsible for error assessment. The field of psycholinguistics using real-time research methods could lend insight into these processes (see Chapter 3). Psycholinguists have developed models of reading and visual word recognition (discriminating between words and artificially designed non-words) in isolation, in strings, and in context. They analysed letter transpositions and the parsing of syntactic challenges (e.g. 'garden paths', local coherences, anaphora resolutions), but not exactly how readers process 'normal' learner errors in a text. Psycho- and neurolinguistics have been engaged in testing the hypotheses to support or falsify theories of mental processes. And Error Analysis, by contrast, has been criticized for being just "a methodology for dealing with data, rather than a theory of acquisition" (Cook 1993: 22). No one could deny that, and as James (1998) responds, "In fact, while some people want theory of Second Language Acquisition (SLA), I am convinced that many others want methodology for dealing with data." The present thesis is more of a "dealing with data" kind, too, and can be defined as a study of the cognitive foundations underlying error perception with the purpose to find out how the reaction to errors is formed.

Error evaluation studies produced hierarchies of error types, but it was not always clear why the order of the categories was so various. Some error gravity scales placed vocabulary errors as most serious, while others found them least disturbing. The names of the error types taken from pedagogical contexts appear to mask the group of factors that, in fact, determine the evaluators' perception of the concrete sentences used in the questionnaires. In my thesis, I attempt to single out these factors and investigate their effects while controlling for other possible influences. Thus, the main focus is on the processing of errors, that is, the time it takes the reader to recognize, repair, and integrate the meaning.

Figure 1.1 below illustrates the process of error perception with examples of underlying factors. The object of my study does not include the production of errors and the influence of the writers' characteristics on the errors they make. For this reason, the writer's part is separated by a dotted line, but it cannot be excluded from the picture because the reader keeps in mind who the writer is. There is evidence that, while processing the text, the reader reflects on the writer and updates the writer's characteristics (the bottom of Figure 1.1). When they are unknown, the reader makes inferences.

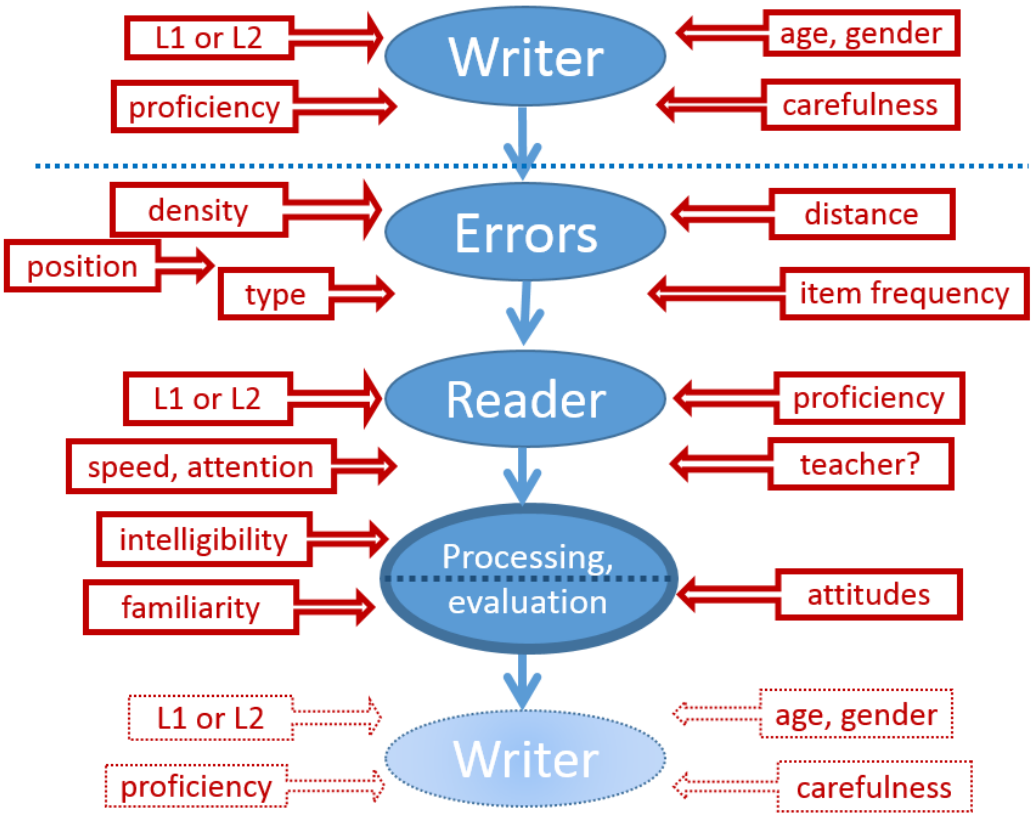


Figure 1.1. *Factors involved in error perception*

Errors in the text have their properties, too. They can be of a physical type, such as density (or error rate), position in the line and in the word, the length

of the text and the word affected by the error, the distance between the mismatching elements, etc. Alternatively, errors can be characterized in terms of quality: the frequency of the item, the linguistic category, the complexity of the construction affected by the error (explored in Chapters 4 and 5).

There are also variables unique to the reader, which may determine his/her reaction to errors: whether the reader is a native speaker or a foreign learner (at what level of proficiency), whether s/he has a special training in linguistics or language teaching, whether s/he is a fast or slow, careless or attentive reader, etc. Besides that, the deviation from the norm can be familiar or unfamiliar to the reader, and the text may have a different degree of comprehensibility. In addition, the perception of the language material may be influenced by attitudes in society, the educational system, and the degree to which the reader adheres to them.

The word "perception" in the title of this thesis encompasses both the cognitive mechanisms of error processing and the reader's impression of error gravity. The main goal of my research project is to find out the correlation between the error processing cost and error evaluation. My starting hypothesis is that:

THE MORE PROCESSING TIME THE ERROR REQUIRES FROM THE READER,
THE GREATER IS THE ERROR GRAVITY.

As previously mentioned, error gravity is a product of conscious ratings in a questionnaire, with an explicit instruction indicating the presence of errors to the participant. During Error Analysis, insights were gleaned that, underneath the attitudinal reactions to errors, there is mere processing difficulty causing irritation possibly due to low comprehensibility. By measuring the processing time for errors without pointing to them and reducing the social background (i.e. who made an error in which context) in a laboratory setting, it is possible to find out the irritation potential, i.e. the processing cost, of the error. Comparing these values to the results of a judgment study (with an explicit social background) for the same errors will allow us to calculate to what extent error gravity depends on the processing difficulty (Chapter 6).

Off-line methods, e.g. a questionnaire, measure how participants interpret a sentence after they have heard or read it completely. Participants can take time to think about the meaning of the sentence, use their explicit knowledge and metalinguistic abilities to make a conscious and controlled decision about its grammaticality (cf. Marinis 2013). On-line methods, by contrast, measure participants' unconscious and automatic responses to language stimuli as they unfold. When the task does not explicitly require acceptability judgment, there is no need to make use of metalinguistic abilities.

"Participants cannot use their explicit knowledge about language because they do not have the time to think about the meaning or grammaticality of the sentence" (op. cit.: 5).

The method used in my thesis for eliciting the processing cost of errors is eye-tracking – measuring the duration of fixations on words in their order. The relevance of this method is derived from the fact that "eye movement data reflect moment-to-moment cognitive processes" (Rayner 1998). Furthermore, eye-tracking has been widely used to study language processing in general as well as the perception of linguistic units, in particular.

There is a trade-off between a clean experimental setting and the contextual validity. Socially oriented researchers criticize psycholinguists for constructing unnatural stimuli, presenting them in isolated sentences, and that, even with sophisticated statistical methods, testing the processing of utterances out of context leads to questionable results. In their turn, cognitive scientists believe that purely qualitative and explorative studies are not formal enough and do not constitute grounds for generalizations. The fact is that naturalistic material is very heterogeneous in its nature. If we use unedited samples of natural language with their contexts to test one variable they all share, it is difficult to exclude the influence of other factors contained in the samples. The effects of our variable may be consumed by the noise of variation. For this reason, cutting out most of the context is an inevitable necessity. In my experiment, the processing times, as well as error ratings, were measured at sentence level. Although, in learner language, an error does not come alone but is often accompanied by others, I used only one error per test item. To maintain a natural flavor, my stimuli are corpus-based, feasible, but altered enough in order to control the variables.

1.2 A sociolinguistic approach to errors

When a foreign language is learned in the classroom, it is often presented as a set of words along with a set of rules for combining the words. If a student succeeds in applying the rules, the output is 'correct'; if the student fails, it is deemed 'wrong'. However, outside the classroom, language use is not just 'black' and 'white'; rather it contains many shades of grey. Impeccably correct language performance is restricted only to particular social areas, such as education, jurisdiction, administration, and the media. It is interesting that the same people who adhere to the standard language at work switch to a non-standard sociolect in informal situations. From the point of view of prescriptive grammar, all non-standard uses are considered 'errors', but from a descriptive

perspective, they are seen as alternatives bearing identity signals. If ungrammatical uses occur with some degree of regularity and systematicity, they are called "features", as in the case of English as a second language (ESL) in indigenized L2 varieties, also called "New Englishes" (Kachru 1985), discussed in section 5.5.1.

A different type of deviations from Standard English is found in countries where English is taught as a foreign language (EFL). The systematically occurring characteristics of interlanguage (Selinker 1972) are called "learner features", or simply "learner errors".

It is important to keep in mind that the notion of error is relative and indicates deviations from the standardized written L1 (first, or native, language) variety. "For example, "he *don't* eat meat" is not an error in English except in the sense of divergence from standard form" (Rifkin and Roberts 1995: 523). Designating an instance of non-standard use an 'error' or a 'feature' is often a question of politics (cf. Gonzalez 1983 "When does an error become a feature of Philippine English?"). The arbitrariness of such a distinction is similar to that between a dialect and a language. Following the playful, yet witty, definitions, "a language is a dialect with an army and navy" and "a language is a dialect with a dictionary, a grammar and a New Testament", a feature is an error used with self-confidence and an identity signal, and an L2 feature is an error that made its way into a dictionary, a grammar book and literature of one of the New Englishes. The transition of the status from an error to an officially recognized norm reflects the stage of "endonormative stabilization" of a language community (Stage 4 according to Schneider 2003) or "institutionalization" as described by Mollin (2007). "It is important that the new language variety, as a carrier of a new regional identity, has lost its former stigma and is positively evaluated" (Schneider 2003: 251). For that reason, it is not politically correct to call features "errors", thus implying that there is only one particular variety of English in the world that is correct. In the present study, all deviations from Standard written English, including L1 non-standard uses and ESL features, will be called "errors" only for convenience.

The body of possible errors in English can be imagined as a continuum, ranging from non-standard features common in English as a native language to learner errors in English as a foreign language, where the boundaries separating L1, ESL, and EFL features are fuzzy. I suppose that L1 features reflect social constraints, EFL learner features are determined by cognitive constraints, and L2 features are located in between. In my sample of 281 sentences with errors (discussed in Chapter 5), I could not find items to be classified as features characteristic exclusively of ESL speakers. For example, the leveling of the differences between the Simple Past and the Present Perfect, e.g. "I've done it

yesterday" or "I did it lately," is both an attested non-standard feature as well as a typical learner error (the perception of it is explored in Chapter 4).

There are two paradigms in the terminology for language speakers which will be followed in this thesis: (1) L1 vs. L2, or (2) L1 vs. ESL vs. EFL. The native language that is acquired first is always an L1. In the SLA tradition, L2 refers to the language learned AFTER the first language has been internalized as a system. In this study, L2 means a foreign language (as e.g. English in Germany or Russia). However, in particular sections where the use of English across varieties is important, the distinction between L1, ESL and EFL will be used to emphasize that, in ESL communities, English is practiced on a daily basis and has more functions than for EFL speakers.

1.3 A cognitive approach to errors

The concept of language and grammar adopted in this study shares the ideas of functionalism, cognitive linguistics and the usage-based model. Unlike the formalist approach and generative grammar, functionalist theories of grammar hold that structural aspects of language are motivated and constrained by functional concerns: "the forms of natural languages are created, governed, constrained, acquired and used in the service of communicative functions" (MacWhinney, Bates and Kliegl 1984: 128). In other words, grammar is shaped by language use. The functionalist approach views language as a dynamic system determined by communication and discourse as well as human cognitive capacities.

Outside Error Analysis and areas affiliated with language teaching, the word "error" is hardly used at all. Even modern SLA sources try not to sound prescriptive or judgmental and thus avoid using "errors", shifting to milder terms such as, for example, "non-English-like patterns" (Gass and Selinker 2008: 42) or "learners' idiosyncratic forms" (R. Ellis 2008: 63). Instead of the term "interlanguages", which suggests an intermediate stage between two 'real' languages, modern researchers speak of "learner varieties" and "try to describe their internal structure in positive terms" (Dimroth 2013).

Cognitive linguistics does not study errors or the perception of errors. Deviations from the norm are discussed not in terms of 'correct' and 'wrong', but 'favored' or 'disfavored'. Some questionable uses are simply acknowledged as 'low-frequent' and, consequently, 'not preferred'. Such an approach has also been adopted in SLA. For instance, R. Ellis (2008: 49) ponders whether or not some "infelicitous uses" can be considered erroneous. On the one hand, the forms are grammatical, but on the other hand,

> "this may not be the form *preferred* by native speakers of the code. […] The probabilistic nature of the use of forms can only be considered if the analyst has access to a corpus of native-speaker language which can serve as a baseline for comparing learners' choice of forms" (ibid.).

Ellis gives the following sentence as an example:

(1.1) One day an Indian gentleman, a snake charmer, arrived in England. He was coming from Bombay.

The preferred form is likely to be *had come*, although *was coming* could also be possible if one wished to emphasize duration. Theoretically, we can find out in a native speaker corpus the proportion of *had come* to *was coming* in similar contexts. The questions relevant to my study are: (1) what is the threshold, i.e. the percentage of the disfavored uses, that would allow teachers to mark it as an error, and (2) are the low-frequent alternatives, in fact, processed significantly slower? In Chapter 4, I will discuss the processing times for matching and mismatching adverbials with verbs in the Simple Past and the Present Perfect investigated for frequency in a corpus.

Construction Grammar (CxG) is a theory that best reflects the modern understanding of language in the mind and in use. This framework does not look at errors as such, yet one can gain an impression of its approach to errors. I will introduce CxG in more detail now because, throughout the thesis, many of my comments are made through its perspective. This approach argues that the grammar of a language is made up of a network of constructions (form-function pairings) instead of "meaningful words und meaningless rules" (Tomasello 2006). As a theory of linguistic knowledge, CxG denies the modular approach, that is, the separation between the mental lexicon and a store for morphosyntactic rules, the so-called "dictionary-and-grammar model". The reason for this is the multitude of 'exceptions' that do not fit in the modules.

> "[O]rdinary language is fully permeated by a large number of idiomatic expressions whose forms and meanings are not entirely predictable on the basis of either the word meanings recorded in a dictionary or the rules of syntax provided by a grammar" (Hilpert 2014: 4).

Instead, there is a single network of form-meaning pairs called "construct-i-con" in which a construction is a language unit at all levels of description. Common grammatical categories such as words, word classes and phrase structure rules are viewed as generalisations at a high level of abstraction, whereas CxG stresses "the importance of low-level generalisations in the representation of linguistic knowledge" (op. cit.: 70).

Those who subscribe to CxG make subtle distinctions between similar structures and look for motivations behind construction alternatives. They believe that constructions are learned on the basis of the input and, though they vary cross-linguistically, they are common due to cognitive constraints and their functions (cf. Goldberg 2011). Most recent studies describe children acquiring language in constructions, in a piecemeal fashion, based on the input from interactions with other people. "A construction is a generalization that speakers make across a number of encounters with linguistic forms" (Hilpert 2014: 9). This observation has brought a dramatic shift towards the usage-based model (e.g. Bybee 2006, 2010) where frequency plays a key role (see N. Ellis 2002 for review). "[P]erceptual salience and type frequency are among the factors that have a significant effect on constructional acquisition" (Boyd and Goldberg 2009: 419).

A construction can be described as a unit with its own specifications for syntax, semantics and pragmatics. For the combination of constructions to be accepted by the language, their specifications should match, i.e. undergo unification (Fried and Östman 2004).

> [P]ieces of linguistic material that do not match ('unify') along any number and types of properties (syntactic, semantic, pragmatic) will not be licensed as possible constructs of a given language (op.cit.: 25).
>
> Unification can take place only on condition that the relevant pieces of information do not conflict. This means that two values either have to match exactly or at least one must be unspecified (op. cit.: 38).

Thus, in terms of CxG, an error is a **unification failure** involving two constructions, and not a violation of one construction. It can be that (1) two adjacent constructions do not fit together, (2) a smaller construction does not fit into a larger construction, (3) a construction is missing (deletion), or (4) an extra construction is added (insertion). As evidence that there are two constructions involved, it is possible to correct an error in two ways, by changing either the first element or the second. For example, the sentence *I am meaning this book and not that one* can be made less awkward by either removing the progressive (*I mean this book and not that one*) or by changing the verb (*I am thinking of this book and not that one*). In a situation of natural communication, it is the context that helps readers identify which element should be altered in order to repair the error.

As already mentioned, sometimes it is difficult to determine whether a deviation from Standard English should be categorized as an error or not. It may look or sound unusual, but at the same time create no difficulty for understanding, as illustrated in the following examples:

(1.2) *a. The Prague I remembered…*
b. There are really two Californias.
c. Bring us three coffees.

Fried and Östman (2004) explain such uses as "shifts" and "examples of stretching the 'blueprint' by failing some of its specifications" (op. cit.: 39). In these cases, "two words are apparently allowed to combine into a phrase even though their inherent specifications are in conflict – either with each other or with the constructions they occur with" (op. cit.: 38).

Such deviations from Standard English meet the economy principle and offer an advantage by producing more compact sentences. Instead of saying *the city of Prague (back then)*, one simply says *the Prague*, instead of *two parts/sides of California,* simply *two Californias*, and instead of *three cups of coffee,* just *three coffees*. Similar examples are discussed by Hilpert (2014: 17) as effects of coercion:

(1.3) *a. Three beers please!*
b. John sauced the pizza.
c. Frank played the piano to pieces.

He states that:

> …constructions may override word meanings, creating non-compositional constructional meanings in the process. The 'morphosyntactic context', that is, the construction in which a lexical item is found, thus has the power to change or suppress certain semantic characteristics of that lexical item. When word meanings can be observed to change within a constructional context, we speak of **coercion effects** (ibid.).

As a mass noun, *beer* should not be used in the plural, *sauce* is not a verb but a noun, and you cannot literally *play something to pieces*, but the constructional context makes the reader trade off some properties of the words and accept them in new functions. Some of these constructions appear to be quite productive: one can not only *sauce the pizza*, but also *pepper the steak* or *butter a toast*.

A syntactic amalgam is another example of the non-standard use of constructions that is very close to be taken for a syntactic error from a prescriptive point of view. In this case, "two constructions are mutually interwoven" (Hilpert 2014: 64), for instance:

(1.4) *a. It was an important enough song to put on their last single.*
b. It's unbelievable the things he can do with the piano!

In the first example, the sources are the clause *It was an important song* and the instantiation of the *ENOUGH To*-INFINITIVE construction *It was important*

enough to put on their last single, and the amalgam sentence is linked to both of them via subpart links.

It is interesting why some constructions undergo the 'shift', or 'coercion', more readily than others and whether economy is the only motivation behind these changes. Native speakers (and near-native learners) must have a special feel because their creative uses are justified and perceived as acceptable, unlike those of beginning learners. The conclusion is that, in the constructionist approach, an error is a conflict between two constructions in which their specifications of use are violated. However, sometimes individual specifications can be violated, which results in utterances with questionable acceptability.

1.4 The structure of the thesis and its application

Following this introduction, Chapter 2 presents an overview of error studies in applied linguistics and, particularly, surveys on error evaluation. It is described how the irritation caused by errors is perceived by different social groups. Chapter 3 gives a psycholinguistic account of processing in L1 and L2 with a special attention to eye movement research. The purpose is to apply the psycholinguistic findings to error processing and to discover the factors involved in the processing of grammatical violations.

From Chapter 4 onward, I present the results of my eye-tracking experiment. It focuses on the effects of distance and frequency on the processing of one particular error type. Chapter 5 investigates the error processing cost based on a mix of different errors and tests the effects of some error characteristics on their processing time. In Chapter 6, I report the results of my web-based evaluation study of non-standard features in English and compare them to the processing cost. For a similar analysis of learner errors, the evaluation data are adopted from another study. The purpose is to establish correlations between the groups of speakers in relation to the task.

The conclusion sums up the most important findings of the present study, integrating them with the previous research on error evaluation and processing. Suggestions are made for the development of the field.

This project is designed primarily for applied linguists who would like to know more about processing but do not have the tools to study it properly. In discussing complex psycholinguistic and statistical matters, I tried to avoid technical language where possible and to explain the effects in simple words. This thesis might also be interesting to psycholinguists with regard to the method of investigating heterogeneous data and surprisingly convergent results.

2. Error Studies in Applied Linguistics

Learners' tendency to transfer native language structures to their foreign language was thought to be one of the important **sources of errors**. To minimize the negative outcome of language transfer, since the 1950s **Contrastive Linguistics** has investigated differences and similarities between every pair of most common world languages, concentrating its findings in **pedagogical grammars**. While Contrastive Analysis compared the native and the foreign language for "consciousness raising", and **Error Analysis** compared the learner's **interlanguage** to the target language, the comparison of the interlanguage to the native language was carried out by the **Transfer Analysis**. As a result, researchers developed **error taxonomies**, conceptual underpinnings of the difference between **errors** and **mistakes** as well as between **grammaticality** and **acceptability**. A number of studies focused on **error gravity** with an aim to produce a **hierarchy** of error types. Surveys in **error evaluation** searched to explain the reaction to errors by the variables of the **readers'** and **writers' characteristics**. This chapter presents the concepts of Error Analysis, an overview of error evaluation studies popular in the 1970s-1990s and similar questionnaire-based surveys carried out to this day.

2.1 Error studies in Second Language Acquisition

2.1.1 Error Analysis

Since the 1960s, Error Analysis has become a popular research field in Applied Linguistics and Foreign Language Teaching. It developed due to the theoretical concepts and empirical research of two hypotheses explaining the nature of errors in second language acquisition. According to the Contrastive Hypothesis established by Robert Lado (1957), errors are caused by differences in language structures and learners' tendency to transfer elements of the mother tongue to the system of the target language. Errors can be avoided by means of a systematic contrastive analysis of the two languages and preventive didactic strategies. According to the Interlanguage Hypothesis, proposed by S. Pit Corder (1967) as "transitional competence" and developed by Larry Selinker (1972), second language acquisition progresses in steps, which are predictable to a certain degree. The development goes through some intermediate states between the native and the target language, so errors appear due to the lack

of competence typical of every learning level. The concept of interlanguage "validates learners' speech, not as a deficit system, that is, a language filled with random errors, but as a system of its own with its own structure" (Gass and Selinker 2008: 14). There are didactic measures, such as correcting and preventing errors, facilitating students' transition from one level to another. These two theories share a few common principles: they acknowledge transfer, claim the predictability of deviant language performance during acquisition, draw their findings from empirical error analysis, and use errors "likewise for diagnosis and evaluation of the process of language acquisition as for the development of therapeutic or corrective language teaching strategies" (Spillner 1991: IX).

The concept of interlanguages as developmental stages is well illustrated in the logo to the "42nd Annual International Conference on Language Teaching and Learning & Educational Materials Exhibition" (25-28th of November, 2016, in Japan). As I interpret it, the learner language at the beginning level (the bottom) can be compared to a zigzag, for the utterances are awkward and perceived as 'bumpy'. The red color signals a high irritation potential. As learners' proficiency progresses, the interlanguages become smoother, and the colors get calmer. A native-like command of the foreign language is portrayed as a smooth blue line, as pleasant for the perceiver as a calm sea. From the point of view of the interlanguage model of second language acquisition, an error is the difference between the current performance and the nearest higher level of proficiency the student is aiming for, and not the target language (cf. James 1998: 8).

Throughout the history of research in Second Language Acquisition (SLA), error analysis has been carried out for many source and target languages, in different learning situations, for various ages of learners and teaching methods. "The attempt to discover more about L2 acquisition through the study of errors was itself motivated by a desire to improve pedagogy" (R. Ellis 1996: 48). According to Corder (1967), the study of errors was significant in three respects: errors tell the teacher what needs to be taught; they tell the researcher how learning proceeds; and they are a means whereby learners test their hypotheses about the L2.

The research in Contrastive Linguistics relied on the following assumptions (Kortmann 2005: 156):

1) foreign language acquisition is different from first language acquisition;

2) L2 is always acquired against the background of a speaker's L1;

3) learners usually find certain features of a foreign language easy to learn, while having difficulties with others.

This last point gave rise to the distinction between positive and negative transfer. The similarities between two languages make the target language easier to learn because they trigger positive transfer in which a native language structure with minimum adaptation fits perfectly with the L2 system. It is the differences that cause difficulties in acquisition, and, without proper awareness, the interference of L1 structures in the system of the target language leads to errors.

Tricky cases for a learner are those which look like similarities on the surface but turn out to be differences. These are usually called "false friends". Relevant examples can be found in phonology, vocabulary, grammar and other domains of language use. For instance, the letter "B" in Russian corresponds to the sound [v] and not [b], as in English. German and English have words with a similar form (such as *argument, art, bad, direction, gymnasium*) but a different meaning. In grammar, German learners of English are tempted to overuse the Present Perfect for all kinds of past tense events or to form a *be*-perfect for motion verbs.

With all the useful findings of Contrastive Linguistics (CL), its role should not be overestimated, especially concerning the prediction of errors. "The catalogue of errors which CL predicts because of the structural differences between two given languages only partially coincides with the errors actually occurring" (Kortmann 2005: 159). For example, it is clear why a Russian learner of English can produce a phrase like *I very good fellow* (Russian has no articles and does not employ the form of *be* in the present tense). But for some reason, Spanish learners of English frequently omit copula verb forms, too, although Spanish has them, or the same type of error occurs in speakers of other languages and even in children learning English as an L1 (Odlin 2003: 3). "Not only did errors occur that had not been predicted by the theory, but also there was evidence that predicted errors did not occur" (Gass and Selinker 2008: 98). It has been shown that only up to one quarter of all errors can be explained by L1 interference (James 1998: 181); for the other three quarters there is a number of other factors. Among them are learner strategies causing errors, e.g. false analogy, incomplete application of rules, the wrong concept of a target language word, undergeneralization and overgeneralization, overlooking "cooccurrence restrictions", or there are classroom-induced errors (op. cit.: 185-189). The causes of learner errors are usually subdivided into interlingual (interference) and intralingual, or developmental, with the latter type prevailing (cf. Touchie 1986). Developmental errors are similar in L1 and L2 acquisition.

Negative transfer causes different types of deviance. The easiest to identify are production errors (Odlin 2003: 36): (1) substitutions – the use of native language forms in the target language; (2) calques – a very close reflection of L1 forms, often a literal translation; and (3) alterations of structures – e.g. hypercorrections (an overuse of transferred features). It is the substitutions and calques that are most often meant by interference and result in what everybody conventionally calls "errors" or "mistakes".

Interference also involves underproduction (and, as a result, underrepresentation) of the target language structures which are different from the L1 structures and thus appear to be an extra challenge for the learners. Students simply avoid difficult, in all senses 'foreign', constructions. For example, Schachter (1974) found that Chinese and Japanese students of English tended to use fewer relative clauses than did those whose native languages had relative clauses similar to English. Often the consequence of underproduction is an overproduction (overrepresentation) of easier constructions, mainly common for L1 and L2. Thus, avoiding relative clauses, "Japanese students may violate norms of written prose in English by writing too many simple sentences" (Odlin 2003: 37). Or, for instance, American learners of Hebrew tend to make apologies more often than it is common in Hebrew, thus transferring their native language habits. Russians are known for overusing *to my mind, you see, you know, well* (Swan and Smith 1987: 159):

(2.1.) *a. 'What time does the film start?' *'To my mind, at seven.'*
*b. *Well, you see, I often go to the Crimea, you know.*

Under- and overrepresentation are also violations of norms, of stylistic or pragmatic nature, and yet, they are rarely classified as errors, rather as "certain stylistic effects" (Kortmann 2005: 158). Foreign language teachers find it a challenge whether or not they ought to correct in the students' essays the kind of phrasing that "results in unidiomatic language use and gives the impression that a native speaker would have expressed the same content or issue differently" (op. cit.: 157). A text evaluation study (Ilin 2008) showed that native speakers are quite sensitive to such deviations, and sometimes they add more foreignness to the text than occasional grammar mistakes.

Language transfer takes place not only in phonetics, phonology, syntax, morphology, and semantics but also in the realm of discourse, or pragmatics. Inexperienced foreign language users tend to decode an L2 utterance as if it was literally translated into their L1 and may thus infer something different from what was originally meant by the speaker. For example, in Russian, phrases corresponding to the English "How are you?" are usually taken not as part of a greeting but as an independent question which presupposes an

answer. Or German learners may use the word *please* when offering something or as a formulaic reply to thanks meaning *Here you are* or *You are welcome*.

Discourse transfer has been most thoroughly studied regarding coherence and the expression of politeness (requests, apologies, greetings, etc.). A "failure to use 'conversational fillers' in other situations may make Russians sound impolite when, in fact, they do not mean to be so" (Swan and Smith 1987: 159):

(2.2.) *a. 'Would you like to go there?' *'No, I wouldn't.'* (for *'Well, I'm afraid I can't because…'*)
*b. *Tell me please how to go to the station.* (for *Excuse me, could you tell me the way to the station, please?*)

Another example is Kaplan's (1966) comparison of the writing done by speakers of different languages where he noticed that the thought pattern in English resembles a straight line ("direct" and "to the point"). Writing in Russian, by contrast, resembles a zigzag and writing in Oriental languages "a widening gyre"; however, there are counterarguments to these observations.

Error Analysis has been criticized for the weaknesses in methodological procedures and limitations in scope. R. Ellis (2008: 60-61) answers to some criticisms, pointing out that they are not essential. Among most serious ones is that Error Analysis was preoccupied with a description of what is wrong in the learner language without looking at what is correct. As a result, it has not created a complete picture and has not offered a theory.

In sum, Contrastive Analysis failed to predict all learner errors, but it created a data base of potential errors which help diagnose and explain deviations from the target language caused by the interference of L1 structures. Although the term "transfer" is most often used for native language influences, it has been acknowledged that individuals can transfer structures from any language they previously acquired (no matter how well). Therefore, the modern term for transfer is "crosslinguistic influence". The interpretation of transfer within behaviorist models of learning gave way to cognitive (mentalist) approaches. It has been shown that most learner errors are developmental and that particular types of them no longer occur as students move on to the next stage of their language proficiency.

2.1.2 The distinction between errors and mistakes

Many lay sources claim that there is no difference between an *error* and a *mistake*: the words are fully synonymous and are often defined through each other. However, some dictionaries distinguish between these terms offering the following explanations. An error suggests the departure or deviation from what is right or correct, a failure to make effective use of a standard or guide, e.g. *an error of judgment*. A mistake implies the wrong action resulting from carelessness, inattention, misunderstanding, haste or taking one thing for another, and it does not in itself carry a strong implication of criticism, e.g. *dialled the wrong number by mistake*. There are specific collocations employing one word or the other, e.g. errors 'occur', but mistakes 'are made'. A native speaker's intuition tells that errors happen to machines, but mistakes are made by humans, *error* is used in the language of science, in a more formal register, errors are repeated systematically, but a mistake is rather an accident and is not serious.

Applied Linguistics has developed its own definitions. Based on the distinction introduced by Chomsky (1965) between competence (the speaker-hearer's knowledge of the language) and performance (the actual use of language in concrete situations), Corder (1967) drew a line between an error as a failure in competence and a mistake as a failure in performance.

> Mistakes are akin to slips of the tongue. That is, they are generally one-time-only events. The speaker who makes a mistake is able to recognize it as a mistake and correct it if necessary. An error, on the other hand, is systematic. That is, it is likely to occur repeatedly and is not recognized by the learner as an error (Gass and Selinker 2008: 102).

Thus, in Second Language Acquisition (SLA) and English Language Teaching (ELT) literature, the distinction between competence and performance failures was interpreted in practical terms: an error occurs when the student does not know the correct form; when he or she knows the correct way, it is a mistake. A possible criterion to tell errors from mistakes is based on the frequency of occurrence: errors are systematic, but mistakes are random. However, this distinction was subject to reasonable criticism. R. Ellis (1985) notes that competence is not directly observable but only through observing the performance, and it is difficult to decide whether the cause of the error appears in the competence or in the performance domain. Besides, mistakes can also be made systematically. Littlewood (1984) doubts that errors and mistakes are clearly distinct in their psychological reality. Moreover, a learner's competence is too unstable to serve as a basis for demarcation between error groups.

While the neutral term for an "unsuccessful bit of language" (James 1998: 1) is **deviance**, James suggests that the "clearest and most practical classification" (op. cit.: 83) is the following:

- **slips** (can be quickly detected and corrected by their agent without external help),
- **mistakes** (can only be corrected by their agent if pointed out),
- **errors** (cannot be self-corrected and require further relevant learning), and
- **solecisms** ('incorrect' according to prescriptive grammar taught in schools).

In spite of the consistency of the suggested definitions, in the literature on Error Analysis, SLA and ELT, the term *error* overwhelmingly prevails. However, in forums and surveys (e.g. Ilin 2008 and studies outside the academia discussed in section 2.3.2), native speakers of English most often use the word *mistake*. Is this because all researchers assume that the deviance is caused by the lack of knowledge, and folk readers explain it as accidental? There is probably more of a stylistic difference: *error* sounds more formal and 'scientific', otherwise *typographical error* does not make sense.

Since most of the bibliographic sources used for this thesis do not consistently observe the distinction between the types of deviance, the tradition of using predominantly *error* in academic writing will be followed; however, if *mistake* is used, it most likely means "a slip".

2.1.3 Grammaticality and acceptability

"The whole concept of error is an intrinsically relational one" (Hawkins 1987: 471). Who defines what should be considered correct and wrong in English? Native speakers' ideas about grammar often operate on intuitions. For example, in his experiment on error evaluation, Hultfors (1986: 52) intended as errors the omission of the definite article and the lower case of the noun *hotel* in the sentence *I am staying at Sheraton hotel for three days*. He was surprised that very few of the 444 British informants identified the sentence as foreign, and when he asked to correct it, only 3 people out of 138 wrote *Hotel* with an upper-case initial, and 13 informants suggested rather *I was staying…, I stayed…, I have been…, I shall/will be…, I'm going to stay…, I am spending three days at…*, etc. (Hultfors 1987: 135). An opposite example is described by Barber (1964: 132) who once borrowed a book written in the 1930s by a distinguished literary scholar. In one place, the word *commonest* was vigorously crossed out by another borrower who had written in the margin "most common!!!" That reader did not know that *commoner* and *commonest* were the grammatically

correct forms popular a few decades before, but for the younger generation, they already looked like an error.

Quirk et al. (1985: 13) point out that speakers acquire the rules of their first language unconsciously, and if they are asked to explain them to a foreign learner, they have great difficulty. The grammatical rules for a language learned in a classroom setting, by contrast, seem much clearer because they have been spelled out. Learners' "heavy exposure to correctness-based instruction" (James 1998: 84) might be one of the factors that give native speakers "the impression at times that foreign learners speak their language 'better' than they" (ibid.). To make rational judgments about grammar, the knowledge of the language must be explicated. The standards are determined by professional linguists, native speakers, who also know ABOUT their language. How do they decide what should be counted as norm? "Since we do not have an Academy of the English Language, there is no one set of regulations that could be considered 'authoritative'. Instead, evaluations are made by self-appointed authorities who, reflecting varying judgments of acceptability and appropriateness, often disagree" (Quirk et al. 1985: 14). On aggregate, these authorities write the prescriptive grammar, defining what should be preferred or avoided in the standard variety, and spread it to the educational institutions.

Analyses of native and non-native speakers' production errors provided a base for looking at the variants in the language use and for deciding upon the terminology. The data collection involved observational and experimental studies (in a natural or manipulative environment), and introspection. The most common measures of deviance are the following.

1) **Grammaticality**, synonymous with well-formedness, is context-free; decisions are based on the grammar of the language, and there are no circumstances where this could be said in this way. "Appeal to grammaticality is an attempt to be objective, to take decisions such as whether some bit of a language is erroneous or not out of the orbit of human whim" (James 1998: 65).
2) **Acceptability** measures whether or not an utterance could be produced by a native speaker in an appropriate context. "To decide on the acceptability of a piece of language we refer not to rules, but to contexts" (op. cit.: 67).

A clear case of an error is a piece of language that is not grammatical and not acceptable, just as unarguably correct is a case that is both grammatical and acceptable. Yet there are examples that can be considered grammatical, but not acceptable. For instance, garden path sentences like *The horse raced past the barn fell* mislead the readers (Borsley 1991: 4), or other artificial examples

used in psycholinguistic experiments (e.g., *I like my coffee with sugar and dog*). And vice versa, there are utterances that are ill-formed but could be judged as acceptable. For example, *He is a not unintelligent person* violates the rules of the English language as a system but is quite comprehensible and could be used in some contexts. One can find plenty of ungrammatical but yet acceptable examples in non-standard varieties.

Ungrammatical and unacceptable utterances are usually produced by beginning learners. Advanced learners' language may sound odd or strange or 'foreign' due to "its tendency to be unacceptable while being grammatical. […] It is unacceptable because it disappoints the hearer's expectation of idiomaticity" (James 1998: 71). Foreign learners are known for breaking the "customary collocations" and for "unusual grammar or phonological configurations". The Construction Grammar approach could explain this by the insufficiency of surface-level constructions in the learners' inventory, while the language instruction feeds students with abstract, schematic constructions. As a result, learners know how to combine words correctly, but they choose other words or other syntactic structures than native speakers would prefer in the same context. One cannot predict that the corresponding question for *What time is it?* should be phrased in Russian as *Which hour [is it]?* (formal) or *How much time [is it]?* (informal). Idiomatic expressions require specific learning; they reflect the "**cooccurrence restrictions** of English, which are not governed by fixed rules but are probabilistic or 'weighted' in unpredictable ways" (op. cit.: 75).

Native and non-native speakers sometimes make the same errors; that is why it is difficult to make assumptions about the status of the writer based on isolated sentences. For example, children often produce erroneous forms based on analogy or overgeneralization in their native language. An error appearing from overgeneralization is found in the following sentence used in Hultfors' experiment: *The house was full of mouses*.

> This sentence made a foreign impression on the informants but was found to be quite easy to understand. In view of the fact that the error exemplified by this sentence is also often made by small children who have English as their first language, the mean for foreignness [3.43, max 5.0] may seem high (Hultfors 1986: 58).

In his study, the participants had been told that all sentences in the questionnaire were produced by foreign learners. If this error is assumed to be produced by a non-native adult, it is not surprising that it gets a high mean for foreignness. Otherwise, the same sentence might be interpreted as produced by an English speaking child, and a foreign impression does not arise at all (though it is unlikely because, by the time children learn to write, they do not

say *mouses* any more). The same holds for *This is the goodest cake I have ever tasted!* Used by a foreigner, it is an error of a learner who regularizes the superlative form of an irregular adjective. Used by a native speaker, a young adult, it is a 'stylistic device' because, as a slang word, *goodest* is "a hilarious way of saying *best*, while being annoying and original!"[1] But then it is not an error in the same sense because the wrong form is produced intentionally. In the situation of perception, unlike in a context-free linguistic analysis, "the same error may be evaluated very differently depending on who made it and where, when and how it was made" (R. Ellis 2008: 60). Depending on the known or the assumed information about the writer, the same error produces a different effect.

It is widely believed that errors cause irritation, which is defined as the result of the *form* interfering with the perception of communication. "The irritation continuum ranges from unconcerned, undistracted awareness of a communicative error to a conscious preoccupation with form" (Ludwig 1982: 275). In other words, a low-irritating error is noticed but does not bother, whereas a highly irritating error makes the reader pause and take time to recover.

There are situations when errors may be noticed but are readily ignored, that is when they are produced by attractive and influential personalities, "the Kissingers or Pavarottis of this world" (James 1998). Page (1990: 105) refers to "sympathetic native speakers" – people who are more interested in the message or its bearer than in its formulation, people who overlook imperfections of form. The cause of such behaviour of the audience should rather be attributed to the personalities who 'cast a spell' with their professional success, so that their poor language achievements become less important.

It has been debated whether irritation correlates with comprehensibility. Some researchers believe that lower comprehensibility entails greater irritation (Johansson 1978, Ludwig 1982). The findings of Kresovich (1988) support the claim that the more an error obscures meaning, the less it is tolerated. Other researchers, e.g. Santos (1984), Vann, Meyer and Lorenz (1984), Rifkin and Roberts (1995) disagree. They regard irritation "more as a function of the expectations and characteristics of interlocutors, who may become irritated by errors even when the message is comprehensible to them" (Santos 1988: 70). In that study, professors found the sentences with errors, on aggregate, highly comprehensible, reasonably unirritating, but linguistically unacceptable, thus making a clear distinction between these concepts.

[1] http://www.urbandictionary.com/define.php?term=Goodest

Rifkin and Roberts (1995) point out that there are errors that are both comprehensible and irritating. On the other hand, Stewart, Ryan and Giles (1985) report that American college students rated British English as difficult to understand but did not downgrade its status, which implies that low comprehensibility may not necessarily be irritating "when the evaluators' social attitudes are taken into account" (op. cit.: 522). In the study by Santos (1988), University professors found double negation (*They wouldn't get nowhere unless they used a translator*) the least acceptable, the most irritating among the learner errors, although comprehensibility was not an issue.

> It seems clear that the reaction here is a *social* rather than a strictly linguistic one and is undoubtedly a transfer from attitudes toward less educated native speakers as well as attitudes ingrained after years of prescriptive education (Santos 1988: 84).

Thus, there is no direct relationship between irritation and comprehensibility, and it has been assumed that social factors play a more important role than linguistic ones in forming the subjective judgment of how offensive the error is to the perceiver.

2.2 Error Gravity

2.2.1 Sociolinguistic factors determining error gravity

Williams (1981) identifies two steps in text perception: (1) whether the rule is violated or not and (2) whether the rule violation is noticed and responded to or not. James (1998) adds two more steps: (3) whether the reaction is positive or negative and (4) to what extent, that is, the measure of (dis)approval – error gravity.

Language teachers often face the dilemma whether they should emphasize and correct particular errors. Studies in error gravity have searched to establish instructional priorities by investigating the opinions of audiences for whom learners may write. James (1998: 205) explains the purpose of error evaluation in the following way: "to prevent obsession with trivial errors and give priority to the ones that really matter."

There were three main research questions of error evaluation studies (cf. R. Ellis 2008: 56). (1) Are some errors more problematic than others? (2) Are there differences in the evaluations made by native and non-native speakers? (3) What criteria do judges use in evaluating learners' errors?

Johansson was one of the first to focus on error evaluation and reactions to non-native English. Explaining the relevance of his study, he says that "error evaluation must be considered a largely neglected field of research" (Johansson 1978: 1), although a need to establish error gravity had previously been pointed out by James (1972: 76) and Robinson (1973: 192), and there had been occasional attempts to do it in a systematic way by Bansal (1969) and Olsson (1972, 1973).

Johansson concentrated on the pragmatic effect of errors, analysing not only the intelligibility of deviant utterances but also the degree of irritation they cause to the readers. As there is no unified standard of correctness in SLA, it is often up to the teacher whether to mark a certain case as an error or not. Johansson claims that not only 'overt' errors (in grammar, pronunciation or spelling) should be regarded as wrong, but also 'covert' errors – discourse violations extending beyond the local sentential context. The criterion of determining error gravity should be its degree of interference with the discourse: the most serious errors are those which make the utterance fully unintelligible. "If the erroneous utterance is fully comprehensible, it could nevertheless have serious consequences from the point of view of communication, e.g. make the receiver tired or irritated" (Johansson 1978: 4). Erroneous utterances that are comprehensible and do not cause irritation can be classified as containing low-gravity errors.

However, a simple ranking of error types is not sufficient. Johansson (1978: 7-8) suggests considering such factors as (1) receiver characteristics (age, education, regional and social dialect, degree of association with people from other countries); (2) the type of language situation (formal or informal, speech or writing, disturbed or undisturbed communication, etc.); (3) the role of the error producer (whether the learner is a tourist, a visiting scientist, a secretary, an interpreter, etc.). On the whole, adding the sociolinguistic dimension extends the research enormously, making it impossible to produce a single hierarchy of error types.

In his own experiments based on errors typical of Swedish learners of English, Johansson explored subjective and objective measures of error gravity, overt vs. covert errors in written English, lexical and grammatical errors, foreign accent and speech distortion, types of phonological errors, by native vs. non-native judges. For a better control of the test situation, he used constructed materials containing specific error types as variables. In most cases, errors appeared in sequences of isolated words or disconnected sentences and were submitted to groups of non-linguistics students at the University of Lancaster, England. His experiments resulted in the conclusion that lexical errors (in a written text) cause more serious problems of interpretation than grammatical

errors. Five repetitions of the same error (no matter lexical or grammatical) were judged to be less serious than five occurrences of different errors. On the 'irritability' scale, articles and spelling errors proved to be at the bottom, whereas wrong prepositions and word order appeared at the top.

Other researchers did not see irritability as a gradient property of errors. "All errors are equally irritating […] one should not expect to be able to establish a hierarchy of errors with respect to irritation" (Albrechtsen, Henriksen and Færch 1980: 395). Instead, James (1998: 222, his emphasis) suggests that "the irritation potential of an error is a reflection of three factors: its **predictability**, the **social relationship** obtaining at the time between speaker and hearer, and the degree to which it infringes **social norms** (as well as linguistic ones)."

Another study on error evaluation was carried out by Petti (1979). He asked Swedish, British and American university teachers to grade approximately four hundred errors. In his final discussion, the focus was mainly on the assessment made by his colleagues, "since linguistic achievement in terms of the traditional contrastive grammar system as taught in Sweden was chiefly being measured" (Petti 1979: 3). He also used disconnected sentences as test material. According to Petti's investigation, certain errors in the use of non-count nouns, verbs, pronouns, subjects, and word-order errors were judged as more serious than articles and some vocabulary errors. However, the results do not present a systematic ranking of error gravity.

A hierarchy of error types was more successfully completed by Hultfors (1986). He undertook a profound investigation of error gravity based on the criteria of acceptability and intelligibility, also accounting for the variation of participants and the 'foreigner role' as outlined by Johansson. For his questionnaires, Hultfors used 75 sentences selected from a corpus of texts produced by Swedish learners of English, where 5 of them were correct (control) sentences "to check the reliability of the informants". The participants consisted of 444 native speakers of British English of different age, gender and social status.

In Questionnaire 1, each sentence was evaluated based on a 5-point scale from "native-like" to "very foreign" and from "very easy to understand" to "very difficult to understand". For the final error gravity index, the value of foreignness was added to that of intelligibility difficulty. After applying relevant statistical methods, Hultfors presented a rank list of 39 error types. The ten most serious error types are shown in Table 2.1[2].

[2] The table was extended compared to the original: the examples of test sentences and full names of the error types were added for clarity.

Rank no.	Error gravity index	Error type	Example of a test sentence
1	7.22	Personal pronoun substituted for reflexive pronoun	*She was standing alone, beside her with rage.*
2	6.97	Omission of *do*-periphrasis	*He thinks not that they know what to do.*
3	6.79	Incorrect use of modal auxiliary	*Can you French?*
4	6.76	Attributive *own* substituted for predicative *own*	*He has an own company.*
5	6.69	*To*-infinitive substituted for *ing*-participle	*Many cities have stopped to expand.*
6	6.54	Incorrect choice of adverb	*He works very hardly.*
7	6.27	Incorrect use of idiomatic phrase	*Excuse me, what is the clock, please?*
8	6.11	Insertion of preposition	*She was here for two years ago.*
9	6.05	Incorrect choice of grammatical subject	*It was little else to do.*
10	5.94	Incorrect comparison of adverb	*He drives badlier than his brother.*

Table 2.1. *The ten most serious error types (cf. Hultfors 1986: 240)*

Speaking about the relevance and application of such an error hierarchy, "the intention has not been to produce scales against which learners' errors have to be judged. The results may, however, serve as a useful guide in the planning of teaching and in the evaluation of learners' errors" (Hultfors 1987: 209, his emphasis).

Questionnaire 2 in Hultfors' study was designed to investigate how the social role of the L2 learner influences the perceived error gravity. The same 75 sentences used in Questionnaire 1 were graded by 118 British participants who had to assign the seriousness of the error depending on the status of its producer.

		SERIOUS	MEDIUM	MINOR
a)	Foreign tourist visiting Britain	☐	☐	☐
b)	Foreigner working as a secretary in Britain	☐	☐	☐
c)	Foreign manager of a big firm on business in Britain	☐	☐	☐

Table 2.2. *Variables testing the 'foreigner role'*

As a result, the "average test error" was judged to be of rather medium gravity when made by a foreign learner working as a secretary in Britain or a foreign manager and of rather minor gravity when made by a foreign tourist visiting Britain (op. cit.: 104). In general, the errors in almost every sentence were

judged as more serious when made by a secretary, less serious when made by a foreign manager, and quite minor if made by a tourist visiting Britain.

In Questionnaire 3, each of the 75 sentences was followed by a line where the participants (138 total) were asked to write their own version of the test sentence, correcting whatever they considered an error in the use of English. The majority of the informants identified the errors as expected by the experimenter. It appeared that six of the eight sentences interpreted differently by ten or more per cent of the informants are found in the upper half of the intelligibility ranking list. The largest number of unexpected corrections was caused by the omission of the auxiliary *do* as in *Became he a dentist?* and *Came he to the party yesterday?* In most cases, the participants changed the sentence from an interrogative to a declarative one. Many of the unexpected corrections altered the meaning of the original sentence.

Hultfors' survey showed that the selected test sentences with errors typical of Swedish learners of English do not cause serious comprehension problems but are likely to produce a strong foreign impression on native speakers of English. Teenagers found most of the test sentences more difficult to understand and thus showed higher demands on learners' English than the older informants did. Adults only had higher demands on the language of a foreign manager. There were no significant differences in the answers of the male and female participants. Besides, people with a higher education, people who often talk to foreigners and people who have lived abroad graded the test sentences as easier to understand and had lower demands on learners' English.

2.2.2 Effects of the rater characteristics

A series of studies were aimed at comparing error evaluations by different subject groups, including native and non-native speakers as well as teachers and 'non-teachers'. In most cases, non-native speakers rated the errors and overall learners' writing more strictly than native speakers did (James 1977, Sheory 1986, Santos 1988, McCretton and Rider 1993, Porte 1999, Hyland and Anan 2006, Marefat and Heydari 2016). In attempts to explain this result, the authors suggest that non-native speaking teachers "may have felt that their own knowledge of the language was being tested, and as a reaction to this, tended to mark more severely" (McCretton and Rider 1993: 182). Another argument was that non-native speaking professors "have attained an extremely high level of proficiency in English and, because of their investment of effort in the language, judge the errors of other NNSs [non-native speakers] more severely than do NS professors" (Santos 1988: 85). In one study, in the

assessment of grammaticality, the native speakers were stricter than the non-native speakers who "have rated both compositions more positively" (Kobayashi 1992: 81).

Garrett and Austin (1993) conducted a judgement study on the use of apostrophe in English, collecting responses from British postgraduates in teaching English as a foreign language, undergraduates at the same university majoring in other subjects, and German undergraduate 'Anglisten'. The authors suggest that the difference was not so much in being a native or a non-native speaker of the language as having explicit knowledge of the relevant linguistic features. In some arbitrary areas where one has to be taught the rule to be able to use the item correctly, foreign learners are more equipped than native speakers. This was indicated by the fact that the English linguistics students from Germany showed the highest accuracy in error detection.

In the study by James (1977), foreign language teachers were operating on a wider scale, ranking more finely than native speaking teachers did. The error types most stigmatized by the non-native speakers were wrong lexical choices (*Christ ascended into *sky*) and wrong uses of prepositions (*He disapproved * my idea, he laughed *with my joke*).

Other studies indicate that non-teachers tend to be more lenient in their judgements (e.g. Schairer 1992, Hadden 1991). "Taken together, a majority of studies suggest that native speaker non-teachers are the least severe judges of L2 error, one plausible explanation for which is that non-teachers just do not notice many learner errors" (Murray 2002: 189).

There are also effects of age and specialization. "Professors in the physical sciences rated the acceptability of the language of the compositions significantly lower than did those in the humanities/social sciences" (Santos 1988: 80). This result was confirmed by Roberts and Cimasco (2008) who additionally found that "the older professors displayed a lower degree of irritation in their ratings than did the younger professors" (op. cit.: 81).

In Hughes and Lascaratou (1982), both groups of native speakers (teachers and non-teachers) were more lenient in judging the seriousness of errors than the non-native speaking teachers. The sentence ratings of the Greek teachers correlated with those of the English teachers at 0.40 and with English non-teachers at only 0.07, while English teachers' and non-teachers' ratings correlated at 0.82 (op. cit.: 178). "In judging the seriousness of errors, the Greek teachers made reference to the 'basicness' of the rules infringed, while the non-teachers depended almost exclusively on the criterion of intelligibility" (op. cit.: 175). The native speaking teachers of English declared the importance of both criteria but showed preference for the intelligibility. Interestingly, native speaking laypersons valued intelligibility three times as

highly as the teachers. This fact led the authors to suggest that students' work must be assessed primarily with reference to the effectiveness of the communication.

There is abundant evidence that native speakers tend to be more tolerant of errors and show readiness to reduce error gravity by thinking of the contexts where this utterance would be appropriate (cf. E. Davies 1983). At the same time, native speakers in another study tended to extrapolate from the experimental contexts and assess the gravity in terms of the severe consequences this error COULD have in real life (Hughes and Lascaratou 1982). It confirms that, for native speakers, the form is important not "for form's sake" (James 1998: 233) but rather as possible interference with intelligibility and the communication process.

As writings are normally rated based on the student's performance at the surface-level (grammar and vocabulary) and discourse-level features (content and organization), non-native speakers are usually harsher toward the surface-level errors and "rely more on rule infringement rather than intelligibility" (Hyland and Anan 2006: 509). Native-speaking teachers in their study placed emphasis on style and appropriateness. Connor-Linton (1995) also noticed that "the Japanese EFL teachers focused on matters of accuracy (content, word choice, and grammar), while the American ESL teachers focused on both intersentential features of the discourse and specific intrasentential grammatical features" (op. cit.: 99).

However, in a single study, native speaking teacher candidates "perceived accuracy-related errors (e.g., the misuse of English vowels, misplacement of adverbs, and number disagreement) to be more troublesome than other elements such as lexical choice or style, let alone content and organization" (Kang and Veitch 2016: 21). Mahoney (2011) investigated the differences in the teacher and student perceptions or written errors and found that the most significant teacher-student gaps were apparent when judgments of error involved phonic distinction, intelligibility, and context breakdown.

> "Many students are not capable of correcting a dictation quiz in a manner similar to that of their teachers. When given an answer key, students perceive what is wrong but not to what degree. Students' marks resemble those of their professors when sentences lack words or contain misspellings, but they stray when evaluative decisions involve context and intelligibility (op. cit.: 125).

The results of the several above-mentioned studies imply that, on the one hand, there might be cultural differences in the evaluative criteria (Brits and Americans clearly set comprehensibility and linear content structure as priorities). On the other hand, focusing on the 'petty' surface-level errors and

ignoring global organization and discourse failures could be indicative of the raters' immaturity. Fellow-students and even teacher candidates, so to speak, do not see the woods for the trees. To balance the criteria, it has been recommended that "language teachers should take a moderate course between consistent correction and non-correction of student errors, offering support without discouraging student efforts to communicate" (Kresovich 1988). Similarly, "rater training programs in Iran should aim at deemphasizing the role of grammar while at the same time highlighting the role of inter-sentential discourse" (Marefat and Heydari 2016: 24).

The large variability of the ratings in the performance assessment and the different strategies of the raters became the subject of the study by Zhang (2016). "The results suggest that more accurate raters are better at integrating information from target essays and are more self-conscious about their rating accuracy" (op. cit.: 37). Alongside the attempts to explain the rating priorities in being a native or a non-native speaker, a teacher or a non-teacher, there are studies based on think-aloud protocols that attribute the rating variation to the individual types. They suggest that there is more variation inside the native and non-native speaker groups than between them. Eckes (2008) classified the raters according to their dominant rating criterion: syntax, correctness, structure, fluency, non-fluency and the "non-argumentation type". Baker (2012) categorized the raters based on their decision-making style into rational, intuitive, dependent, avoidant and spontaneous.

To summarize the results, Rifkin and Roberts (1995) made a review of 28 error gravity studies (1977-1995) and were somewhat disappointed at the inconsistent findings of the error investigations that "make it difficult to point confidently in any one direction and proclaim it the route for improving native/nonnative interaction" (op. cit.: 512). Indeed, error gravity research was established with an urge for applicability of the results in the teaching English as a foreign language, but the findings turned out to be quite divergent and, among other things, revealed a gap between native and non-native speaking teachers, possibly due to differing values or even cognitive constraints.

2.2.3 Hierarchies of error types

Some researchers noticed that the judges were consistent in the ranking of the error types, which suggests that the categories have some psychological validity (cf. James 1977). One of the explanations for the tendencies in foreign language teachers is that they must have developed a system of principles for the evaluation that guide them in marking students' writing. Vann, Meyer and

Lorenz (1984) note that most people see errors "in relative rather than absolute terms" (op. cit.: 437), it is therefore sensible to order them according to their importance.

All error gravity hierarchies presented in this section have been adapted for better comparison and organized in the same way – from least severe error types at the bottom to most severe ones at the top. To highlight the differences, lexical and syntactic errors are marked in red and blue. It helps to follow the different positions these error types take in the hierarchies.

Sheorey (1986) presented an error gravity scale in three large categories.

		Most serious:
	Less serious:	verb-form errors
Least serious:	articles	lexical errors
spelling	prepositions	

According to Tomiyana 1980, errors in articles are easier to correct and therefore less crucial for communication than errors in connectors.

Teacher candidates in the study by Kang and Veitch (2016) showed the following preferences.

		Most serious:
	Less serious:	adverb misplacement,
Least serious:	style,	misspelling of English vowels
the error in the	singular/plural	
existential	distinction,	
expression		

Delisle (1982) offers a hierarchy of error types based on the evaluations by German 13-17 year-olds of written sentences produced by English speakers learning German (op. cit.: 40).

vocabulary
verb morphology
gender
word order
case endings
spelling

The utmost gravity of lexical errors was also elicited in the study by Santos (1988).

Below is an error gravity hierarchy based on the responses of native speaking English teachers (Hughes and Lascaratou 1982: 178).

pronouns
vocabulary
verb forms
plural, spelling
word order
concord
prepositions

Vann, Meyer and Lorenz (1984) asked Iowa State faculty members to rank "foreign student errors" on a 5-point acceptability scale. The study resulted in the following error gravity scale (op. cit.: 431).

word order
it-deletion
tense
relative clauses
word choice
subject-verb agreement
pronoun agreement
preposition
spelling-2 (deletion and substitution)
comma splice (connecting two complete sentences with a comma)
article
spelling-1 (differing from standard American spelling: 1 British, 1 colloquial)

The next hierarchy was elicited from native and non-native speakers of English in the study by James (1977: 124).

transformations
tense
concord
case
negation
articles
word order
lexical errors

Likewise based on the combination of the native and non-native speakers' results, McCretton and Rider (1993) suggested there must be a "universal hierarchy of errors" (op. cit.: 183).

concord
verb forms
prepositions
word order
negation
spelling
lexis

While we may look for explanations why lexical and word order errors are changing places from hierarchy to hierarchy, there are a few other problems concerning not the results but rather the methodology of arriving at these scales, which affects their interpretation and application. One problem (also discussed by others) is that the categories, i.e. the error types, are not directly comparable due to the inconsistent terminology. For example, what James (1977) calls "case" (*He disapproved my idea, He laughed with my joke)* corresponds to McCretton and Rider's (1993) "prepositions"; his "tense" (*I am born in 1942, He didn't found the purse he lost*) corresponds to their "verb forms", etc. Besides, particular error types are included in the sample by some researchers and excluded by others. It is clear that the names of the categories and their contents are arbitrary, i.e. every researcher made his/her own choice, and the label chaos only causes inconvenience.

A larger problem, however, is that the labels were so tightly dependent on the concrete examples that there is no certainty that the elicited position on the error gravity scale can be extrapolated to other sentences that could be labeled in the same way. In other words, the error types assigned to the sentences often do not get to the roots of what exactly is wrong there. It seems that the researchers have noted different 'symptoms' on the surface of the sentences but rarely looked into the 'cause of the disease'. For instance, the category of negation in the hierarchies is too broad: it can include double negation or auxiliary deletion, and each of these types may entail a different gravity for different reasons. As another example, it is not surprising that the sentence *She was standing alone, beside her with rage* received high error gravity scores (see Table 2.1 above), but it is unlikely because a "personal pronoun substituted for a reflexive pronoun". A number of L1 non-standard features involve modifications of personal and reflexive pronouns, and there is little chance that this could be THE MOST serious error type, especially for learner errors. Rather, this sentence was supposed to have a partially filled idiomatic construction (*beside herself with rage*), a low constraint and a low frequent one, which reduces the chances of quick repair. On the surface, *beside her* looks grammatical, making the reader expect a verb or an existential *there* following it, but instead, the prepositional phrase makes the sentence almost unintelligible. Alternative to 'pronoun substitution', this stimulus could be marked as 'morpheme deletion' based on the missing *–self*. But to be more accurate, it is a misconstructed idiom with such characteristics of it as low frequency and low predictability. Without looking at the core of the problem, such error hierarchies remain somewhat useless.

The most important limitation of the hierarchies, as I see it, is that such explorative and casuistic studies have failed to identify a driving force (or

forces) behind error gravity. Ideally, the motivations behind error ranking should be integrated into a linguistic theory. The categories used in the hierarchies are well-known in foreign language teaching and marking of students' writings. They may well reflect the learning priorities as established by teachers in the conscious metalinguistic decisions. The researchers had defined them prior to the evaluation procedures and only had to order them in accordance with the responses. Roberts and Cimasco (2008) criticize those methods and point out that one should not direct native speakers to the prior established categories but rather let them evaluate the text in a naturalistic manner, i.e. mark whatever they perceive.

From a psycholinguistic perspective, these hierarchies lack the single criterion (the factor) that would connect the error types into one logical sequence. The error types look as if they were taken from different taxonomies. Such heterogeneity could be possible if the hierarchy was produced as a result of a multifactorial analysis with a few competing variables. It would be more valuable than the hierarchies themselves if these competing factors were named. What drives error gravity upwards? Is it unintelligibility? Is it the size of misordered elements? Or instruction-induced attitudes? What reduces error gravity? It is locality? Or highly constrained contexts? How does item frequency work? These types of questions need to be answered with regard to the perception of errors.

2.3 Evaluation of native speakers' errors

2.3.1 Writer's identity

Error analysis has been carried out not only for foreign language learners but also for native speakers. In the 1920s, the pioneers of the native speaker error analysis distinguished between the normative approach to errors (involving attitudes and standards) and the functional approach (violating clarity, economy and expressiveness). Prescriptive grammar started with the Dictionary of English Normative Grammar as a collection of typical errors referring to the standards of the 18th century. Nevertheless, there has always been indeterminacy as to what to consider an error in English. James (1998: 11) describes how A.L. Jones in 1966 assembled a corpus of 386 essays of Malayan teacher training college students "to detect all the non-Standard-British-English features". The native speaker judges failed to decide on the deviance or non-deviance of certain features. For instance, 32 of 128 items stigmatized by one judge were acceptable to the other five. About 6% of the

original 'deviances' were in fact acceptable. Besides, the native speaker judges failed to reach consensus on the ideal correction of most errors.

It has been noticed that Americans seem to be less critical in their judgments and more careful in their comments. According to a native speaker's intuition, Americans are generally more hesitant about judging language for a few reasons.

1) The existence of British English – there are Americans, perhaps many, who consider British English to be 'more right' than American English, or to sound more educated or more sophisticated. Something what Americans consider wrong might suddenly sound very correct when pronounced with an RP accent.
2) The existence of such a large community of English-speakers around the world, all with different standards. Watching films from other English-speaking countries, or even reading books, one is exposed to 'foreign' possibilities in using English. Even in the US, the simultaneous existence of Black English, Standard English, and local accents or dialects means that people's exposure to forms they wouldn't say themselves is relatively high.
3) With political correctness, Americans often find it wrong to criticize other ethnic/social groups in favour of anything that the white mainstream society represents.

The standards of language performance applied to native speakers are higher compared to foreign learners. Rubin and Williams-James (1997) used the same essay with six kinds of errors and pseudonyms identifying the writers as Anglo-American, Danish or Thai and submitted each version to native speakers of English. The raters favoured the 'Thai' writer and mostly disfavoured their 'Anglo-American' counterpart. For the latter, the authors assume, there must have been 'they should have known better' attitude.

In the study by Janopoulos (1992), university faculty were asked to rate sentences containing errors commonly made by non-native speaking writers. Half the faculty were told they were rating learner errors, and the other half were made to believe the sentences had been produced by native speakers. The results indicate that the faculty were generally more tolerant of 'non-native speakers'' errors than when they perceived that the errors were produced by native speakers.

Another example is based on a text written in Russian by a native speaker of German and evaluated by 100 Russian informants (Ilin 2008). The writer's native language was identified correctly in the 16% of the questionnaires. It is interesting that 21% of the participants did not recognize a foreign learner but thought that the letter was written by a Russian native speaker. These participants mostly focused on the meaning rather than the language of the text, making assumptions about the circumstances of why he was writing. They often admitted the writer's poor grammar but regarded it as

lack of education – *he is illiterate, ignorant, was a bad student at school* (15 comments), or that *he is crazy, needs to see a psychiatrist* (3), or *he emigrated long ago* (2), or *he is a dialect speaker* (1). A few times the author had a chance to hear the immediate comments on the text and the writer before they were written down on paper. In the beginning, the judges were very critical of the writer, identifying him with unprivileged native speaker groups. But when, in the middle of the text, they realized that this oddness could appear due to the non-native speaker status of the writer, they said that IN THIS CASE, it is a very good level of proficiency for a learner, and the writer must be a very nice and educated person. This suggests that non-native authors can, in fact, be taken for native speakers, but their text may still not look 'perfect'. There is a risk that these imperfections may be attributed to undesired characteristics. Therefore, it may be of advantage to introduce yourself as a foreign learner in the very beginning (e.g. say what country you are from). In this case, the writer will be regarded as a well-educated person and appreciated for having made the effort to learn a foreign language.

Wall and Hull (1989) collected teachers' corrections and opinions about a student's essay. "Nearly three-quarters of their 140 classifiable responses (104, or 74.3%) said the errors in this text were serious because they got in the way of effective communication of meaning. […] A second set of responses (16, or 11.4%) said the errors in this text were serious primarily because they represented some problem with the student's education" (op. cit.: 277).

In his *Phenomenology or Error*, Williams (1981) identifies such error types in which no rule is violated, but the reader responds anyway, and those in which a rule is violated, but the reader does not respond. One can easily find errors in students' essays because the reader is determined to find them there. At the same time, one may fail to notice errors in a scholarly article because this is where readers do not expect them.

Roberts and Cimasco (2008) and Kang and Veitch (2016) found no significant effect of the writer's identity on the holistic rating of student essays. However, in the matched-guise experiment by Kang and Veitch (2016), there was a significant difference in the comments affected by the ethnic identity of the writer. When the author of the essay was thought to be Chinese, the recommendation was to place more focus on grammatical accuracy. For the Spanish guise, the teacher candidates suggested reading extensively and expressed "a form of empathy" (op. cit.: 20). The authors believe that teachers and raters are positively affected by the writer's culture they themselves were exposed to (e.g. Spanish in the US) and show more tolerance to such L1 interference in English. They conclude that raters attribute the same global scores but reach them based on different criteria influenced by the writers'

identity. Similarly, Connor-Linton (1995) makes "an important methodological point" that quantitative similarities in ratings may mask significant qualitative differences in the reasons for those ratings.

> Clearly, stereotypes exist, and they feed into attitudes and evaluations of groups and individuals. […] When asked to simply "think about" an ethnic or racial group (Callan & Gallois, 1983) or even look at pictures of ethnic group members (Fishman, Rattner, & Weiman, 1987) and then rate members of that group according to a list of personality attributes, subjects behave in a predictably negative or positive manner (Roberts and Cimasco 2008: 126).

In the evaluation process, on the one side of the scales is the readability of the text, but on the other side is the social context, i.e. what kind of person has written or could have written this text under which circumstances. There is evidence that the writer's identity (whether it is a native speaker or a foreign learner, his or her nationality, etc.) has an effect on the perception of written errors. When there is no personal acquaintance, the attitude to language learners depends not on who they are in fact but on who they are assumed to be in the deficit of the relevant information.

2.3.2 Error evaluation outside the academia

Reactions to errors are also studied in applied fields such as the teaching of writing skills for business communication (e.g. Leonard and Gilsdorf 1990, Beason 2001, Brandenburg 2015). It has been a separate research culture where the SLA "comprehensibility" corresponds to "clarity", and error gravity is measured in terms of how much errors "bother" ("serious" errors are consequently "bothersome"). The terms "errors" and "mistakes" seem to be used interchangeably. Writing instructors distinguish between three types of deviations (based on Shoebottom 2016):

1) "mechanical" errors, including orthography (spelling, capitalisation) and punctuation, partly overlap with "slips";
2) "grammar" errors involve morphology and syntax;
3) "usage" errors comprise cases that are grammatically possible in English but are misused in the context (often refers to vocabulary).

"Usage errors" is a popular term with various meanings. Native speaker students and teachers discuss it in forums and do not converge to the uniform definition. There is an interesting opinion, for example, that grammar mistakes (non-academics definitely prefer "mistakes" over "errors") refer to a violation

of the rule applicable to all members of the class (≈abstract constructions?), whereas usage mistakes involve 'single-word rules' (≈surface, low-level constructions?). For example, *he enjoy dancing* is a grammar mistake because all verbs require the suffix *-s* in the 3rd person singular, but *he enjoys to dance* is a usage error because it is the property of the verb *enjoy* that it requires a gerund. After having studied several web-based sources, I came to the conclusion that native speakers' most common deviations from the norm are called "usage errors" as they violate the 'official grammar' as a prescribed system. Grammar mistakes, by contrast, are usually made by foreign learners.

Hairston (1981) describes how she compiled a questionnaire (not for academic purposes) with 65 sentences containing common native speakers' errors and collected the responses from 84 top-level professionals in different non-academic areas in the USA. Each sentence had to be marked for one category out of three: "Does not bother me," "Bothers me a little" and "Bothers me a lot". The first striking result was that women were 'bothered' much more extensively than men, which indicates that women's attitudes toward language are more conservative than men's. The following details present the most "outrageous lapses" elicited in the study, according to business people.

> Respondents of both sexes reacted most strongly against errors that were so glaring they might be called "status markers". The most egregious example was "When Mitchell moved, he brung his secretary with him." Seventy-nine out of eighty respondents indicated that the sentence bothered them a lot. Other substandard verb uses such as "When we was in the planning stages," "Calhoun has went," "Jones don't think it's acceptable" also brought overwhelmingly negative responses. Readers very strongly disapproved of two other kinds of errors that might be called "status markers": double negatives and beginning a sentence with an objective pronoun. For example, "There has never been no one here like that woman," and "Him and Richards were the last ones hired" (Hairston 1981: 796).

Again, the "status markers" mean that particular native speakers' errors indicate lack of education. Note that the negative responses were elicited to errors that in English linguistics enjoy a status of non-standard features.

Out of the error examples discussed in this thesis (in Chapters 5 and 6), the high seriousness in Hairston (1981) was attributed to using an adjective for an adverb (*He treats his men bad*), the "colloquialism" *would of* in place of *would have*, and lack of subject-verb agreement. The next low-seriousness errors bothered only a few people: *different than* instead of *different from* and omitting the apostrophe in the contraction *it's* (op. cit.: 797).

Business people also evaluate language accuracy against global measures. While there is only one situation when no kind of error is allowed,

that is, in applications for a job, "the theme that dominated the written comments was professionals' concern for content; they care even more about clarity and economy than they do about surface features" (Hairston 1981: 798). One senior vice president of a computer company wrote that the difference between the winners and the "also-rans" at the top levels of business was the ability to communicate effectively. Another person hated "literary acrobatics, no matter how grammatical, that tend to obscure meaning." Only one bank president wrote, "I don't let such things bother me."

Beason (2001) conducted a qualitative in-depth study of how errors affect the "ethos" of the writer by submitting a questionnaire and interviewing fourteen business people. He claims that

> errors must be defined not just as textual features breaking handbook rules but as mental events taking place outside the immediate text. Defining error as simply a textual matter fails to forefront the "outside" consequences of error, especially the ways in which readers use errors to make judgments about more than the text itself (op. cit.: 35).

Giving an overview of the previous studies, Beason mentions that syntactic errors such as fragments (incomplete sentences not containing one main clause) and fused sentences (also called run-on sentences – two main clauses not separated by punctuation) are deemed to be especially bothersome. Misspellings fall in the middle range, although the results are not consistent. Some studies claim that homophone errors do not particularly stand out; others found that *you're [your]* was extremely bothersome while such misspellings as *recieve* were only mildly irritating. In the study by Wall and Hull (1989) without predefined categories, only 1.4% of responses referred to misspellings.

In Beason's investigation (2001: 41), five error types were evaluated on a 4-point scale from "not bothersome at all" (1) to "extremely bothersome" (4). Fragments confirmed to be the most irritating errors with a mean score of 3.00. With all the expected 'mildness' of spelling errors, they appeared second most serious (2.70), followed by word-ending errors (2.59). The other syntactic error, fused sentences, was judged to be less severe (2.48). The least bothersome error identified by the participants was unnecessary quotation marks (2.30).

Spelling mistakes were also included in my stimuli discussed in Chapter 5, therefore, it is interesting to zoom in to the distribution of the scores within the spelling category in Beason (2001: 42).

metods (3.00)
they're [their] (2.93), *aboutt* (2.93)
recomendations (1.93)

The author describes *recomendations* as a "misspelling many people might easily produce or overlook" (op. cit.: 40), *they're* as a homophone error, *aboutt* and *metods* as "two misspellings so glaring they could be typographical". It is noteworthy that only one missing letter is a common feature in the least and the most irritating misspellings. The difference is that, in *recomendations,* the deleted letter is the missing double in a long word, whereas in *metods,* it is a single letter in a short word that would, of course, be critical for word recognition.

Beason's study is particularly valuable because of the opinions of his subjects expressed in the course of an interview. The errors were inserted in a business document, so the business professionals could well relate to them.

> The subjects frequently accounted for even the most negative scores *not* by discussing their confusion as readers, but by commenting on the image the error creates of the writer. […]
>
> The interviews suggest, in fact, that the extent to which errors harm the writer's image is more serious and far-reaching than many students and teachers might realize. At times, the subjects stated in very general terms that errors affect a person's credibility as a writer or employee (op. cit.: 48).

The following characteristics were attributed to the writer based on the errors:

- hasty (readers sympathize about time pressure but are still bothered),
- careless (one subject refused the services of student interns whose errors indicated to her that they would not proofread carefully),
- uncaring (shows inappropriate attitude to writing of documents, unimportance to the writer, no respect to the reader),
- uninformed (no knowledge of conventions for formal English).

While some of the subjects were lenient with what they deemed accidents, others viewed them as more bothersome because the writer, in the reader's estimation, essentially decided to ignore a problem that could have been easily fixed. "Thus, error gravity is not necessarily determined by whether an error is perceived as an accident rather than a knowledge problem" (op. cit.: 52).

Errors can affect the image as a business person. The writer may be assumed to be a faulty thinker (particularly syntactic errors such as fragments and fused sentences indicate limited reasoning skills), not a detail person (someone who makes an error in writing a word is likely to make an error in writing a number, which is not acceptable in working with money), a poor oral communicator, a poorly educated person (resulting from the writer's inability to learn rather than ineffectual teaching), or a sarcastic, pretentious, aggressive

writer (the impression based on the use of unnecessary quotation marks to emphasize words).

Moreover, the readers extrapolated the effects of errors from the individual writer to the group of people he or she might represent. Errors harm the company's image of which the writer is a poor representative. A company with such an employee may lose customers. If the company has to go into court, the opposing attorney will take advantage of the errors and turn them against the company by making its bad portrait for the judge and the jury.

The participants of the study not only shared their own negative impressions caused by the errors but also guessed how other people could be bothered by them. "This guesswork might partly account for the diverse reactions subjects often had to the same errors" (op. cit.: 57). The interviewees stated that they, in fact, would not make certain judgments, but they know other people would.

In the end, Beason encourages teachers to explain to the students that the variation of reactions to errors is so large that an error that seems minor and benign to some people may be quite offensive to others. Consequently, all errors are worth attention, and not only those impeding the meaning. The writer's ethos can be endangered by errors. Teachers may be lenient to minor errors, but in the non-academic community, errors can affect people more severely and bring negative consequences. Many mistakes can be eliminated by proofreading.

In a recent study by Brandenburg (2015), the participants were asked to read a text in a web-based survey: one version with six errors and one correct control. In the first version, the readers only noticed two errors. However, the ranking for the writer was statistically significant between the two versions, "suggesting that the presence of errors can affect the writer's ethos" (op. cit.: 74).

The conclusion can be made that an error is not only an infringement of textbook rules; it contributes to the image of the writer. Errors bother to a different degree, but the main thing is that they do.

2.4 Summary

Studies of errors in applied linguistics based on learner data dispersed the beliefs inspired by behaviourist theories that errors are mainly caused by transfer from the native language. It appeared that most errors are common for first and second language acquisition and reflect the developmental steps towards the target language. Learners work out and test their hypotheses about the structure of the language, so that the errors in their own right fit in the system at its stage of development.

Although there was a shift in the scholars' view of errors from an abuse of the standard variety to the full legitimization within the interlanguage, as far as perception is concerned, errors never lost their irritating impact. The effects of errors were investigated based on questionnaires. The studies aimed at discovering the differences in the ratings between particular subject groups and depending on the status of the error producer. It turned out that non-native speaking teachers rate most strictly, while native speaking non-teachers are most lenient. An important finding that relates to the analysis in Chapter 5 is that "NNS [non-native speaker] judges seem to be especially hard on morphological and functor errors in comparison to NS [native speaker] judges. However, they tend to evaluate lexical and global errors less severely than NS judges" (R. Ellis 2008: 57).

The gradable effect or errors inspired multiple surveys on error gravity and an urge to build a hierarchy of error types to be used as reference in language teaching. The results turned out to be so divergent and incompatible that a universal hierarchy did not seem to be possible. This fact sets a demand for the search of the driving forces behind error gravity. However, researchers with a pedagogical focus have been content with eliciting opinions about errors, i.e. exploring their social meaning. Regardless of the causes, most deviations from the standard norm of written English trigger a negative reaction which should not be underestimated.

3. A Psycholinguistic Approach to Error Processing

Errors irritate not only because they infringe grammar rules or social norms but also because they interfere with **processing**. Some researchers investigated correlations of error ratings in questionnaires with **factors** underlying the psycholinguistic mechanisms of error perception. Comparing the reactions of native and non-native speakers, it is necessary to find out to what extent differences in the **cognitive organization of the L1 and L2** may affect error processing. Psycholinguists working with **real-time methods**, i.e. self-paced reading or **eye-tracking**, conducted experiments on **word recognition** that can be extrapolated to the perception of local errors. Recent psycholinguistic and neurolinguistic studies focused on **cognitive processes** involved in reading, including effects of **bilingual processing**, sensitivity to grammatical violations and differences related to the task. There are also effects of accent, **error rate** and **error type** on the probability of **internal repairs** while processing implausible utterances.

3.1 Attempts to predict error gravity

There have been attempts to figure out factors determining error gravity independent of the status of the reader and the writer. Although *error gravity* is by no means a psycholinguistic concept, some researchers within the Error Analysis framework discussed the perception of errors in cognitive terms, trying to predict it based on factors involved in language processing. In these studies, the data were collected by means of questionnaires eliciting the overall evaluation of the text after the processing had been completed. As an off-line method, a questionnaire does not have the advantages of modern on-line techniques such as eye-tracking or brain imaging, which allow to extract the readers' reaction during the task. Nonetheless, the following attempts to explain the ratings by processing constraints deserve attention.

James (1998) suggests a criterion of the **rule range**, predicting a correlation between the error gravity (EG) and the proportion of the sentence that is distorted.

> This criterion is observed by teachers marking written work, who are more severe when they have to underline or strike out a whole clause rather than just a single word or morpheme. This reaction

> might be linked to a desire to quantify EG (in terms of the length of a red-ink line), but could more plausibly be related to the different degrees of the processing effort needed on the part of the reader or listener to undo the error (James 1998: 208).

Another hypothesized factor is the **word frequency** as used by native speakers: an error in a more frequent item is more serious than an error in a less frequent one (op. cit.: 210, no references to studies have been made). The frequency of committed errors – error **density** – has been found to correlate with error gravity. Zola (1984) reported that even minor spelling errors in highly predictable words disrupted reading when they occurred too often.

Santos (1987) investigated the reactions of native speakers to learner errors based on the **markedness** theory of the Prague School linguists. "The marked member contains at least one more feature, morpheme, or rule than its unmarked counterpart, and its contextual distribution is specified. The unmarked member has an unspecified and thus considerably wider range of distribution" (op. cit.: 208). For example, the form of the indefinite article *a* is unmarked, whereas the form *an* is marked, i.e. more complex from the linguistic and psycholinguistic perspectives, being restricted to the use before a noun beginning with a vowel sound. Santos hypothesized that errors "reflecting the unmarked-to-marked direction will arouse a greater degree of irritation" (ibid.). For example, *With an great effort I sent him inside* (*a -> an*) would be more irritating than *Such a event happened to me* (*an -> a*) because the unmarked form *a* is generally more expected. Similarly, *I believes* should be rated as more serious than *He believe* (1st/3rd-person singular ending), just as *We took a vacations* more serious than *Life consists of dramatic event* (singular/plural NP).

As a result of the experiment, Santos found a significant effect of markedness for three of the five error types, which supported his hypothesis in the cases of the singular/plural NP, *a/an* and the infinitive/gerund errors (e.g. *I asked him coming over* was more irritating than *I asked him to come over, not to know my brother would come home early*). However, there was no effect for errors involving 1st/3rd-person singular endings and the active/passive constructions. The latter error type was rated lower than the others in both conditions (e.g. *Humor is consisted of events that happen out of their ordinary context* and *It forbids to drive faster than 65 mph*). The analysis of the marked and unmarked categories together revealed that syntactic errors were rated significantly lower than morphological ones. Besides, the syntactic errors were "overlooked" only a single time, whereas the morphological ones – fifteen times. The least detected errors were the "plural made singular" (i.e. deletion of the *-s*) and the "*an* made *a*" (i.e. deletion of the *-n*).

An important factor employed for the explanation of the seriousness of grammatical errors is **rule generality**.

> If grammar is more general and predictable, and lexis more idiosyncratic and fine-grained, […] surely it implies that grammar errors are more serious *linguistically* than lexis errors. […] The **passive** rule, for instance, applies to all transitive clauses, while the oddity of *We fried some *milk* derives from the incompatibility of a particular verb with a particular noun object (James 1998: 207).

However, if one postulates the ultimate value of meaning and sentence comprehensibility, the error gravity scale turns upside down. Olsson (1977) carried out an intelligibility experiment consisting of a 419-word text with 24 inserted grammatical and lexical errors. In the form of a cloze test, it was distributed to secondary school students (total 371) in England, Scotland and the USA who were to fill in 40 blanks of the text. Olsson concluded that the degree of intelligibility of erroneous utterances was quite high, although it is influenced by the degree of deviance and relative plausibility of competing interpretations. But grammatical errors impair intelligibility to a lesser extent than semantic errors (cf. Olsson 1972).

> You can achieve communication without grammar, so wrong grammar is not serious. Relating this to noticeability, we must conclude that in life-or-death situations we do not bother to notice grammar errors (James 1998: 219).

Gass and Selinker (2008) discuss a similar problem in terms of the **range of choices**.

> There is a more limited number of grammatical possibilities (or grammar rules) in language than there are vocabulary items or possible pronunciations. That is, if a learner fails to mark agreement or puts items in the wrong order, there is a greater likelihood that an NS [native speaker] can fall back on his or her grammatical knowledge to make sense of what a learner is saying. However, if a learner uses an inappropriate or nonexistent vocabulary item, the NS may be sent down a comprehension path from which there is little possibility of return (op. cit.: 312).

Thus, potentially, the rules of grammar can be applied to more items, and their violation affects the language system more deeply than a misuse of an individual word. However, practically, an error is most often constrained, the number of competing solutions is limited, and grammar offers a smaller paradigm of forms that could fit. In natural language, the preceding context, not necessarily verbal, gives some clues as to what should be expected. The stronger these constraints are, the smaller is the competition and the easier is the resolution. I believe, both grammar and lexical errors have equal gravity

potential because both can be highly or low constrained, and only by analyzing the probabilities of misinterpretations in particular cases, one can predict the kind of error that would be judged as more serious.

The concentration of some researchers on the formal characteristics of errors led Rifkin and Roberts (1995) to questioning the validity of the results. "By continuing to ignore the variability introduced by phenomena such as stereotyping or content, and by focusing only on linguistic detail, error gravity research has privileged a conception of error as a purely linguistic phenomenon" (op. cit.: 513). As shown in the previous chapter, there were other researchers who analyzed errors in very different contexts, focusing on the qualitative descriptions where the adherence to formal variables would have been beneficial.

3.2 Processing concerns underlying second language acquisition

After the behaviourist approach to the causes of errors did not find support in the learner data, i.e. the predominance of transfer was disproved, researchers focused on the role of mental processes in language acquisition. They discovered some common features for first and second language acquisition, namely that they go through the same developmental stages, but at a different rate (Gass and Selinker 2008: 37). Learners create systematicity and develop rules which do not necessarily correspond to the adult (or target) language. Both first and second language learners make use of the overgeneralization of grammatical morphemes, and correcting their errors is not always effective.

In spite of the similarities between first and second language acquisition, there are also differences, and some of them are fundamental. "In normal situations, children always reach a state of "complete" knowledge of their native language" (op. cit.: 164), whereas adult second language learners usually attain only a transitional stage, an interlanguage. Another difference is the outset of the language learning: non-native speakers have specifications of the situation available for the form, while native speakers have to figure it out. Children build up a language system, whereas adult learners already have a complete system of their L1.

The differences in the mental system for the native and non-native language are reflected in Krashen's distinction between acquisition (a subconscious process, usually proceeding in a natural environment) and learning (a conscious process, typical of the classroom setting). This explains why errors are perceived differently in L1 and L2.

> We are generally not consciously aware of the rules of the language we have acquired. Instead, we have a "feel" for correctness. Grammatical sentences "sound" right, or "feel" right, and errors feel wrong, even if we do not consciously know what rule was violated (Krashen 1982: 10).

In the learned system, by contrast, the correctness of the utterance is checked against the knowledge of the rules. This is how error detection is mostly carried out in L2.

The idea of a sequential acquisition/learning (some linguists do not observe this distinction) of language structures was developed by Pienemann (1998) and became known as the Processability Theory. Its major premise is that "at any stage of development the learner can produce and comprehend only those second language (L2) linguistic forms which the current state of the language processor can handle" (Pienemann 2013). The developmental trajectories go across two dimensions: stages of development and interlanguages. "In this paradigm, each stage represents a set of grammatical rules that share certain processing routines, and each interlanguage variety represents a specific variant of the grammatical rules" (Pienemann 2008: 10). It can be concluded that errors are made due to the unavailability of the relevant processing resources at a given stage of the second language development.

The phenomenon of language transfer is also explained in cognitive terms. It has been suggested that the native language establishes a set of associations with their particular strengths. "These associations can possibly interfere with the establishment of an L2 network" (Gass and Selinker 2008: 221). In other words, a second language learner has to reset the associations according to the target language, and before this process is complete, the learner relies on his L1 network, which facilitates the transfer.

However, VanPatten and Keating (2007) found that, in processing tense forms, learners begin with universal processing principles and only then resort to the processing strategy characteristic of their L1. The transition to the strategies of the target language is achieved as the learner becomes more proficient. The universal strategy is the reliance on the semantic rather than grammatical information. For example, in VanPatten and Keating's study, learners relied on adverbs while interpreting tense.

As for the processing strategies, learners need to figure out the relationships between words in a sentence (e.g. which element will fill the subject slot), while the cues for this decision may be different in L1 and L2. One way of operationalizing this problem is offered by the Competition Model, "a functionalist approach to grammar that accounts in a principled way for

probabilistic outcomes and differential 'weights' among competing and converging sources of information in sentence processing" (Bates et al. 1982). This model suggests that:

> [l]anguage processing involves competition among various cues, each of which contributes to a different resolution in sentence interpretation. Although the range of cues is universal (i.e., the limits on the kinds of cues one uses are universally imposed), there is language-specific instantiation of cues and language-specific strength assigned to cues (Gass and Selinker 2008: 221).

A "breakdown" in the normal use of cues, i.e. an error, results in a competition between the alternatives and a decision which cue is preferred. For instance, to determine which element fills the subject slot in the sentence *The grass eats the cows*, one can base the interpretation on the word order (a very strong cue for native speakers of English), rely on morphology (agreement with the verb), or trust the meaning and animacy (i.e. the interpretation that the cows eat the grass). It has been demonstrated that when the direct interpretation is not plausible, meaning-based cues are generally preferred to word order. As the L2 proficiency increases, learners gradually adopt the appropriate cues of the target language.

Another factor causing differences in language processing by native and non-native speakers is working memory. Its function is associated with the temporary storage and processing of incoming information needed to execute complex cognitive actions (Sagarra 2013). It is well-known that working memory capacity is limited and varies across individuals. Learning a foreign language after puberty is a challenge for the cognitive resources because it requires additional computation and activation compared to a monolingual or an early bilingual mind. Domain-specific models (Just and Carpenter 1992, Baddeley 2007) claim that L2 learning is constrained by the limitations of the working memory capacity, whereas domain-free connectionist models explain the limitations by the general ability to control attention in the face of distraction (cf. Sagarra 2013). Working memory capacity is measured in reading span tests in which subjects are asked to recall increasingly longer items.

The relationship between the working memory and proficiency in the L2 is reciprocal: a large memory span is both a prerequisite for and a consequence of successful language acquisition. On the one hand, a high phonological short-term memory capacity facilitates learning a foreign language (cf. Service 2013). It has been shown that working memory plays a role in error detection. In Sagarra's (2008) study, only higher-span beginning learners were sensitive to tense agreement violations. On the other hand, higher proficiency results in a

larger memory span. For instance, in the study by Gass, Roots and Lee (2006), there was a significantly stronger correlation between L1 and L2 working memory measures for advanced speakers than for low proficiency learners. The release of the working memory capacity in advanced learners possibly occurs due to the automatization of the processes that require metalinguistic control in beginning learners.

L2 learners are generally slower than native speakers in performing the same tasks. What kind of processes occupy the learner's working memory that produce the slowdowns while reading in L2? As functional neuroimaging studies with bilinguals have shown, linguistic tasks involve not only the area storing representations but also activate other areas responsible for working memory, response inhibition, response selection, decision making, language selection and switching, conflict monitoring, attention, and error detection (Abutalebi 2011: 125). The engagement of these areas takes place to a greater degree in L2 processing. Similar mechanisms were described from the linguistic point of view by Krashen (1982) and are known as part of his monitor model.

Indefrey (2006), who analyzed 30 neuroimaging studies, argues that processing in L1 and L2 involves the same areas in the brain but with a different strength of activation. "Bilingual speakers with late L2 onset, low L2 proficiency, or little L2 exposure tend to show stronger brain activation for processing the L2 than the L1" (Bultena and Dijkstra 2013: 1). This indicates that the same brain areas are involved in processing L1 and L2 independently of age of acquisition and proficiency. A number of studies have demonstrated that, over the course of language acquisition, the brain undergoes qualitative changes and that, with increased proficiency in the L2, responses to morphosyntactic violations become more native-like (Roberts et al. 2016: 10).

There are three observations drawn from the studies on the bilingual lexicon (cf. Kroll and Bogulski 2013): (1) the L1 is active while learners process words in their L2, (2) L1 activity persists even at a high level of proficiency in the L2, and (3) in proficient learners, the L2 is activated when words are processed in L1. "The persistence of cross-language interactions even in the face of structurally distinct languages suggests that the lexicon is fundamentally an open system at a relatively abstract level" (op. cit.: 4). It has also been shown that even when bilinguals intend to produce an utterance only in one language, with the presence of rich context in that language, there is still activation of the other language not in use.

3.3. Introduction to eye movement research

The study of language processing largely relies on eye movement data. For a comprehensive review of the most prominent studies and findings involving eye movement experiments see Clifton et al. 2007. The history of eye-tracking research goes back to the 19th century and can be divided into three stages. The first is the study of how eyes move in reading, the second deals with word recognition, and the third stage (to the present day) focuses on cognitive processes behind reading.

Eye movement research is based on the "eye-mind hypothesis" (Just and Carpenter 1980) arguing for the immediacy of interpretation, i.e. that participants are processing exactly the object they are looking at for as long as their fixations last.

To study language processes, cognitive psychologists not only examined the normal reading but also used various techniques and unusual tasks, e.g. pressing a button to display every new word or to decide whether it is an existing word. Some experiments involved changes on the screen while presenting the language material. In spite of the very interesting findings, it was acknowledged that "all of these tasks are somewhat unnatural in relation to the normal silent reading process (i.e. readers do not normally do such things)" (Rayner et al. 1989).

As a result of numerous eye-tracking studies, different models of reading have been developed. All theories are based on the assumption that processing difficulty positively correlates with fixation duration. "It is clear that when text is difficult, readers fixate longer, move their eyes a shorter distance with each saccade, and make more regressions" (Frazier and Rayner 1982).

Here are some basic facts about reading and, specifically, findings on silent reading in alphabetical left-to-right languages like English. When viewing a text, "the eyes are either relatively stable (fixating) or moving between fixations (saccading). When the eyes are moving, we still experience perceptual stability" (Shillcock 2007: 89). There is a false impression that the eyes are moving smoothly over the page – as a matter of fact, they jerk from one fixation point to the next, and a reader receives meaningful input only while the eyes fixate between each saccadic movement. Figure 3.1 shows an example of reading patterns for two randomly chosen sentences from the eye-tracking experiment discussed in this thesis (extracted from Data Viewer). The circles represent fixations, and the arrows point to the direction of saccades. The larger size of the circles corresponds to the longer fixation duration indicated by the number in milliseconds. The frame (in pink) assigns the fixations to the corresponding words; the words are then grouped into regions.

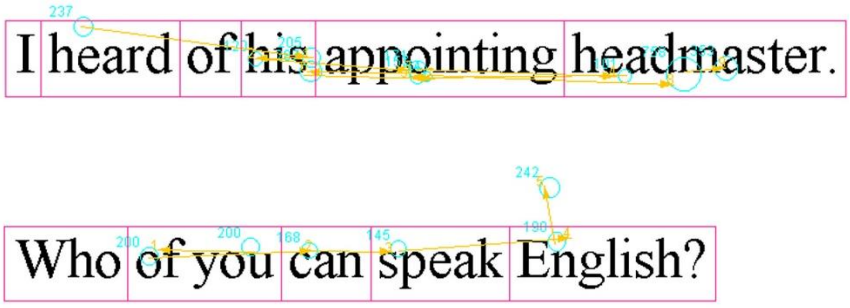


Figure 3.1. *Reading samples illustrating fixations and saccades*

Skilled readers make saccades of about 7-9 letter spaces and fixate individual words for about 200-250 msec. Long words are usually fixated more than once, while shorter words, especially frequent function words, can be skipped altogether. "[W]ords that are highly predictable – given the preceding sentential or textual context – are skipped more frequently and are characterized by shorter fixation durations than words that are less constrained by the preceding context" (Roberts and Siyanova-Chanturia 2013: 215).

If the preceding text causes difficulties and demands more reading time, it may affect the processing of the target region.

> [The] results indicate that there is immediacy of processing — that the reader begins processing a word on encountering it. Moreover, further analyses indicate that the reader generally finishes the processing of that word (as far as it can go) before reading the next word. If people did start to look at the next word while completing the processing of the previous word, there should be some influence of the properties of word N on the duration spent on word N + 1 or even later words. We will call this a "spill-over" [sic] effect (Just, Carpenter and Woolley 1982).

A spillover effect has been found, for example, for processing low-frequency words. It would also be relevant for processing errors. Therefore, it is common practice to analyze the reading times not only for the area of interest, but also for two regions (constituents) preceding and following the target.

One of the advantages of the eye-tracking technology is the possibility to separate the stages of processing by analyzing early and late time measures. The following measures were used in the present study:

(1) the *first pass reading time* (FPRT) – the sum of all fixations in a region when the region is read for the first time, e.g. the adverbial *lately* in Figure 3.2 (the first, second and third pass apply only to the region of interest; the fixations marked in red are included in the time measure illustrated by the figure):

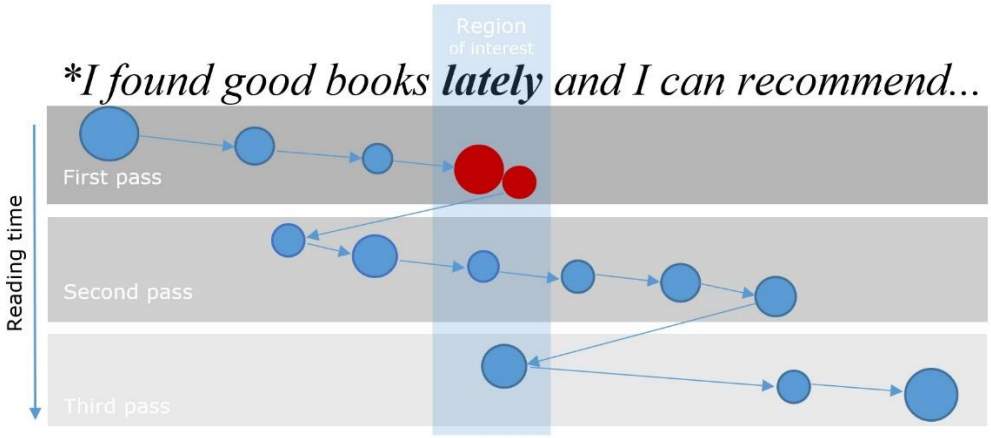


Figure 3.2. *Example of calculating the first pass reading time*

(2) the *regression path duration* (RPD) – the sum of all fixations in a region and the regressions to the preceding words prior to moving on to the next word to the right (Konieczny et al. 1997),

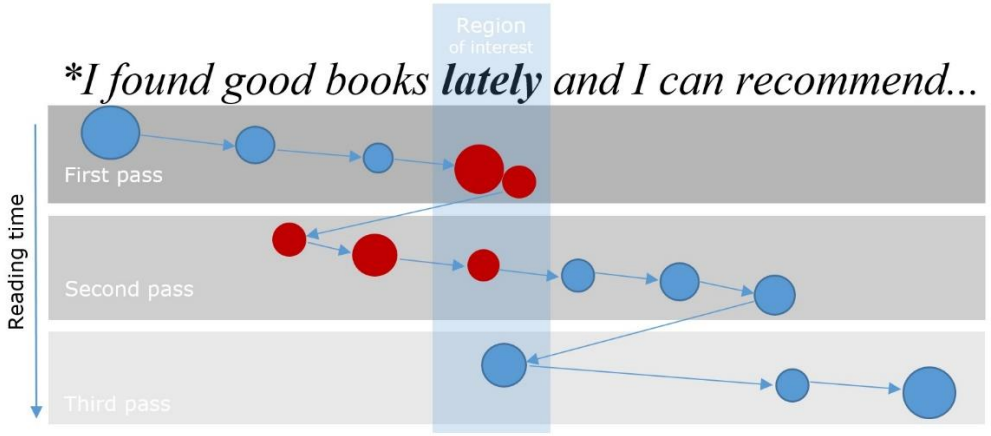


Figure 3.3. *Example of calculating the regression path duration*

(3) the *total reading time* (TRT) – the sum of all fixations in a region including the first pass, the second (third, etc.) pass and regressions to it.

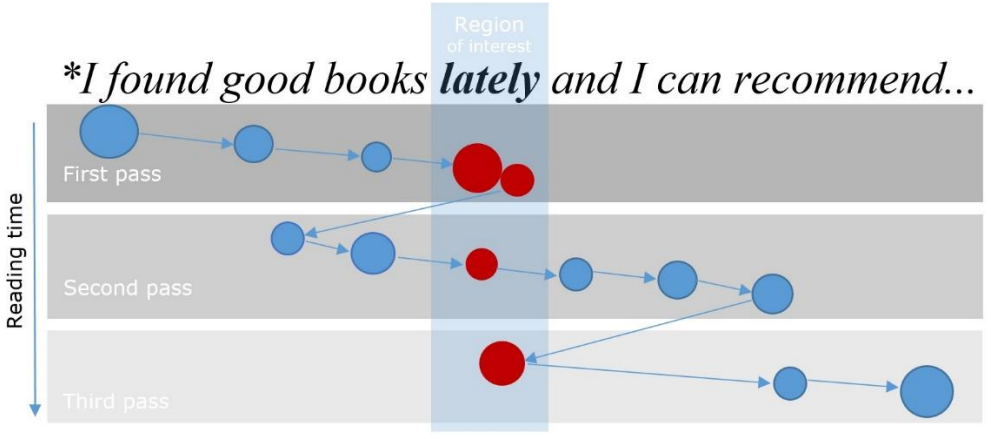


Figure 3.4. *Example of calculating the total reading time*

In order to obtain a complete picture, both early and late measures need to be analyzed. Early measures are found to be sensitive to lexical access and early integration of information, whereas late measures are associated with reanalysis, discourse integration and recovery from processing difficulties (cf. Roberts and Siyanova-Chanturia 2013: 217). Besides duration measures, one can also use fixation counts and the probability of regressions, which also reflect an increased processing load.

3.4 Word recognition and its implication for error processing

Early psycholinguistic models of reading can roughly be divided into those which accept phonological mediation in reading and those which do not. The debate has been whether the word meaning is computed directly from the orthography, and the retrieval of the phonological information follows it (e.g. Seidenberg 1985), or whether "phonetic activation nonoptionally occurs (prelexically) during lexical access" (Perfetti et al. 1988). The dual-route model of Patterson et al. (1985) seems to have found a compromise. It posited that known words are recognized as wholes, on the bases of their shapes (logographic reading), and that, in skilled readers, the visual logogens directly activate the meaning. Unfamiliar words are read by means of the spelling-sound rules, which presupposes phonological decoding. There is strong evidence that readers appeal to phonology of a written text even when reading silently. "Readers also appear to generate phonology for nonwords even when the task does not explicitly require it" (Gathercole and Baddeley 1993: 178).

Many studies have supported the phonological mediation hypothesis (e.g. Van Orden 1987). Lesch and Pollatsek (1993) tested the priming effect of *bare* followed by the target *lion* that is semantically related to a homophone of the prime (*bear*). With a 50-ms exposure, the target was related to the prime homophone. With a 200-ms exposure, however, this priming disappeared. This effect was taken as evidence for the phonological activation of meaning, and besides, it showed that the longer exposure allows sufficient time for a proper spelling check.

> It is now well established that the recovery of phonological structure is a mandatory phase of print processing, and that a phonological code is used as a routine procedure for lexical access and for accessing meaning (Frost and Ziegler 2007: 108).

One of the most convincing arguments for the direct visual access to meaning is the "word superiority effect" which showed that letters can be more

accurately recognized in the context of a word than in isolation (Reicher 1969). For example, subjects are more accurate at recognizing *D* in the context of *WORD* than in the context of *ORWD*. It has often been interpreted as evidence that word recognition involves the use of word shape information rather than recognition of the letters of which the word is composed, as non-words can only be read letter by letter. McClelland and Johnston (1977) demonstrated that letters are identified faster in the context of pseudowords (pronounceable strings of letters sharing characteristics of legal words, e.g. *mave*) than in non-words (letter strings violating the spelling rules of the language, e.g. *amve*). It indicates that the reason for the word superiority effect is not the recognition of familiar word shapes, but rather the existence of regular letter combinations (Larson 2004).

The effect of the word shape was in fact elicited a few times, but still remained controversial. Haber and Schindler (1981) found that misspellings that changed the overall shape of a word were more likely to be detected than misspellings consistent with word shape. This effect was larger for function than for content words (of equal length), suggesting that function words are more often identified holistically due to their high frequency and predictability. Similarly, Monk and Hulme (1983) found that alternations to word shape were noticed more often than alternations preserving word shape. At the same time, Abramovici (1983) confirmed that errors were more readily detected in content words than in function words but could not replicate that errors changing the overall word shape were detected more easily than errors preserving word shape. In addition, no evidence was found that the position of the error in the word affected its rate of detection. Paap, Newsome and Noel (1984) tested the relative contribution of word shape and letter shape and found that the entire effect is rather driven by letter shape. Beech and Mayall (2005) demonstrated that readers are more accurate at recognizing words based on the external letter features (lines inside the letters removed) than internal letter features (outer lines removed), which indicates priority of the word contour.

Some studies suggest that word recognition is based on the analysis of letters. Although it has been shown that THE EYE RAEDS PEFRECLTY WEHN THE WRODS ARE WIRETTN IN SCUH A WAY (e.g. Johnson, Perea and Rayner 2007), which evidences against the model of letter-slot coding, the reason why we recognize words with letter transpositions is that they "are perceived as being very similar to their base words" (Rastle 2007: 72).

The neural network models of reading by Seidenberg and McClelland (1989) and Plaut et al. (1996) presented word recognition processes as simultaneous decoding of the constituent letters in their positions (which all together contributes to the word shape). The model "starts out with no

knowledge about the relationship between letters and pronunciations, only that letters and sounds exist" (Larson 2004). After a few rounds of training, the machine can read a few high-frequency regular words. After many rounds of training, the model will be able to read not only words it has seen before, but unfamiliar words as well.

There are a few factors facilitating processing, e.g. frequency and age of acquisition. "The time taken to recognize a word is reduced both when that word has a high printed frequency and when that word was acquired early in life" (Rastle 2007: 75).

Thus, word recognition studies have discovered a set of processing strategies and experimentally confirmed them for different tasks (McNorgan et al. 2015). The strategies do not necessarily contradict, but rather complement each other. Depending on the task and quality of the input, readers seem to be switching between a few modes of word recognition. They can be depicted as a continuum (Figure 3.5) from most primitive serial letter-to-sound mapping (children learning to read, processing non-words, rare words with irregular spelling, foreignisms) to most speedy holistic retrieval (function words). Experienced readers recourse to the holistic visual decoding if the word is short, frequent and predictable enough. Otherwise, the recognition mode is shifted to the left, and the processing requires letter shape evaluation. If the graphic image is not easily mapped onto a phonetic representation, the mode is shifted to the serial letter-to-phoneme analysis.

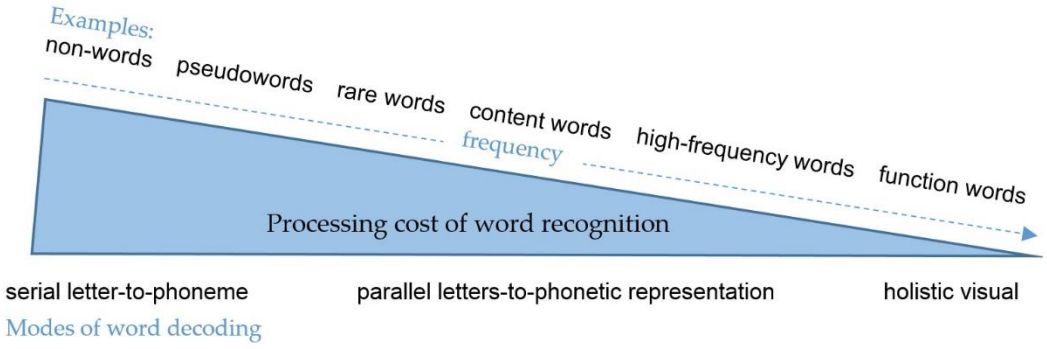


Figure 3.5. *The word recognition continuum*

Thus, the perception of a text proceeds sequentially as a word-by-word verification with the corresponding units stored in memory. There is evidence for the different kinds of these mental units – it depends on how one imagines the representations of phonology, orthography and meaning, and the relations between them. In some cases, the verification of the word spelling is carried out based on the 'snapshots' of the visual memory processed through feature analysis and modified into abstract letter/word identities. However, in most

cases, the word spelling is verified by means of phonological mediation and possibly semantic control, i.e. by circulating through the whole chain of the word representation (as argued by connectionist models).

Drawing on the findings of psycholinguistic experiments, it can be hypothesized how native speakers process texts containing errors. In particular, the models can explain how readers deal with spelling mistakes (letter transpositions, omissions or substitutions). That is, having seen familiar words as whole units in the beginning, readers proceed by analysing the individual letters and their order, verifying them with the corresponding orthographic representation. "*Given sufficient time*, skilled readers can decide with a remarkable degree of accuracy (perhaps 100 percent) whether a visually presented letter string (e.g. BALSE, FALSE) is a known word or a non-word" (Rastle 2007: 73, her emphasis). Thus, readers are quite good at identifying misspellings, but only when they are given enough time for processing. Should they be reading in a hurry, there is a probability for spelling mistakes to be overlooked.

An error may also escape from the reader if it is not directly fixated. It has been shown that the position of fixations is determined by the length of words and the perceptual span.

> The effective processing window thus defined for English is typically three or four letters left of fixation and fourteen or fifteen letters right of fixation, but with accurate recognition of only seven or eight letters to the right (Shillcock 2007: 96).

It implies that seven or eight characters on the right side of the perceptual span are not seen as clearly as those to the left. Hypothetically, errors falling on the right side of the perceptual span have a chance to remain unnoticed and, as a result, cause no processing difficulty to the reader. This might be relevant for local errors, e.g. spelling mistakes, wrong word formations or inflections, i.e. errors occurring within a few letters. Taking the misspelled or ill-formed word for the correct one would be more likely if these words are orthographically similar, on condition that the reader has a clear understanding of the contents and, because of high expectations, automatically repairs the word containing the error. However, due to the preview of the area to the right of the fixation, the likelihood of skipping an error is small.

It seems to be almost impossible to predict which word or part of the word may fall in the low-focus area, but there is a pattern.

> Longer words have a more marked OVP [optimal viewing position] curve, although performance at the OVP itself is comparable for shorter and longer words; for long words there is a sharper drop

> from the OVP to the poorest processing, which occurs at the right end of the word (Nazir 2000; cited by Shillcock 2007: 91).

The implication for error detection is that readers pay less attention to the end of the word, e.g. wrong inflectional morphemes. Furthermore, skipping is especially frequent when the text reads smoothly.

> Depending on the text, the task, and the language, around one third of words may not receive a direct fixation, yet they contribute to the reader's understanding. A word may be skipped in part simply because it is short and/or occurs close to the current fixation (op. cit.: 97)

Thus, if the reader skims through the text, carried by the flow and picking up only the important information, s/he may skip short function words containing an error and fail to notice wrong endings falling on the weak part of the visual span.

3.5 Real-time processing in the non-native language

A critical debate in L2 processing research has been whether the bilingual mental lexicon is language selective or nonselective.

> One possibility is that lexical representations from each language are stored and accessed separately (i.e., lexical processing is language selective). However, most recent research converges on the idea that words from both languages are integrated into one lexicon; that is, the bilingual mental lexicon appears to be language nonselective (Roberts and Siyanova-Chanturia 2013: 221).

The fact that words from different languages are stored in a common database is also supported by neuroimaging studies (cf. Bultena and Dijkstra 2013). It has been demonstrated that processing a word in one language activates representations from another language. Part of the evidence for nonselectivity comes from the cognate facilitation effect. Words of different languages that share the same historical origin are recognized and produced faster than noncognates. Presumably, reading a word in one language involves activation at the orthographic, phonological and semantic levels, which spreads to the representations of both languages and thus facilitates the reaction. The cognate facilitation effect was elicited even in unilingual sentences (Duyck et al. 2007), which suggests that "the top-down cue of language may not be enough to completely inhibit activation of the bilingual's other language, at least in low-constraining sentence contexts" (Roberts and Siyanova-Chanturia 2013: 223). Cognates also differ from matched control words in the produced brain waves,

e.g. the N400 component, marking semantic integration (cf. Bultena and Dijkstra 2013: 2).

Unlike cognates that are processed faster, words that have the same orthographic form but a different meaning (i.e. interlingual homographs or "false friends") are processed slower. It is assumed that the lexical activation of the forms in two languages does not converge on the same meaning and produces an interference effect.

The influence of the context is shown to be an important factor in processing language anomalies at a sentence level, i.e. lower vs. higher semantic constraints.

> [M]any studies that employ translation and word recognition tasks have found that crosslinguistic activation effects are eliminated when words are presented in highly constrained sentence contexts, at least with very proficient learners (e.g., Van Hell & De Groot, 2008; cited by Roberts and Siyanova-Chanturia 2013: 223).

It has been found that, although two languages share one mental lexicon, when the semantic context is highly constrained, any cross-language ambiguities are resolved during the later stages of comprehension processes.

Some inconsistencies in the results of individual experiments on L2 processing may be due to the different levels of proficiency of the participants, e.g. foreign language learners vs. members of bilingual communities. "It is therefore possible that cross-language activation effects are more strongly present in learners with less L2 proficiency or experience" (op. cit.: 224).

In Frenck-Mestre and Pynte's (1997) study, English-French bilinguals read structurally ambiguous sentences. The analysis demonstrated that both L1 and L2 readers had slowdowns on the ambiguous parts compared to the corresponding parts of non-ambiguous sentences. This finding, as well as results of other studies examining reactions to "garden-path" sentences (e.g. Juffs and Harrington 1996), indicate that processing in both L1 and L2 is incremental, that is, readers attempt to interpret and integrate the meaning into their current analysis word by word.

Some of the differences between processing in L1 and L2 are explained by the effect of transfer from the native language. The results of Frenck-Mestre and Pynte's (1997) experiment showed that "bilinguals hesitated when reading in their second language at points in the sentence where their native language presented conflicting lexical information" (op. cit.: 119). Roberts and Liszka (2013) examined the sensitivity of German and French learners of English to tense and aspect mismatch between a fronted temporal adverbial and the

inflected verb, e.g. *Last week, James has gone swimming every day*[3]. In a judgment task, both subject groups assessed the mismatching items as less acceptable than the controls. However, in a self-paced reading experiment, only the French L2 learners were sensitive to the mismatch conditions in both the Simple Past and the Present Perfect. The fact that the German learners of English did not show any processing cost for either mismatch type is explained by influences of the learners' first language: "only those whose L1 has grammaticized aspect (French) were sensitive to the tense/aspect violations on-line, and thus could be argued to have implicit knowledge of English tense/aspect distinctions" (Roberts and Liszka 2013: 413).

There are also studies that found no effect of the learner's L1 on the processing in L2 in such cases where the transfer would be expected (e.g. Havik et al. 2009, Papadopoulou and Clahsen 2003). The switching of the processing strategies of the learners, i.e. following their L1 cues or inhibiting them, is explained by learners' abilities to postpone online commitments to a particular analysis in the absence of lexical-semantic information (cf. Roberts 2013: 3). If there is enough input for plausibility and subcategorization, learners' processing is incremental and native-like. In case of the insufficient input, learners may suspend their analyses to prevent false interpretations in recovering from which L2 learners have more difficulties than native speakers.

Contrasting data from off-line judgment studies and real-time processing revealed some differences in the perceptions related to the task. For instance, Roberts, Gullberg and Indefrey (2008) elicited an effect of L1 during the paper-and-pencil task. When asked for their preferred referent for the subject pronoun, the German learners of Dutch patterned with the native speakers of Dutch, overwhelmingly choosing the most recent, sentence-internal referent for the pronoun, whereas the Turkish learners chose this referent only 55% of the time. However, in an eye-tracking experiment, both groups of L2 learners patterned together and differently from the native speakers. It was suggested that the differences in the results of the tasks are caused by the fact that L2 learners are having more difficulties than native speakers in the computing and integrating grammatical and discourse-pragmatic information at the same time, which is necessary for pronoun resolution. The integration of information from multiple sources during real-time comprehension causes processing slowdowns in L2 readers.

> [E]ven though learners may show the same patterns of interpretation as those of native speakers in tasks that measure

[3] Temporal adverbials locating the action wholly in the past, as *last week*, are normally used with the verb in the Simple Past (*went*), and not the Present Perfect. The use of verb forms with the past tense reference and temporal adverbials will be thoroughly discussed in Chapter 4.

> ultimate interpretation, they may have arrived by a different route: Thus investigating how L2 learners process the TL [target language] in real time can inform us of the nature of the human sentence processing mechanism in general (Roberts 2013: 7).

Eye-tracking experiments have also been used to investigate L2 learners' sensitivity to ungrammatical input during real-time comprehension. "Data from such research can inform one of the major debates in L2 acquisition: whether or not (or to what extent) it is possible to acquire nativelike grammatical knowledge if the L2 is learned after puberty" (Roberts and Siyanova-Chanturia 2013: 227). In a self-paced reading study by VanPatten, Keating and Leeser (2012), "non-advanced" learners of L2 Spanish were compared to native speakers of Spanish in the perception of grammatical violations. Both groups demonstrated sensitivity to the subject-verb inversion and adverb misplacement. However, only the native speakers reacted to the violations of subject-verb agreement in person and number. The fact that the learners were sensitive only to the syntactic errors but not to the morphological ones was taken as evidence for a "representational problem for morphology" in L2.

Researchers have been particularly interested in the sensitivity to the phenomena not instantiated in the learner's native language. For instance, in an eye-tracking study conducted by Keating (2009), adult learners of Spanish (L1 English) read sentences in Spanish with gender agreement violations. Besides varying the grammaticality, the critical constructions differed in distance between the NP and the modifying adjective. This tested the learners' ability to hold the relevant noun in working memory while processing the intervening material until the dependency is established. As a result, longer reading times for the mismatching condition indicated that "gender agreement is acquirable in adulthood […] and that the distance that separates nouns and adjectives affects the detection of gender anomalies in the second language" (op. cit.: 503). In particular, Keating found that advanced learners reacted as strongly as native speakers to the violations of gender only within adjacent words, i.e. at a close distance. Intermediate and beginning learners were not sensitive to gender violations at all, which indicates that the acquisition of abstract gender becomes successful towards higher proficiency. However, native speakers were sensitive to gender agreement violations across phrase boundaries and even across clause boundaries. The fact that non-native speakers' sensitivity is restricted only to local errors is assumed to be "due to deficits in processing, not underlying competence" (op. cit.: 529).

Keating's study supports the shallow structure hypothesis proposed by Clahsen and Felser (2006a and 2006b). Based on a series of psycholinguistic experiments, they argue that processing in L2 is fundamentally different from

processing in L1 in that non-native speakers do not compute a full syntactic analysis during sentence comprehension but, instead, rely more on lexical-semantic and pragmatic information. Native speakers, both children and adults, are guided to the same extent by the syntactic as well as lexical and pragmatic information. The "shallowness" refers to the distance separating the parts of the sentence that stand in a grammatical relationship. For example, the L1/L2 processing differences for morphological errors (noun plurals) were not as dramatic as the differences for syntactic dependencies or relative clause attachments. But more importantly, the shallowness refers to the deficits of deeper layers of analysis in L2. For example, to understand a sentence, non-native speakers process the predicate-argument structure based on the thematic roles or other lexical and pragmatic information, whereas native speakers compute representations containing hierarchical details and more abstract elements of the syntactic structure.

According to Randall (2007: 81), the higher frequency and longer fixation duration of non-native speakers indicates a lack of storage capacity in working memory due to (1) the necessity to pay attention (and thus store) both content and function words; (2) the need to consciously use syntactic rules to continually process the text; and (3) the lack of automatic word recognition strategies (the longer fixation times indicate the time spent decoding the words). "Thus, native speakers, through their knowledge of the language structure, are able to take in and 'chunk' more information at one time. One way this chunking and information load reduction is realised [sic] is through the knowledge of the syntactic structure of the language" (ibid.). This explains the view on reading as a "psycholinguistic guessing game" (Goodman 1967) in which native speakers have more advantages over non-native speakers. In his seminal paper, Goodman argues that reading is not a precise process involving exact, detailed perception and sequential identification of letters. Reading is rather a process of using syntactic and semantic knowledge to make successive predictions about the text and verifying them.

The distinction between these two modes of reading (precise vs. guessing) is reflected in the bottom-up and top-down models. "Top-down reading is related to a global procedure, easy for native speakers or advanced learners of a language, whereas bottom-up reading is related to analytical decoding, and it is driven by a process that results in meaning and proceeds from the units to the whole" (Luque-Agulló and González-Fernández 2012: 471). While reading in L2 is thought to be dominated by bottom-up processes, especially at a low proficiency level, more recent approaches allow an interaction between the data-driven and concept-driven reading modes.

3.6 'Error' perception studies in cognitive sciences

It has been shown in sociolinguistic studies that the information about the producer of the text affects the perception of grammatical violations in reading (e.g. matched-guise techniques discussed in Chapter 2). As for psycholinguistic studies, there is a temptation to think that, in a laboratory situation, we are free from the influences of the writer, i.e. the perception starts from zero, and the participant builds up his own image of the writer, if at all. This is not the case. The language of the experimenter functions as the prime and influences the perception. In an eye-tracking study by Konieczny, Hemforth and Sheepers (1994), there was an effect of the person who conducted the experiment. When the instructions were explained by a non-native speaker, an Irish learner of German speaking with an accent, the subjects repaired the target stimuli more readily, assuming that the low-frequency syntactic structure could be a learner error. When the experiment was conducted by a native speaker of German, the participants were holding on to the literal, though more complex, interpretations. This made the authors suggest a distinction between reanalyses and internal repairs, the latter being a "constraint-relaxation process on the incoming material, making it fit into the already interpreted structure" (ibid.). It appears that participants attribute the language material of the experiment to the person running the study. The feedback received during my eye-tracking experiment confirms that, too. For example, a few participants wondered whether it was I who wrote the sentences for the experiment, and when I said "no", they confirmed that they would not expect those kinds of errors from me.

The participants' readiness to repair erroneous utterances produced with a foreign accent was also elicited in a series of experiments reported by Gibson et al. (in press). They found that native speakers of English give foreign-accented speakers "the benefit of the doubt" and interpret the apparently implausible utterances in a plausible way (as mentioned in Chapter 1, cognitive scientists do not call them *errors*, therefore, this word in the title of this section appears in inverted commas). For example, a sentence like *The mother gave the candle the daughter,* in spite of the grammaticality of the double-object construction, is repaired into *The mother gave the candle to the daughter*. The results of the study are interpreted within the framework of the noisy channel model which I am going to introduce below.

Classical language comprehension models presume that the input is perfectly formed and clean. However, our everyday life is full of noise in all senses. There is noise in the environment: in the street where oral communication takes place or on the phone with possible voice distortion, the utterances are acoustically corrupted. In addition to that, the message itself

can be ambiguous; it may contain an error or a typographical mistake, or the comprehender may not focus his/her attention on the utterance and misunderstand it. The possible reasons why we are successful at communication at all are that the linguistic signal is redundant and that there are diverse information sources helping us infer the intended message (cf. Levy 2008).

> In a noisy channel model, the comprehender considers the probability of the utterance according to syntactic knowledge and according to semantic and discourse knowledge, and does error correction accordingly. The theory predicts non-literal interpretations of utterances when the veridical form of the utterance has low probability under either kind of knowledge (Futrell and Gibson 2016).

It has been assumed that human comprehension is driven by expectations of what is likely to be communicated and by the information on how the message may be corrupted by noise. The relevant linguistic and world knowledge, e.g. the frequency and probability of particular grammatical constructions and the plausibility of the meanings, may lead comprehenders to interpretations that differ from the literal meaning of the utterance. The likelihood of such interpretations and reliance on world knowledge increases with the perceived noise rate (Gibson et al. 2013). Applied to the present study, this would mean that the more deviations from the norm are detected in the text, the more likely it is that the reader feels free to repair the low frequent grammatical structures and correct implausible utterances into those fitting the reader's expectations.

In oral communication, it is likely that the accent of the L2 learner is associated with a high probability of errors. In the ERP study by Hanulíková et al. (2012), the P600 effect indicative of processing syntactic errors was reduced for accented speech. It suggests that errors in learner language are expected by default, and listeners do not react to them in the same way as when the speaker is native.

However, the comprehenders' readiness to make allowance for the noisy (accented) input and repair the utterance for a more plausible interpretation is limited. If the implausible sentence is too different from the plausible one, i.e. if the size of repair is too large, listeners trust their ears and stick to the literal interpretation, no matter how implausible it may be. This conclusion follows from the study of Gibson, Bergen and Piantadosi (2013) investigating the inferences for three syntactic alternations:

1) the double object (DO)/prepositional phrase (PO) object;

(3.1) *a*. DO, plausible: *The mother gave the daughter the candle*.
b. PO, plausible: *The mother gave the candle to the daughter*.
c. DO, implausible: *The mother gave the candle the daughter*.
d. PO, implausible: *The mother gave the daughter to the candle*.

2) the transitive/intransitive;

(3.2) *a*. transitive, plausible: *The tax law benefited the businessman.*
b. intransitive, plausible: *The businessman benefited from the tax law*.
c. transitive, implausible: *The businessman benefited the tax law*.
d. intransitive, implausible: *The tax law benefited from the businessman*.

3) the active/passive alternation.

(3.3) *a*. active, plausible: *The girl kicked the ball*.
b. passive, plausible: *The ball was kicked by the girl*.
c. active, implausible: *The ball kicked the girl*.
d. passive, implausible: *The girl was kicked by the ball*.

After hearing the sentence in one of the four conditions, participants were requested to answer a question like *Did the daughter receive something/someone?* (referring to Examples 3.1). As a result, the implausible sentences were interpreted as their plausible counterparts much more for the double object/prepositional object and transitive/intransitive alternations than for the active/passive alternation. Also when errors were present in the filler materials, i.e. under a higher error rate, the implausible active/passive alternations were interpreted literally most of the time. Gibson et al. suggest that the explanation could be in the perceived likelihood that the implausible utterance appeared due to noise. In case of the DO/PO and transitive/intransitive constructions, the difference between the implausible and the plausible utterances is only in one edit, namely insertion or deletion of a preposition. In contrast, repairing the implausible active/passive alternations involves two edits (insertion or deletion of an auxiliary and a preposition). Obviously, the participants found it less likely that noise could have corrupted the implausible active/passive constructions to such a large degree and were not ready to repair them.

The same materials were used in a later experiment by Gibson et al. (in press) with a new variable – accented or native speaker pronunciation. The effect of the presence of a foreign accent turned out to be significant, but only for the DO/PO and transitive/intransitive alternations. The accented pronunciation significantly reduced the literal interpretations and facilitated the repair of the implausible DO/PO and transitive/intransitive sentences into the plausible ones. However, accent did not affect the perception of the active/passive stimuli; in all conditions, they were interpreted literally most of

the time. Besides the effect of the accent, the probability of plausibility-based inferences for DO/PO and the transitive/intransitive alternations increased depending on the type of corruption. Participants waived the literal interpretations of implausible sentences more readily when the preposition was missing (in the DO and the transitive constructions) than when an unnecessary function word was present (in the PO and intransitive constructions). This means that deletions invite internal repairs to a greater extent than insertions do.

In the discussion of these findings, Gibson et al. wonder why, despite the advantage of being understood better, speakers with an accent are perceived to be less credible (Lev-Ari and Keysar 2010, Livingston, Schilpzand and Erez 2014), less educated (Fraser and Kelly 2012), less intelligent (Fuertes, Potere and Ramirez 2002) or less hirable (Huang, Frideger and Pearce 2014). By contrast, if a native speaker utters something implausible, the comprehenders just deal with it, and the speaker remains misunderstood. The authors suggest that it might have to do with the time needed to adapt to a speaker (time pressure increases "the benefit of the doubt").

I believe this apparent contradiction could rather be explained by the expectations of the proficiency level regarding the status of the speaker (cf. the "foreigner role" discussed in Chapter 2). A tourist visiting the US and speaking with a strong accent will definitely benefit from the doubt and enjoy the advantage without negative judgements of his/her credibility, intelligence or education. But if this person tries to do business or get hired for a position involving public relations, the attitudes are likely to be different. The psycholinguistic experiments have only demonstrated that native speakers ARE ABLE to understand and plausibly interpret learners' erroneous utterances relying on world knowledge. This implies that both tourists, foreign business partners or job applicants will be understood. However, if one asks about evaluation, this is where sociolinguistics comes in, i.e. the perception depending on the social role of the speaker.

3.7 Summary

The attempts to investigate factors involving psycholinguistic mechanisms of error processing (e.g. error size, item frequency, error rate, the level of abstractness) by means of off-line judgment tasks were not quite successful. For example, the effect of markedness on error gravity was not robust. Besides, it could not be decided whether grammatical errors, on aggregate, are more serious than lexical errors, or vice versa. The problem might be that processing mechanisms should rather be investigated by means of real-time (on-line) methods, and the variables should be properly operationalized.

Based on psycholinguistic mechanisms of word recognition, it can be hypothesized that the least processing cost is produced by local errors occurring within a few letters. The nature of the error may be various: spelling, less likely grammar, and least likely lexical. There is a probability for such errors to be overlooked. However, this can happen only in fast reading and if the incorrect word resembles the correct word. In slow and attentive reading, all kinds of errors are normally detected.

Psycholinguistic and neurolinguistic studies have shown that there are similarities and differences between processing in L1 and L2. Diverse mechanisms and circumstances of first and second language acquisition result in the attainment of implicit knowledge for L1 and explicit knowledge for L2. Consequently, error detection in a foreign language is based on the verification with rules, whereas native speakers use their "feel". With greater proficiency in L2, reactions become similar to those of native speakers. Besides, L1 and L2 share the same mental lexicon. However, similar comprehension results are often achieved by different routes. Processing in L2 appears to be slower in speed, shorter in span and shallower in depth, compared to that of native speakers. This is caused by the fact that learners' use of multiple resources while processing puts a strain on the working memory. Lacking probabilistic knowledge for large syntactic constructions, non-native speakers focus on the relations between adjacent words and compensate their inability to control the relations at large distances by relying on semantic and pragmatic information.

Finally, there is much evidence that listeners and readers switch their psycholinguistic settings for communicating with different speaker groups, i.e. they adjust their expectations, readiness for internal repairs and evaluations to the relevant group. This discrepancy between negative evaluations for speakers with an accent and a greater willingness to make sense of what they are saying indicates that the relationship between judgment and processing difficulty is very complex, and that the correlation between them may not be as strong as intuitively expected.

4. The Effect of Distance in Processing Wrong Temporal Adverbials

This chapter explores the effect of **distance** on the **reading times** for matching and mismatching conditions, i.e. **tense errors**, in **native** and **non-native speakers** of English by means of an **eye-tracking experiment**. Results of a previous judgment test and a **corpus study** conducted for this thesis provide solid grounds for the expectation of latencies in reaction to mismatching temporal adverbials with past time reference. The effects of adverbial and verb form **frequencies** are investigated as well. The results are analyzed across the **L1-varieties** of English and compared to those of German learners.

4.1 The Distance Hypothesis

The research question investigated in this chapter has two mutually exclusive hypotheses, which are both psychologically plausible and will be tested. The first hypothesis holds that the further away the second part of the error (~ the disambiguating item) stands from the first part, the more cognitive work it requires to repair the sentence. Presumably, the reader has to go back to the first part of the error and try to solve the problem by correcting either the first or the second item. This hypothesis is supported by locality theories, e.g. Gibson (1998):

> (1) the longer a predicted category must be kept in memory before the prediction is satisfied, the greater is the cost for maintaining that prediction; and (2) the greater the distance between an incoming word and the most local head or dependent to which it attaches, the greater the integration cost.

Besides that, in situations of normal reading (without errors), processing the dependent element should be easier because it is anticipated based on the properties of the head, e.g. the verb (Konieczny et al. 1997). In the present experiment, the error cannot be anticipated, therefore, longer reading times are predicted.

The second hypothesis, on the contrary, predicts no difference or even more rapid processing of the mismatching item standing further away from the first item for the following reasons. There is a common observation that reading times usually do not increase toward the end of clauses or when clauses get longer; in fact, reading often even speeds up toward the end of sentences

(Konieczny 2000). Moreover, there are also limitations of the working memory in that by the time readers get to the second mismatching element, they might lose sensitivity for the grammatical form of the first element, if it stands a few constituents away.

Thus, the question is whether growing distance between two mismatching parts increases the reading times for the second part where the error is discovered. If distance between two mutually dependent elements does not influence processing, the reading times should not be significantly different for short and long sentences, as well as there should be no interaction between distance and the grammaticality of the sentence.

To investigate these hypotheses, I manipulated the distance between the verb form and the temporal adverbial with past time reference. English, like many other languages, has two competing forms to refer to past time: the periphrastic Present Perfect and the synthetic preterite (Elsness 2009). The past tense form of the verb (also called the Simple Past or the preterite) is usually specified by such temporal adverbials as *yesterday,* a certain number of *days/weeks/months/years ago, last week/month/year, the other day*, etc., whereas a verb in the Present Perfect is used with *lately, in/over/for the last* (number of) *week/month/year(s), so far, for now*, etc. Compare, for example:

(4.1) *a. It **snowed** here **a week ago** and we had a chance to play snowball fight.* (Item *10c* in the present experiment, Appendix 1)
*b. It **has snowed** here **for a week now** and we had a chance to play snowball fight.* (Item *10b*)

If the temporal adverbial is meant to strengthen the tense-and-aspect meaning of the corresponding verb form, then the wrong adverbial should cause a mismatch. In teaching English as a foreign language it is usually marked as an error.

(4.2) *a. *It **has snowed** here **a week ago** and we had a chance to play snowball fight.* (Item *10a*)
*b. *It **snowed** here **for a week now** and we had a chance to play snowball fight.* (Item *10c*)

Students whose native languages do not have a formal distinction between the Simple Past (SP) and the Present Perfect (PP), e.g. Russian, or whose language has a different use of the corresponding tense and aspect forms, e.g. German, require special training to learn how to use these forms and temporal adverbials correctly when speaking or writing in English. Yet their use may vary across varieties. Combinations of the SP and such adverbials as *just, yet, already* have become widespread in American English (e.g. Biber et al. 2002, Huddleston et al. 2005) and for this reason were not used in the present experiment.

In order to check whether particular combinations of the tense forms and temporal adverbials can be contrasted as correct and wrong, I extracted the frequencies of their co-occurrence from the Corpus of Contemporary American English (COCA) containing more than 400 million words of text (as of 2010 and 450 million words as of 2013, when the corpus study was carried out). According to its creator (M. Davis 2008), the corpus is "equally divided among spoken, fiction, popular magazines, newspapers, and academic texts". It includes 20 million words each year from 1990-2015, and the corpus was also updated regularly (e.g. contains 520 million words as of March 2017).

The corpus frequency of the verbs (n=40, chosen randomly) in the SP or the PP used with temporal adverbials mostly confirm the rules presented to English learners in the classroom. However, counterexamples were found, too.

(4.3) *a. We saw some rain at the beginning of the weekend, but it'**s been beautiful yesterday**. (2001 SPOK)*

*b. It'**s been** one year **ago** since Hurricane Andrew swept across South Florida… (1993 SPOK)*

*c. I'**ve** just **been** in Natal **a week ago**… (2006 FIC)*

*d. All right I know you may think you'**ve heard** this story **years ago**… (2004 SPOK)*

*e. …it should be a state that George Bush **has won** two weeks **ago**. (2000 SPOK)*

*f. And while tragically we'**ve lost** our first soldier there **yesterday**, the mission seems to me still to be appropriate. (1993 NEWS)*

*g. What **was** wrong with him **lately**? (2003 FIC)*

*h. …young kids who said **in the last month** they **tried** a drug. (1996 SPOK)*

*i. The business of telling people what **happened in the last week** is just about gone… (2009 NEWS)*

*j. I mean, the movie's as entertaining as anything I **saw in the last year**. (1996 SPOK)*

In COCA, such examples appear in small numbers (0.45% with SPs and 11.14% with PPs), compared to the traditional use, and are mostly found in spoken sources and in the news. A more detailed analysis of the frequency distribution of the tense forms and the temporal adverbials is presented in section 4.4.2.

Evidence that preferences truly exist can be found in the corpus examples where (presumably native) speakers switch from PPs to SPs to use them with appropriate temporal adverbials.

(4.4) a. *That's the Constitution. People have **asked** me, **asked yesterday**, "Isn't that inconvenient for the president?" (1990 SPOK)*

*b. The President **has said, said yesterday**, has said before, and I'm quoting him… (1998 SPOK)*

*c…. and **we have heard** about other things. **We heard yesterday** about cooperation in competition matters. (2001 ACAD)*

*d. And, you know, **you've seen -- I saw last year** one program -- I forget which it was… (1992 SPOK)*

*e.…to visit red-hot Hawks, **who have won** four straight; Norcross **won** 28-14 **last year**. (2007 NEWS)*

*f. Chambliss **has said** -- and he **repeated** it **last week** -- that he is not running for statewide office. (1997 NEWS)*

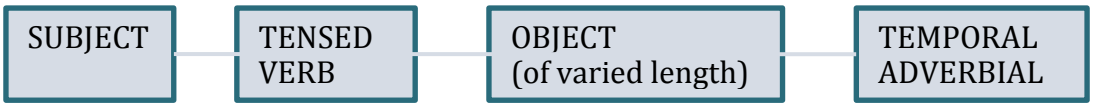
*g. We'**ve lived** through hurricanes and **last year** we **lived** through the seaweed which was just terrible. (1990 SPOK)*

*h. Ross Perot **has gone** on television, **last week** in an interview **said**,' I never hired Ed Rollins. (1992 SPOK)*

Summing up, it needs to be mentioned that grammar textbooks usually present rules where the use of temporal adverbials is seen as correct or wrong. However, corpus data show that language use does not entirely fulfill the rules but reflects the PREFERENCES, i.e. more frequent tendencies in a particular discourse situation and/or variety. Thus, the use of 'wrong' adverbials is possible, but infrequent and disfavored, which is why an increase in the processing times can be expected.

4.2 Experimental design

All stimuli relating to the Distance Hypothesis (320 sentences, Appendix 1) had the following structure:



There were 40 different sentences (item number 1-40), each having eight versions (conditions *a-h*) including four target sentences (containing the mismatch) and four control sentences (with a matching adverbial), according to the following scheme:

Target:	Past ~ *lately*	vs.	Perfect ~ *yesterday*
Control:	Past ~ *yesterday*	vs.	Perfect ~ *lately*

Four sentences had a short distance between the verb form and the adverbial, and the other four sentences had a longer distance, usually by means of adding an adverb and a prepositional phrase.

(4.5) *a. *I'**ve found** good books **last year** and I can recommend you some.*

*b. I'**ve found** good books **lately** and I can recommend you some.*

*c. I **found** good books **last year** and I can recommend you some.*

*d. *I **found** good books **lately** and I can recommend you some.*

*e. *I'**ve found** very good books on ancient history **last year** and I can recommend you some.*

*f. I'**ve found** very good books on ancient history **lately** and I can recommend you some.*

*g. I **found** very good books on ancient history **last year** and I can recommend you some.*

*h. *I **found** very good books on ancient history **lately** and I can recommend you some.*

In all sentences, the critical region, i.e. the part of the sentence where the reading times are analyzed, was the temporal adverbial (in case it consisted of a few words, e.g. *five years ago*, it is the whole phrase). In psycholinguistic experiments, the critical region should not include the last word in the sentence because of the "wrap-up effect". It shows itself as slowdowns at clause and sentence boundaries and is thought to be the point where the reader constructs a higher-level meaning representation and no longer retains the actual words of a clause or a sentence (Field 2004). To minimize the influence of this effect in the present experiment, the sentences had to be extended beyond the temporal adverbial. Even a comma after the critical word could encourage the readers to make an extra pause (Hirotani, Frazier and Rayner 2006), so commas between the clauses were avoided.

Additionally, the reading times for the words preceding the temporal adverbial were controlled for the spillover effect. The calculation took into account the reading times for two regions (constituents) preceding the temporal adverbial.

The dependent variable is the reading times (in milliseconds) at a particular region. Three measures were computed from the fixation reports: the first pass reading time (FPRT), the regression path duration (RPD), and the total reading time (TRT). In the data evaluation, I concentrated mainly on the regression path duration with the assumption that it is most suitable for studying the perception of errors. This measure reflects the time that readers need in order to finish processing the target region, also if they want to go back to the previous parts of the sentence and read them again, before they are ready to move on to the next word.

The main predictors (independent variables) are:

(1) the matching or mismatching temporal adverbial,
(2) the distance between the verb and the adverbial (short/long), and
(3) the language (L1/L2), to account for the differences between native and non-native speakers.

The stimuli were presented in different sentence types, which resulted in additional variables: statements vs. questions, presence vs. absence of negation, clause type (main vs. subordinate), verb type (regular vs. irregular), the form of the auxiliary (full vs. contracted), the length of the target region (in number of characters), the position of the target region in the sentence, the numerical order of the stimulus ("trial"). These factors were statistically controlled (added as "fixed effects").

The statistical model also included so-called "random factors", i.e. the ones which cannot be controlled for in experimental design and for which the

software makes relevant adjustments (e.g. intercepts and slopes for the subjects and items). Among them are: the participants' reading speed, possible difficulty of the sentence, individual reaction to the predicting variables, etc. The adjustments of the random effects serve to increase the generalizability of the analysis.

The sentences were presented according to a Latin square design, which means that a participant saw the sentence only in one condition out of eight (from "a" to "h"). The order of sentences was rotated in a way that the participant received a list with the first sentence (item) in condition "a", the second sentence in condition "b", the third sentence in condition "c", etc. (see Table 4.1), to ensure that all conditions are equally presented to the subject.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8…
List 1	a	b	c	d	e	f	g	h
List 2	b	c	d	e	f	g	h	a
List 3	c	d	e	f	g	h	a	b
List 4	d	e	f	g	h	a	b	c
List 5	e	f	g	h	a	b	c	d
List 6	f	g	h	a	b	c	d	e
List 7…	g	h	a	b	c	d	e	f

Table 4.1. *Example of a Latin square design*

The experiment consisted of two parts with different goals and experimental designs. The stimuli for the Distance Hypothesis were mixed with other sentences in proportion of about 1 to 7, where the seven sentences belonged to the part of the experiment discussed in Chapter 5. The order of sentences was randomized into 48 lists, so that most participants were presented with a unique sequence of stimuli.

4.3 Procedure

Before participating in the experiment, the informants were asked to fill out a questionnaire with their personal information: age, gender, area of expertise (major subjects studied), knowledge of foreign languages, details of their reading experience and literacy acquisition. In the eye-tracking laboratory, the participants received an instruction in German and in English. The text of the instruction was printed out and was always available to the participant. The procedure was orally explained as well.

The eye movements were recorded with an SR Research EyeLink 1000 sampling gaze data with 1 data point per millisecond. Participants sat at the computer with their chin resting on the head support, holding a gamepad with both hands. Each sentence was presented alone in the middle of a 20-inch flat screen with the resolution of 1600 by 1200 pixels. The distance between the eyes of the participant and the presentation screen was 600 millimeters.

Setting the equipment, adjusting the camera and going through the tests took about 5-15 minutes. The experiment began with three training items to practice reading and answering the questions. They looked similar to the experimental items but were not evaluated.

The training items were followed by 330 sentences of the experiment. The participant had to press a button on the gamepad as soon as s/he had finished reading the sentence. The stimulus sentence was followed by a comprehension question with two suggested answers: *yes* and *no*. The participant pressed the corresponding button on the backside panel of the gamepad. Reading and answering the questions took about 45-55 minutes. The German subjects needed approximately 10 minutes longer than the native speakers.

To determine the level of proficiency in English, the German participants were asked to complete a small web-based subset of the TOEFL test[4]. The English test consisted of 20 sentences, each with 4 multiple choice questions and tasks to select the wrong answer or to choose the best answer. The participants scored from 8 to 20 out of 20.

The eye-tracking experiment was carried out with 61 participants. Later, the data for 5 of them were removed, and the analyses reported below include 56 participants (Appendix 2): 32 native speakers of English (57%) and 24 German learners of English (43%), aged from 19 to 33. There were 44 female (79%) and 12 male (21%) participants. Among the native speakers, I had representatives of the standard varieties of English: UK (12 subjects – 38%), USA (17 subjects – 53%), and Canada (3 subjects – 9%).

Based on the information available in the questionnaires, the informants were coded for whether or not they had studied a language as a major subject ("yes" for 36 subjects – 64%).

Language	Language as major	
	yes	no
Native	21	11
Non-native	15	9

Table 4.2. *Number of subjects majoring in language studies*

[4] http://www.stuff.co.uk/toefl_structure_1.htm

4.4 Results

4.4.1 Distance between the verb and the temporal adverbial

To evaluate the effect of distance in the grammatical and ungrammatical sentences, mixed-effects models were computed in R using the lme4 package (Bates et al. 2014). The dependent variable is the regression path duration for the temporal adverbial where the mismatch was presumably discovered. The time measures were conditionalized (zero values were treated as not available) and logarithmically transformed. The outliers have been removed.

The results are presented in the form of the t-value. The difference from zero indicated by the t-value should prove that the measure has not appeared due to sampling error. The "null hypothesis" holds that there is no relationship between the measured phenomena. It has been conventional to reject the null hypothesis at the highest significance level of 5% ($p<0.05$[5]). It means that, according to the statistical assessment, the probability that the observations do not reflect a pattern but just chance, is less than in 5% of cases. The results with the probability of 1% or smaller ($p<0.01$, $p<0.001$) are accepted as highly significant. However, the t-value alone indicates significance: the results with the $t > 2.0$ or $t < -2.0$ are statistically significant.

Does distance matter for the processing of mismatching adverbials? The general answer to this question can be derived from Figure 4.1 below (the whiskers represent confidence intervals[6]).

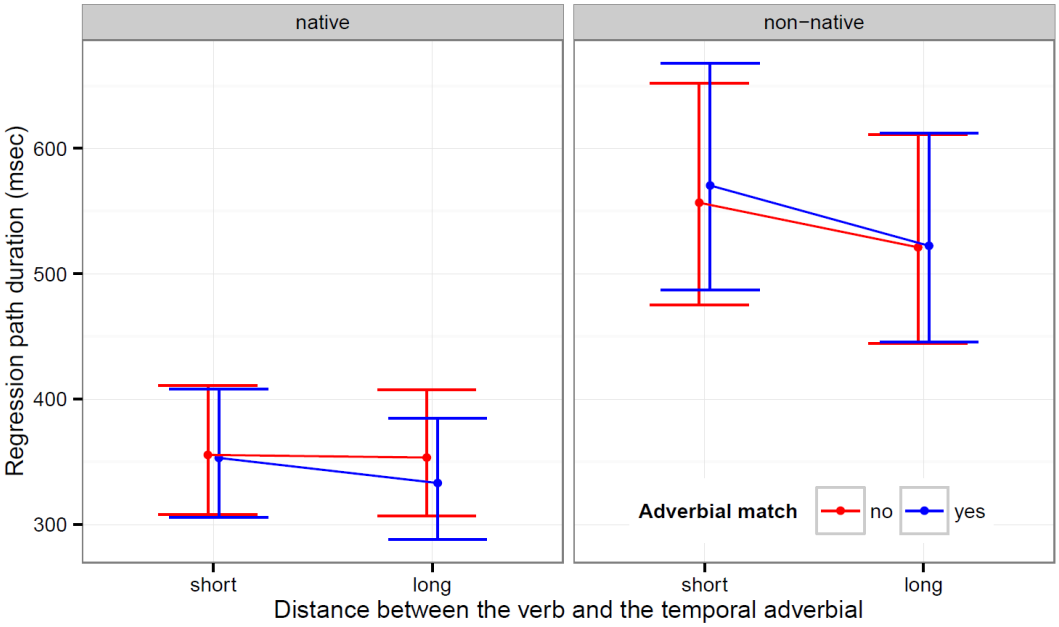


Figure 4.1. *RPD for short and long sentences*

[5] For a technical reason, linear regression models did not automatically return p-values, therefore, in most cases, only t-values are reported.
[6] This plot and other similar images were produced in R with the package *ggplot2* (Wickham 2009) based on the package *effects* (Fox 2003).

In Figure 4.1, the difference between the reaction to the matching and mismatching adverbials is not significant in both subject groups (the interaction of adverbial match and the language, t = -1.099). There is also no difference between the reading times for short and long sentences. Below are the details of the mixed effects model.

	Estimate	Std. Error	t value
(Intercept)	5.341e+00	5.591e-01	9.552
length[S.short]	2.732e-02	5.246e-02	0.521
adv_match[S.yes]	-4.916e-03	1.043e-02	-0.471
Language[S.native]	-2.208e-01	3.153e-02	-7.002
log(reg.length)	4.080e-01	4.746e-02	8.596
rpd.m1	4.584e-05	3.368e-05	1.361
rpd.m2	8.774e-05	2.942e-05	2.982
ia.index	-7.822e-03	1.046e-01	-0.075
trial	-5.717e-05	1.469e-04	-0.389
log(Verb_Form_Frequency)	-2.993e-02	1.168e-02	-2.562
log(adv_freq)	-2.214e-02	7.879e-03	-2.810
age	1.779e-03	9.714e-03	0.183
sex[S.f]	4.691e-02	4.183e-02	1.121
haender[S.links]	1.381e-02	5.330e-02	0.259
Language_major[S.no]	8.802e-02	3.666e-02	2.401
auxil_form[S.contracted]	-1.494e-02	2.700e-02	-0.553
auxil_form[S.full]	-2.153e-02	2.496e-02	-0.862
clause_type[S.main]	1.176e-02	4.680e-02	0.251
clause_type[S.sub_adv]	-6.527e-02	6.034e-02	-1.082
negation[S.no]	5.395e-02	4.181e-02	1.291
sentence_type[S.question]	-1.403e-01	5.556e-02	-2.526
verb_type[S.irregular]	1.563e-03	3.031e-02	0.052
length[S.short]:adv_match[S.yes]	9.304e-03	1.042e-02	0.893
length[S.short]:Language[S.native]	-1.123e-02	1.237e-02	-0.907
adv_match[S.yes]:Language[S.native]	-1.159e-02	1.054e-02	-1.099
length[S.short]:adv_match[S.yes]:Language[S.native]	3.812e-03	1.038e-02	0.367

Table 4.3. *The effects of the object length model (using the sum contrasts)*

In this analysis, native and non-native speakers showed the same pattern in the perception of distance in that a temporal adverbial standing closer to the verb, numerically, took a bit longer to read. However, statistically, the effect of object length is not significant (t = 0.521). In spite of the same pattern, native and non-native speakers differ significantly in their general reading speed (main effect of language, t = -7.002). The German readers were about 200 msec slower in both conditions. Non-native speakers' use of more time can be explained by their need to decode each word and involve multiple resources for processing (as described in Chapter 3).

All in all, absence of the object length effect in this analysis (interaction of length and adverbial match, t = 0.893) does not mean that distance does not matter for the analysis of ungrammatical sentences. It might only be that a mismatching temporal adverbial is too slight a deviation (t = -0.471) to create sufficient grounds for a distance effect to manifest itself.

There is a tendency that subjects studying a language as a major subject showed faster reading times (t = 2.401). However, a separate analysis in this subgroup did not reveal any significantly different reaction to the object length (t = -0.481) or to the adverbial match (t = -0.288), or the interaction of the two (t = -0.432), and the effect of native language remained highly significant (t = -6.447).

An interesting observation is that adverbials are processed faster if they occur in questions (t = -2.526). I hesitate to call this an effect because there were only 3 questions against 37 statements in the sample. Despite this minority, questions were read significantly faster. A suggested explanation might be the influence of prosody. Consider the following examples.

(4.6) Questions:

a. Have you bought any stamps lately and how many do you have by now? (Item *22b*)

b. Did you ask her about her work yesterday and is it going well? (Item 33c)

Statements:

c. It has snowed here for a week now and we had a chance to play snowball fight. (Item 10b)

d. He took some medication last year and it really seemed to help. (Item 32c)

It is possible that in a question, by default, the reader is more likely to be interested in the action taking place or not (or other circumstances indicated by a *wh*-question word), so that the object receives the nuclear tone. Thus, the temporal adverbial becomes unstressed and is read less prominently. In a statement, however, the time specification may be the center of attention (e.g. 4.6-c). Based on the canonical information structure, the adverbial appears at the end of the clause and is likely to be the rheme, the new information. In that case, the prosodic weight falls on the temporal adverbial and it would take more time to read. Certainly, there may be counterexamples to this suggestion, motivated by the preceding context. It should be studied which prosodic pattern is generally preferred in context-free sentences and whether unstressed positions result in shorter reading times.

Another significant result is the spillover effect of the region located two constituents prior to the temporal adverbial (t = 2.982). In the short sentences, it is the verb; in the long sentences, it is often the direct object. The influence of the immediately preceding region is not significant (t = 1.361).

The order of the stimulus sentences (trials) as such had no effect (t = -0.389). The reading times for the adverbials could not be influenced by such grammatical properties of the sentences as presence or absence of negation (t = 1.291), the clause type (main : subordinate adverbial, t = 0.251; main : subordinate relative, t = -1.082), the verb type (irregular : regular,

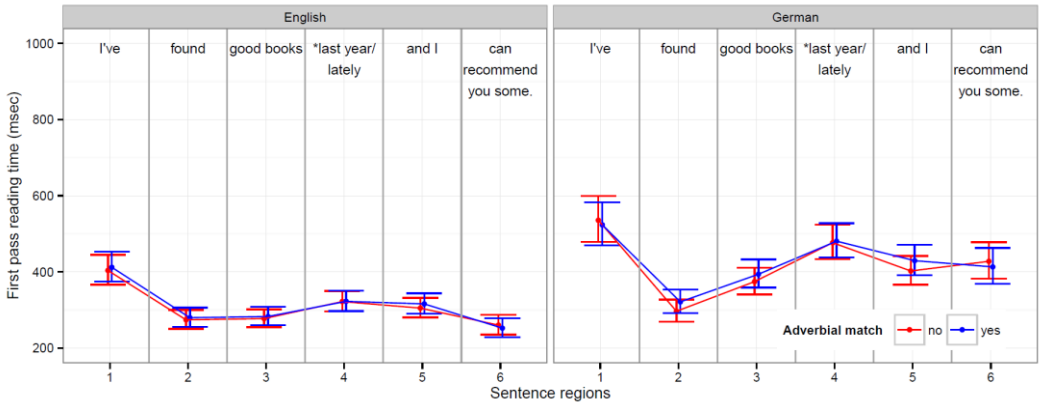
t = 0.052), or the auxiliary form (contracted : full, t = -0.553; contracted : no auxiliary, t = -0.862). The variation in the subjects did not play any important role either: age (t = 0.183), gender (t = 1.121), the dominant hand (left : right, t = 0.259).

The results reported above were based on the analysis of the reading times only for the temporal adverbial. To be sure that the effect of the adverbial match and the object length did not appear in other regions of the sentence (e.g. in the region following the adverbial), I visualized the reading times for the whole sentences in all conditions and in all measures. In the short condition, each sentence was divided into the following regions:

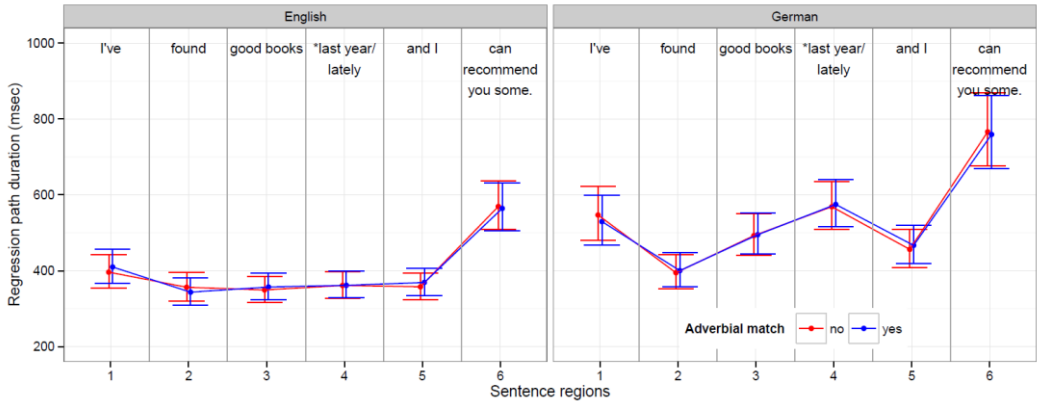
1) the subject and, if applicable, the auxiliary;
2) the verb;
3) constituents standing between the verb and the temporal adverbial (an object or an adverbial modifier);
4) the temporal adverbial;
5) two words following the temporal adverbial (mostly *and* + subject);
6) the rest of the sentence.

Figures 4.2 *a-c* below show the reading times for the short condition where I previously analyzed only Region 4. The labeling of the regions is based on an example sentence.

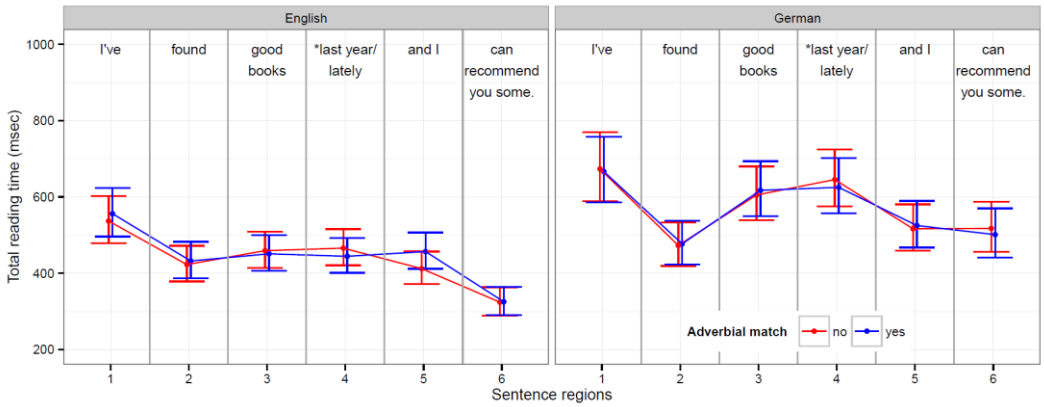
a.



b.



c.



Figures 4.2. *Time measures for whole sentences in the short condition*

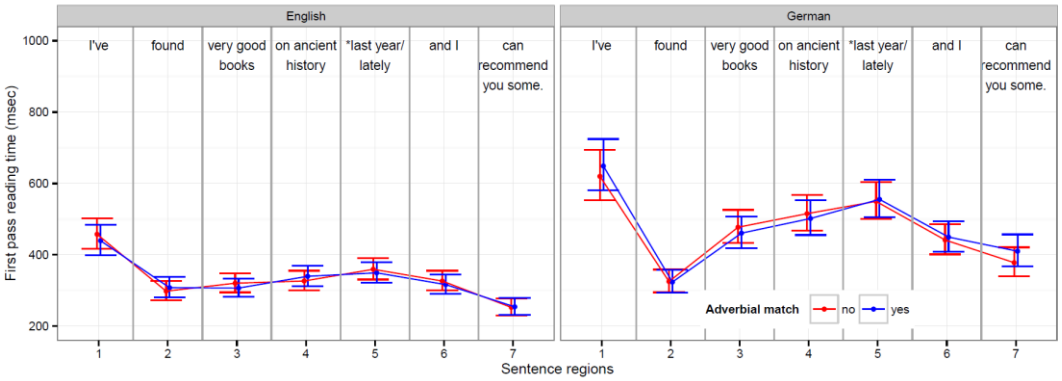
The diagrams clearly show that there were no significant differences in the processing of the grammatical and ungrammatical sentences (except in the TRT for Region 5 in native speakers). The rise towards the temporal adverbial in the FPRT and the TRT can be viewed as a clause wrap-up effect (integration costs), and the highest reading times in the final region in the RPD as a sentence wrap-up effect.

In the long condition, each sentence was divided into the following regions:

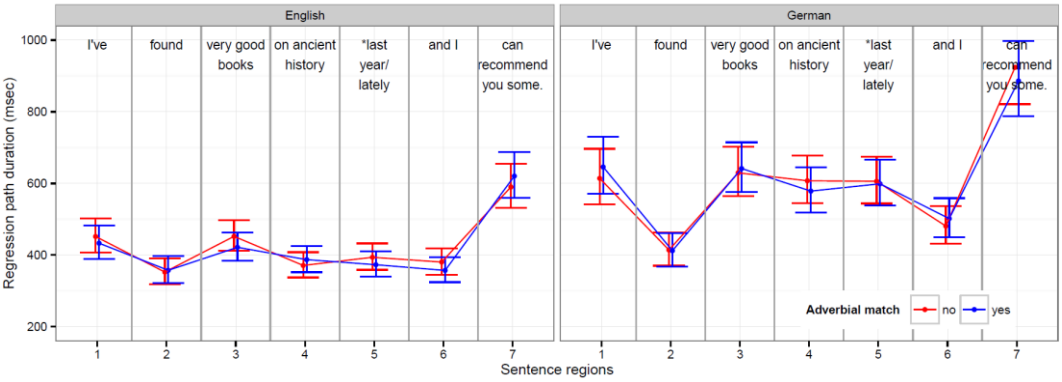
1) the subject and, if applicable, the auxiliary;
2) the verb;
3) the first constituent following the verb (an object and/or a modifier);
4) the constituent preceding the temporal adverbial (a prepositional object);
5) the temporal adverbial;
6) two words following the temporal adverbial (mostly *and* + subject);
7) the rest of the sentence.

Figures 4.3 *a-c* show the reading times for the long condition where I previously analyzed only Region 5. They also demonstrate no significant differences in the perception of sentences with matching and mismatching adverbials. Again, there is a sentence wrap-up effect in the RPD.

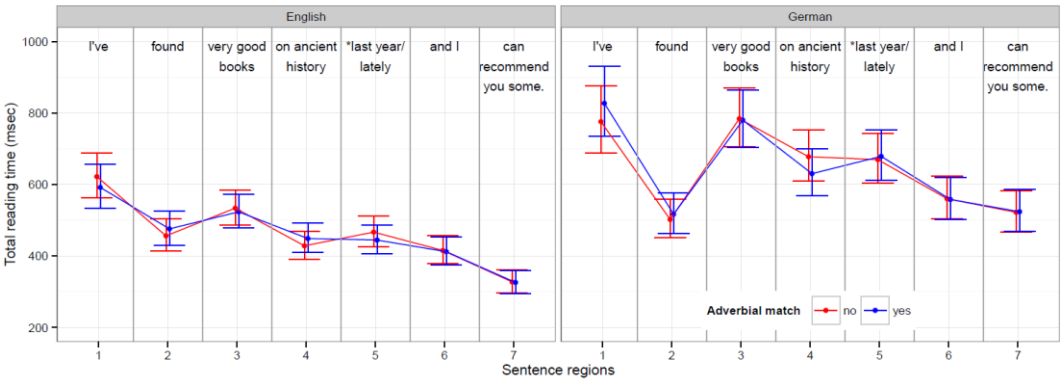
a.



b.



c.



Figures 4.3. *Time measures for whole sentences in the long condition*

Thus, extending the object and increasing the distance between the verb and the temporal adverbial did not affect the reading times at any region of the sentence to a significant extent. In the following sections, both short and long sentences will be analyzed together, and the object length will be statistically controlled.

4.4.2 Frequency effects in the perception of temporal adverbials

Frequency effects have been found in many areas of language use (for review see, e.g. N. Ellis 2002, Gries and Divjak 2012). Clifton et al. (2007) summarize some frequency effects associated with eye movement studies. There is abundant evidence that the frequency of a fixated word influences how long readers look at it (Inhoff and Rayner 1986, Rayner and Duffy 1986). Based on this, I hypothesized that the highly frequent adverbials are processed faster, and this might influence the perception of distance or of the mismatching adverbial. Firstly, I extracted the token frequencies of the adverbials from the Corpus of Contemporary American English. The adverbials were searched for exactly in the same form as they appear in the stimulus sentences.

The token frequency of the adverbials is very diverse (see Figure 4.4), from very common adverbials (*last year, last week*) to rare ones (*for 3 years now, in the last 5 years*). Among the highly frequent adverbials are mostly those which are used with the SP, except for *so far* – the commonest PP adverbial. The less frequent PP adverbials, *lately* and *by now*, follow quite far behind but are also found in the SP-dominant part. *In the last year* opens the less frequent half of my adverbials, which is strongly PP dominant. The least frequent SP adverbial in my selection is *at 10 a.m*. It is obvious that SP adverbials are found more often due to the predominance of the preterite, and that exact time specification results in less frequency. It is important here to have a wide distribution of adverbial frequencies, which provides conditions for a frequency effect to show itself.

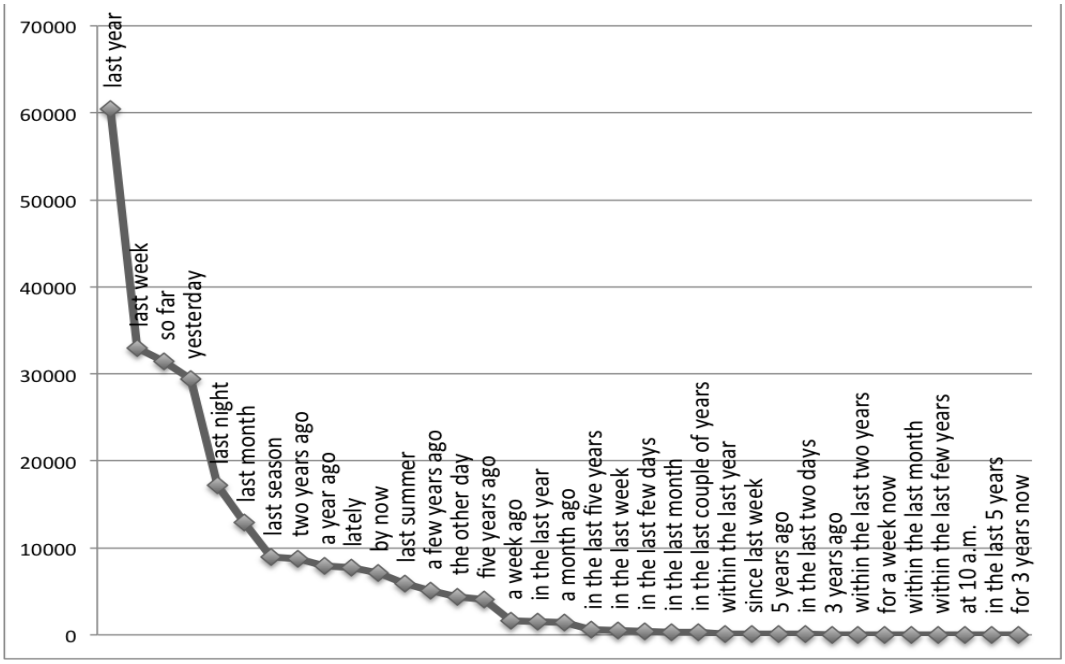


Figure 4.4. *Absolute frequencies of the selected adverbials per 450 million words*

In the previous section, the adverbial frequency was added as a fixed effect and was statistically significant. In the new model, I checked for its interaction with other predictors. The main effect of the logarithmically transformed adverbial frequency proved to be significant (t = -2.797, sum contrasts), but no interactions were found.

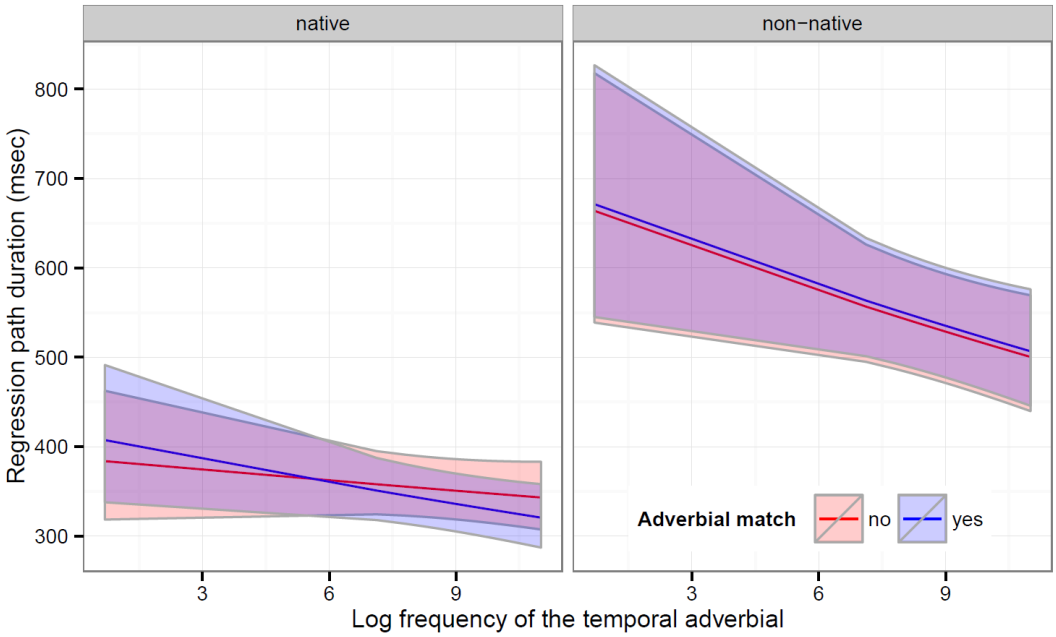


Figure 4.5. *The effect of adverbial frequency*

Figure 4.5 shows that highly frequent adverbials are processed clearly faster in both native and non-native speakers. Frequency does not have any impact on the perception of the mismatching adverbials (t = -0.640). In the diagram, there is a slight indication that native speakers need more time for processing high-frequency adverbials in the mismatching condition, but this three-way interaction is not significant (t = -0.723).

Looking for other frequency effects, I hypothesized that highly frequent **verbs** may facilitate the processing of temporal adverbials, especially in the past tense, and low frequent verbs may slow it down. Similar spillover effects were previously found in relation to word frequency, although only at a distance of one word (n + 1).

> Inhoff and Rayner (1986) and Rayner and Duffy (1986) found that fixation times were 30-90 msec longer on low-frequency words than high-frequency words. However, when they examined the fixation time on the next word […], they also found that its fixation time increased by 30-40 msec. It seems that the processing associated with the low-frequency word spilled over onto the processing of the next word in the text. Such spillover effects are quite pervasive… (Rayner et al. 1989)

In the next model, I will test whether or not the frequency of the verb in the SP or in the PP is pervasive enough at a distance of more than two words to affect the reading times for the temporal adverbial in native and non-native speakers.

Verb form frequencies were extracted from COCA as well. The verbs used in the experiment (n=40) were chosen randomly. The use of the verb in the PP was calculated as the sum of two columns: (1) the number of tokens with the auxiliary *has* and *'s* together[7] and (2) the number of tokens with the auxiliary *have* and *'ve* together[8]. To calculate the occurrence of the verb in the SP, I added the following two columns: (1) the number of past tense forms in affirmative sentences[9] and (2) the number of infinitive forms in the 3-word prior context of the auxiliary *did* for negations and questions[10]. The final value means the absolute frequency of the particular verb in the SP or PP against a corpus of 400 million words (as of 2010), which should also reflect the verb frequency in general.

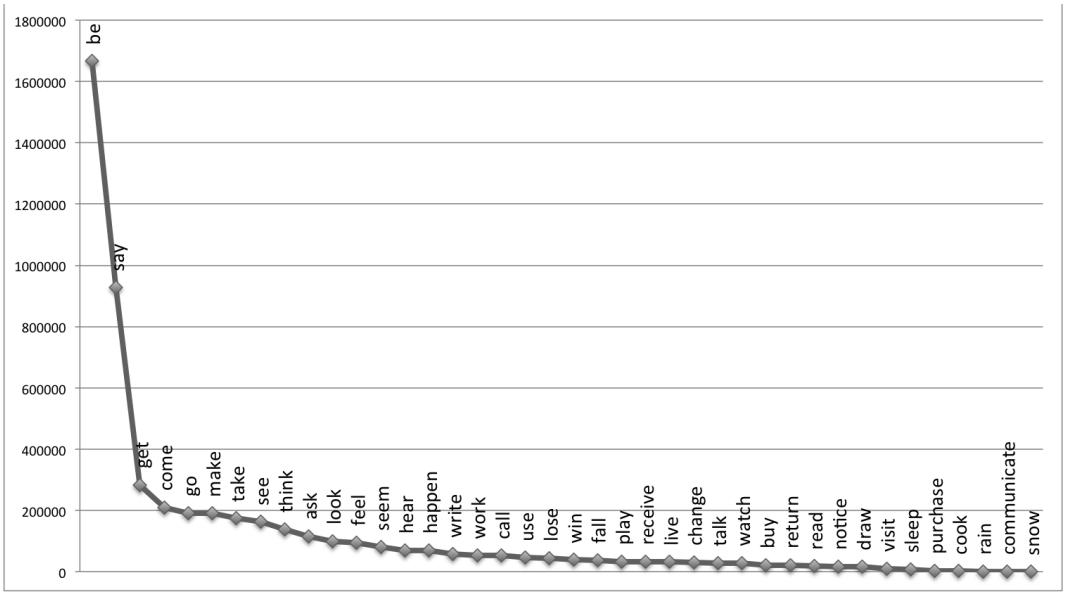


Figure 4.6. *Verb frequency with past tense reference (SP+PP) per 400 million words*

The selection of verbs included very frequent ones (*be*[11]*, say, get*) as well as low frequent verbs (*snow, communicate, rain*). Biber et al. (2002) acknowledge the main verb *be* as the "most important copular verb in English" and discusses

[7] Formula: -[vb*] verb.[vvn] in the context of [vhz] 3;0. Meaning: "exclude all forms of the verb "to be" (to remove the passive) in the context of the auxiliary "have" in 3rd person singular up to 3 words ahead (to allow negations, adverbs and questions).
[8] Formula: -[vb*] verb.[vvn] in the context of [vh0] 3;0. Meaning: "exclude all forms of the verb "to be" in the context of the auxiliary "have" up to 3 words ahead.
[9] The past tense tag [vvd].
[10] Formula: verb.[vvi] in the context of "did" 3;0.
[11] To omit the auxiliary uses of the verb "to be", the search formula excluded any verb forms following the *be*-form: been -[vv*], was -[vv*].

it separately from the lexical verbs, which makes it difficult to compare the frequencies. Their chart of the most common lexical verbs in English starts with *say*, *get* and *go* (op. cit.: 110), where *say* is also considerably ahead of *get*, but *go* following it very tightly. To have a better idea of the frequency distribution, I zoomed in to the lower part of the chart by excluding the three most frequent verbs (for the illustration purposes only).

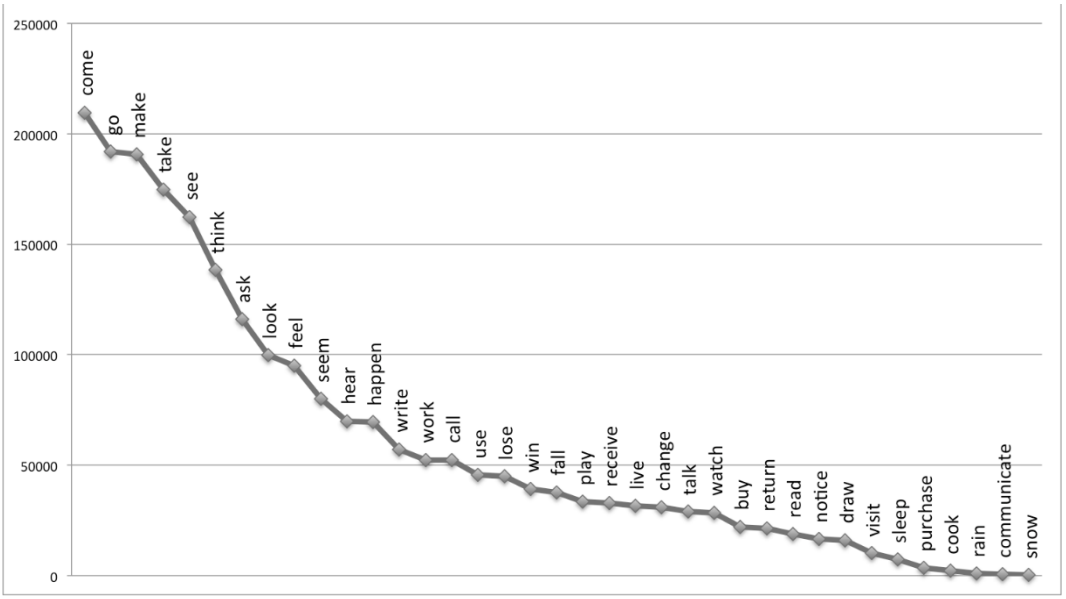


Figure 4.7. *Verb frequency with past tense reference (SP+PP)*

The order of the verbs in descending frequency is presented by Biber et al. (2002) in the following way (the selection used in my experiment): *go, think, see, make, come,* and *take*. It appears that *come* is more commonly used with the past tense reference than other frequent verbs; unlike *see* and *think* whose high frequency is provided by other tenses. In general, my frequency data for the past are in line with the uses in total.

This corpus study also proves the observation that English prefers the preterite (cf. Elsness 2009), especially American English, for which the corpus accounts. It has been found that the proportion of verb forms with past tense reference is, on the average, 86% in the SP to 14%[12] in the PP (Appendix 3). In the SP, the variation is between 53% (the verb *change*) and 99% (*seem*). Figure 4.8 shows the percentage of the verb tokens in the SP. If read reversely, it indicates how often the verb is used in the PP (from 47% in the case of the verb *change* to 1% in the case of *seem*). It confirms that verbs have a bias for a particular tense form, which should be statistically controlled for in a psycholinguistic experiment.

[12] This number could be underestimated due to the difficulty of searching for the PP, i.e. disambiguating the PP from other functions of the past participle.

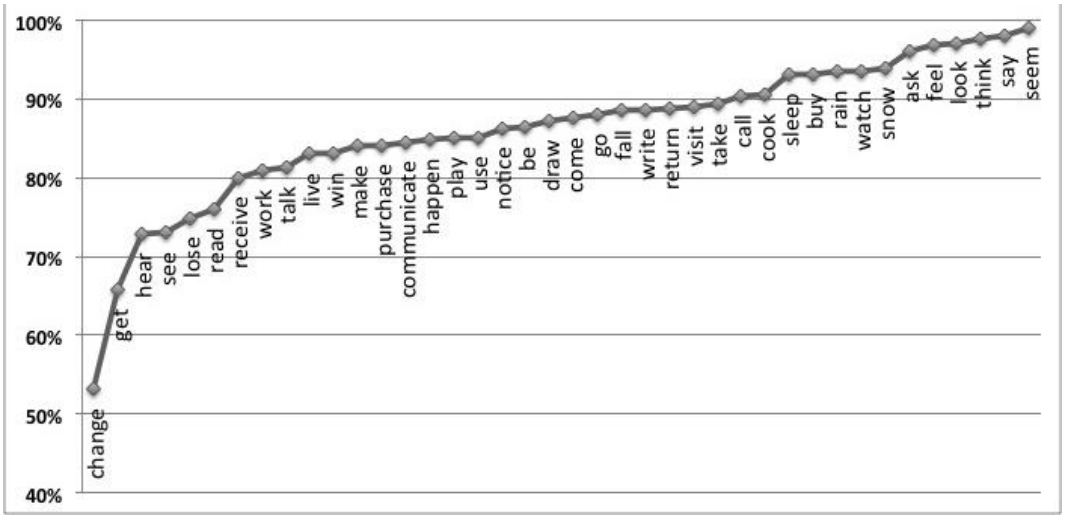


Figure 4.8. *The proportion of the Simple Past to the Present Perfect*

Broadly speaking, the frequency of uses in the SP and the PP is a logical consequence of the verb meaning. Taking, for example, the extremes: *has changed* is much more common than *has seemed* due to the dominance of the perfective (telic) meaning of the verb *change* and the imperfective (atelic) meaning of the verb *seem*. Out of the most frequent English verbs in general, *get* and *see* incline towards the PP more than others, whereas *think* and *say* strongly favor the SP. Other frequent verbs – *be, come, go, make,* and *take* – are around the average.

For the present experiment, I extracted not only the absolute frequencies of PPs and SPs of the verbs but also their use with particular temporal adverbials. The review of previous corpus-based studies (Schlüter 2006) shows that temporal adverbials are used in fewer than 50% of all PPs, with results ranging between 45% for American English (AmE) and 29% for British English (BrE). I searched for my selection of verbs in the context of such adverbials as *yesterday*, *ago*[13] and *last week*, *last year, last month*[14]. Found together with the SP, their number accounted for the matching adverbials. For the number of mismatching examples, I looked for these adverbials following the PP[15]. Correspondingly, the PP forms found with *lately*[16] and *in the last week*, *in the last year, in the last month*[17] in one clause were taken as the number of

[13] Formula: search for "yesterday|ago" in the context of verb.[vvd] 5;0 or verb.[vvn] 5;0.
[14] Formula: search for "last week|year|month" in the context of verb.[vvd] 5;0 or verb.[vvn] 5;0.
[15] Considering the wide use of the past participal in functions other than the PP, extracting the PPs with particular adverbials could not be done automatically in one formula, especially in the context of five words. Therefore, the PP-uses of the verbs with the mentioned adverbials were counted manually out of the first thousand of tokens and this number was multiplied by the number of thousands in the initial result.
[16] Formula: search for "lately" in the context of verb.[vvd] 5;0 or verb.[vvn] 5;0.
[17] Formula: search for "in the last week|year|month" in the context of verb.[vvd] 5;0 or verb.[vvn] 5;0.

correct uses. The same adverbials found together with the SP were considered wrong (Appendix 4).

As previously mentioned, the proportion of non-standard uses of adverbials with the SP is very small – 0.45% on the average. The leading verbs here are *watch*, *notice* and *seem* scoring over 2%. However, PPs are more often found with 'mismatching' SP adverbials – in 11.14% on the average. The leader is the verb *say* with 52.08%. Such examples as *has said yesterday* can be attributed to the "hot news" perfect. Besides "current relevance", this PP function is "pragmatically motivated – it marks the past situation as salient due to its surprise value" (Schwenter 1994). Other verbs whose frequent use of the PP with an SP adverbial can also be found in the press are *win* (42.86%), *ask* (40%), and *make* (27.45%).

(4.7) a. *I mean, it should be a state that George Bush* ***has won*** *two weeks* ***ago****. (CNN_LiveSat)*

b. *We'****ve asked*** *for it two weeks* ***ago*** *and we have yet to receive it. (MSNBC_MeetPress)*

c. *He went on to say,' I'****ve made*** *mistakes years* ***ago*** *and I've learned from those mistakes.' (NBC_Today)*

Elsness 1997 reports a few cases of PPs with the SP time specification. Most of them are "confined to the science category and are particularly frequent in BRPRINT [printed British English] science" (op. cit.: 130). These are examples of temporal specifiers referring to a particular publication (the year) or to a preceding section of the same text (number of the page, table or chapter). The fact that the time reference is given in parenthesis indicates that the relation between the verb form and the temporal specifier is different in this case. One single example of straightforward use of a mismatching adverbial was found in a spoken corpus: "I mean I've sent the first letter about three months ago…"

While my corpus study showed that temporal adverbials are rarely used with the other tense form, it was demonstrated in the previous section that mismatching adverbials, on aggregate, do not increase the processing times, i.e., they do not behave as other 'real' errors (discussed in Chapter 5). It is therefore interesting to notice how verb form frequency affects the perception of non-standard (disfavored) uses. Figure 4.9 illustrates the estimated reading times for matching and mismatching temporal adverbials depending on the verb form frequency.

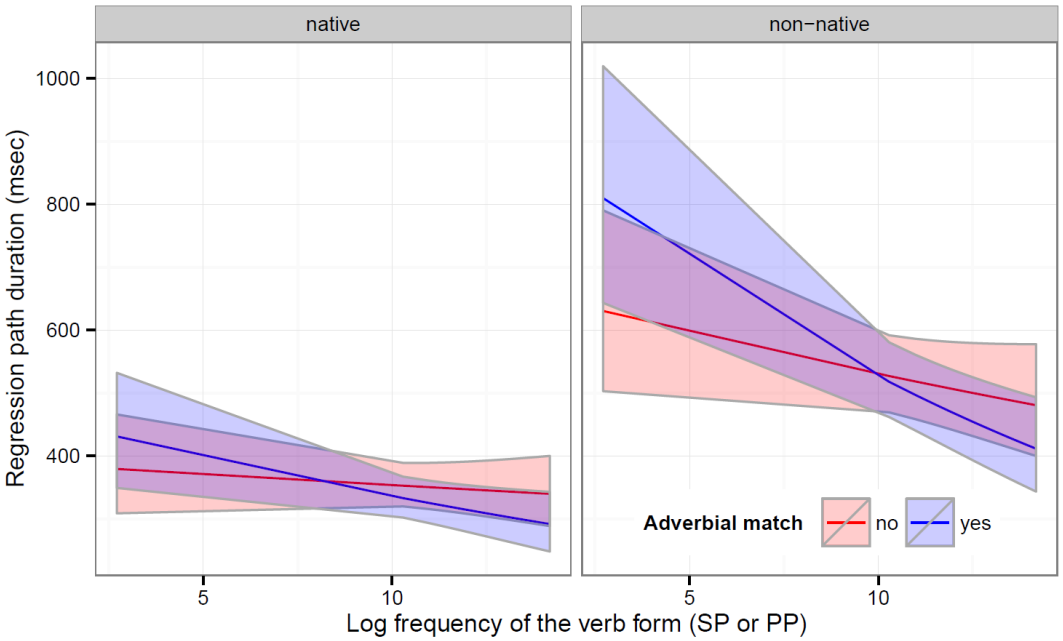


Figure 4.9. *RPD for temporal adverbials depending on the verb form frequency*

As the diagram shows, the slope for processing the matching adverbials is always steeper, which indicates that readers get tremendously faster as the verb form becomes more familiar. However, if these highly frequent verb forms are followed by a mismatching adverbial, the latter requires more processing time. Readers are puzzled seeing a very familiar verb form in an unfamiliar setting.

What stands behind "frequent verb forms"? Below are the names of the verbs in the SP and PP belonging to the most frequent third part of the sample, where the raise for the mismatching adverbials is observed (the right part of Figure 4.9).

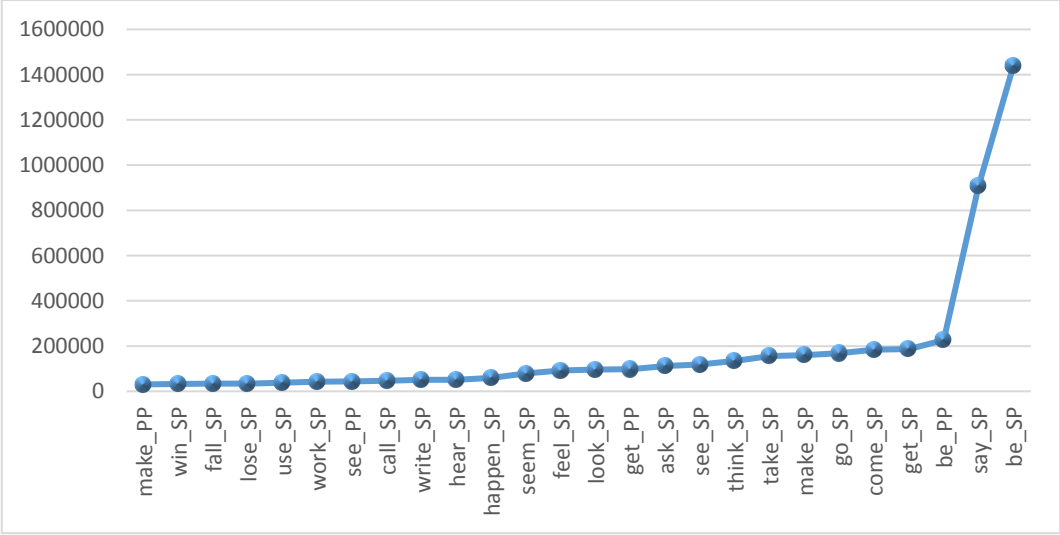


Figure 4.10. *The most frequent verb forms (in SP and PP) in the sample*

As expected, most frequent are primarily SP forms and individual PP forms of common verbs which incline towards PP more than others (*be, get, see and make*). The corpus study resulted in a very low number of these forms used with a mismatching adverbial. The PP forms that can be attributed to "hot news perfect" (say_PP, win_PP, ask_PP – with an attested 'error rate' of over 40%) belong to the left part of the verb form frequency scale. It confirms that these PP forms used with an SP adverbial do not strike as odd and do not complicate processing. Below are the details of the verb form frequency model.

	Estimate	Std. Error	t value
(Intercept)	5.290e+00	5.586e-01	9.469
log(Verb_Form_Frequency)	-3.157e-02	1.171e-02	-2.696
adv_match[S.yes]	1.350e-01	5.204e-02	2.595
Language[S.native]	-3.110e-01	5.604e-02	-5.551
log(adv_freq)	-1.236e-02	8.648e-03	-1.430
rpd.m1	4.875e-05	3.360e-05	1.451
rpd.m2	8.850e-05	2.905e-05	3.046
length[S.short]	1.601e-02	5.206e-02	0.307
log(reg.length)	4.409e-01	4.871e-02	9.052
ia.index	-2.940e-02	1.048e-01	-0.280
trial	-5.227e-05	1.419e-04	-0.368
age	1.733e-03	9.618e-03	0.180
sex[S.f]	4.689e-02	4.136e-02	1.134
haender[S.links]	1.490e-02	5.274e-02	0.283
Language_major[S.no]	8.841e-02	3.627e-02	2.438
clause_type[S.main]	1.441e-02	4.641e-02	0.311
clause_type[S.sub_adv]	-6.543e-02	5.983e-02	-1.094
negation[S.no]	5.503e-02	4.150e-02	1.326
sentence_type[S.question]	-1.452e-01	5.523e-02	-2.630
verb_type[S.irregular]	-2.658e-03	3.011e-02	-0.088
auxil_form[S.contracted]	-1.609e-02	2.696e-02	-0.597
auxil_form[S.full]	-2.075e-02	2.495e-02	-0.832
log(Verb_Form_Frequency):adv_match[S.yes]	-1.495e-02	5.434e-03	-2.751
log(Verb_Form_Frequency):Language[S.native]	9.757e-03	4.931e-03	1.979
adv_match[S.yes]:Language[S.native]	-3.806e-02	4.794e-02	-0.794
log(Verb_Form_Frequency):adv_match[S.yes]: Language[S.native]	2.731e-03	4.988e-03	0.548

Table 4.4. *The effects of the verb form frequency model (using the sum contrasts)*

Also in this analysis, the German learners are generally much slower than the native speakers (t = -5.551), especially in low frequent verbs, but show the same pattern. Verb form frequency proved its significance (t = -2.696) confirming that more frequent items are not only processed faster but also catalyze the processing of the next few words. It is only in this model that adverbial match got a significant effect (t = 2.595) in both subject groups, which demonstrates that mismatching adverbials can add processing cost. However, most important is the significant two-way interaction of verb form frequency and adverbial match (t = -2.751). It is interesting to notice that adverbial frequency has less effect (t = -1.43) on the processing of adverbials in this model than verb form frequency.

There is no indication that adverbial match is perceived differently in L1 and L2 (t = -0.794). As in the previous models, the region located two constituents away complicates the processing of the adverbial (t = 3.046), whereas the immediately preceding region does not have a significant spillover effect (t = 1.451). Again, students studying a language as a major subject, presumably more experienced readers, tend to be faster (t = 2.438), and the adverbials coming in questions require less time (t = -2.630).

Finally, I checked whether the spillover effect of the region located two constituents prior to the adverbial is caused specifically by the verb. This was the case in the sentences of the short condition:

(4.8)	*-2*	*-1*	*adv*	
Short:	*I* |	***found*** |	*good books* |	***last year*** |	*and I…*
Long:	*I* ***found*** |	*very good books* |	*on ancient history* |	***last year*** |	*and I...*

As mentioned before, the effect of the "-2" region (rpd.m2) in both conditions together is t = 3.198. In short sentences, the frequency effect and the "-2" effect on the adverbial come from the same region – the verb. The subset analyses of the short and long conditions for the same model returned the following results.

Condition	**Short**	**Long**
The -2 region	verb	object (adverb + noun, or adverb + prepositional phrase)
rpd.m2	t = 1.864	t = 4.344
log.Frequency	t = -1.805	t = -2.362

Table 4.5. *The effects in the subset analyses*

It is clear that the object has a larger spillover effect and inhibits the perception of the adverbial more strongly than the verb form frequency works to make it faster. In the short condition, we observe two competing influences of equal power coming from the same region. On the one hand, the plain reading times for the verb slow down the integration of the adverbial, and on the other hand, the verb form frequency drives towards faster processing. In long sentences, the frequency drive is even stronger. The fact that, in the end, the slowdown influence of the verb region prevails is likely to be due to the larger number of low frequent verbs in the sample. Thus, the significant spillover effect of the -2 region in the analysis of all sentences together is determined by the object and strengthened by low frequent verb forms.

4.4.3 Perception of adverbials over the course of the experiment

The order of the stimulus sentences in the experiment was randomized, and each observation was marked with an ordinal number ("trial") as the item appeared in the individual list. By analyzing the effect of trial, we can see how matching and mismatching adverbials are perceived in the course of the experiment.

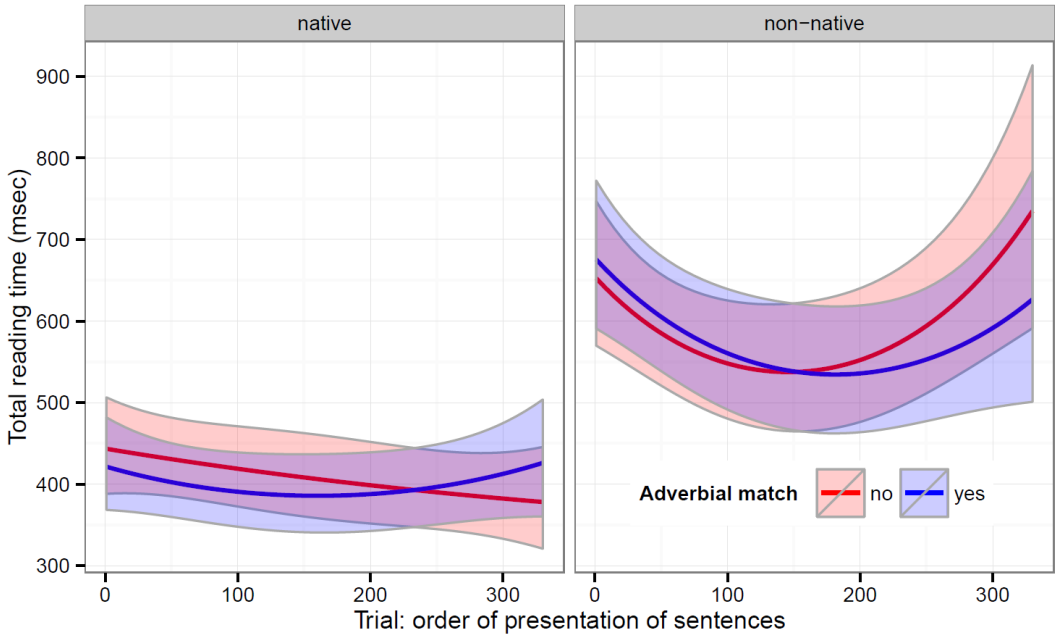


Figure 4.11. *Perception of adverbials over the course of the experiment (polynomial)*

The main (simple) effect of trial demonstrates whether or not there was a slowdown or a speed-up towards the end of the experiment, so-called effects of familiarization or fatigue. In this model, trial has a significant quadratic main effect (poly(trial, 2)2, t = 3.304), which means that, in general, participants become faster in the first half of the experiment (familiarization phase) and become slower towards the end of the reading session (show fatigue). However, in one place, perception deviates from the general pattern: native speakers gradually get adjusted to the mismatching adverbials and process them faster. Non-native speakers, on the contrary, increase their reading times for mismatching adverbials towards the end of the experiment. The 3-way interaction of trial (in the linear form), adverbial match and native language is significant (t = 2.130).

It was also hypothesized that mismatching adverbials may disturb and add processing cost only in the first 10-20 sentences before readers get adjusted and no longer react to them in the rest of the task. It is partially true in native speakers, although, numerically, the greater processing cost of

mismatching adverbials extends even beyond 100 sentences. German leaners, however, in the first half of the experiment make less distinction between the grammatical and ungrammatical uses, but slightly prefer the latter. Broadly speaking, mismatching adverbials are slightly more difficult for native speakers at the beginning of the reading session, and for non-native speakers at the end.

4.4.4 Reaction to mismatching adverbials in English L1-varieties

It has widely been reported that the SP and the PP forms are used differently in BrE and in AmE. Biber et al. (2002) state that BrE uses perfect aspect much more than AmE does, therefore past tense forms appear in contexts where BrE would naturally use the PP. This difference is especially noticeable when the sentences include *yet* or *already*. Huddleston and Pullum (2005) also note that colloquial AmE differs somewhat from BrE. In AmE, *already* prefers the SP, whereas in BrE it is used with the PP. "Americans understand the use of the perfect in such contexts, but use it less frequently" (op. cit.: 159). According to Elsness (2009), the reason for this difference is that, although the reference is clearly to past time, this time is not very precisely defined, which leaves considerable scope for individual judgment. In such cases there appears to be a distinct tendency for AmE to select the SP and BrE to select the PP, so that, on the whole, the latter verb form is more frequent in BrE than in AmE.

Hundt's (2009) data show that the overall diachronic development within both varieties is not significant.

> …the general trend to be observed is that of a slight decrease of PPs in both varieties of English, which is led by AmE. PPs in AmE started out at a lower level in the 1960s, and despite the more dramatic decrease in BrE, written AmE in the 1990s still uses significantly fewer PPs than BrE. So we are dealing with relatively stable regional variation, overall. The evidence based on the tagged version of the corpora thus does not support Elsness' (2009) findings, i.e. a narrowing of the gap between BrE and AmE differences in the use of this grammatical construction (Hundt 2009: 48).

In the elicitation test reported in Elsness (1997), British and American informants judged the acceptability of sentence pairs with the two verb forms and varied temporal adverbials. Both varieties agreed in the use of the SP with the time wholly located in the past (*yesterday, last night*) and found the PP unacceptable in this case.

		BrE	AmE
1a	*I **have seen** John **yesterday**.*	1.4	1.4
b	*I **saw** John **yesterday**.*	5.0	5.0
2a	*I **have seen** Mary **last night**.*	1.3	1.4
b	*I **saw** Mary **last night**.*	5.0	5.0

Table 4.6. *The acceptability ratings* (Elsness 1997: 217)

However, for other adverbials, significant differences in the acceptability ratings have been discovered. For example, speakers of AmE found the use of the PP with *recently* and *just* less favored. For the British informants, *yet* and *already* were perfectly acceptable only with the PP, whereas the American participants found them compatible with both tense forms. "Very generally it can be said that the more distinct the indication of past time, the more marked the preference for the preterite" (ibid.), and most differences between the BrE and AmE are in the acceptability of the PP.

Canadian English (CanE) has often been reported to stand between BrE and AmE in the development of past tense markers. According to Yao and Collins (2012), on the one hand, CanE follows AmE very tightly in the diachronic path of the PP decline (with BrE remaining most 'conservative'). On the other hand, in the use of the SP or PP with particular adverbials (*already, yet, always, never, ever*), CanE patterns more closely with BrE than with AmE. And in the ratios of the overall use of the SP to PP, CanE groups together with the Australian and New Zealand English, which are found in the middle between BrE and AmE.

The present experiment investigates which of the three English L1 varieties (BrE, AmE or CanE) is more sensitive to mismatching adverbials. As it appeared in earlier studies, US speakers have less clear-cut boundaries between the categories specifying the past time because the same temporal adverbials can be used with both SP and PP, whereas UK readers have distinct preferences. Figure 4.12 shows the reading times for the L1 participants from the UK, US, and Canada compared to the German participants with English as a foreign language.

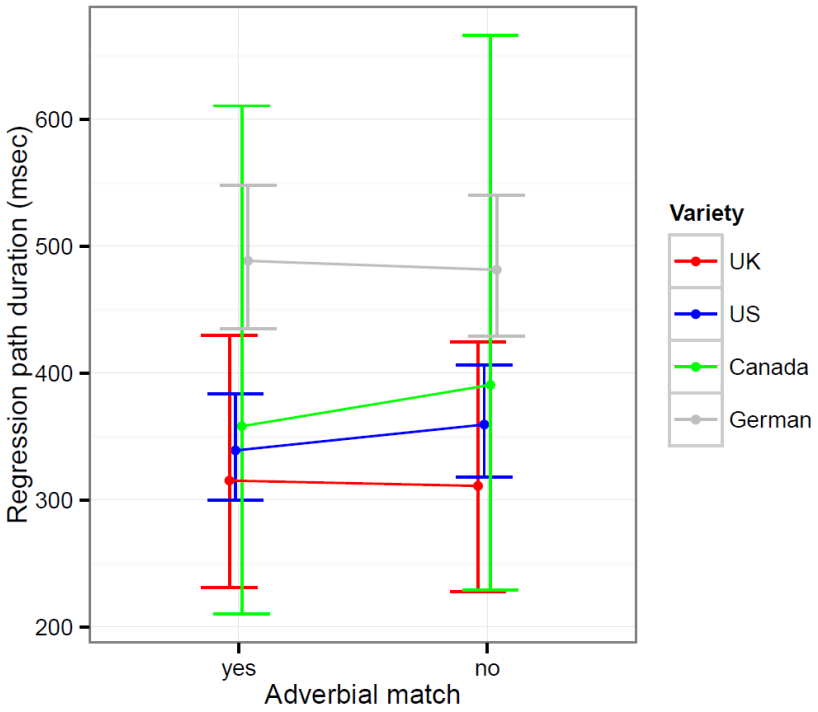


Figure 4.12. *Perception of adverbials in English L1 varieties*

The UK participants not only proved to be faster readers in general but surprisingly showed absolutely no distinction between the matching and mismatching adverbials. The US participants were in fact disturbed by the mismatching adverbials but not to a significant degree. The reaction of the Canadian speakers is not significant either.

Based on the existing literature, I had expected the largest increase in the reading times for the mismatching adverbial in the BrE speakers. Due to the ongoing 'leveling' of the SP and the PP with some adverbials in AmE, I had expected a very small or no increase in the reading times for the American informants. And I was sure to find the Canadian participants in between. However, there appeared to be no significant difference between BrE and AmE in the perception of the adverbials with the SP or PP (the fact that Americans were generally slower is not important) and CanE diverging from both of them. It is difficult to explain such a reaction of the CanE speakers to mismatching adverbials other than by insufficient sample: there were only three Canadian informants compared to 12 British and 17 American participants.

On the whole, all English L1 varieties clustered together, compared to L2, and two of them were to a certain degree delayed on the mismatching adverbials, but neither of the interactions is statistically significant. The similar reaction of BrE and AmE speakers in this task may be explained exactly by the fact that there were no adverbials where these two varieties usually disagree (*yet, already, recently*, etc.).

4.5 Summary

The pedagogical practice of teaching English as a foreign language maintains the rules of combining particular temporal adverbials with verb forms in the Simple Past or the Present Perfect. The violation of these rules, especially for such clear-cut cases as the adverbials locating the action wholly in the past, is usually referred to as a tense error, for example:

> …it is often difficult consistently to locate the error in what the learners have said or written, or in what they should have written. Take the error in *We *have visited London last weekend.* We have a tense error here… (James 1998: 93).

This practice has good grounds. In a judgment study by Elsness (1997), sentences like *I have seen John yesterday* and *I have seen Mary last night* were unanimously rated extremely low by both BrE and AmE speakers. In addition to that, my corpus study confirmed the existence of clear preferences for the SP and PP forms to be used with 'matching' adverbials. The proportion of the 'mismatching' adverbials is only 0.45% with the SP and 11.14% with the PP. Yet, these 'unacceptable' and low frequent uses did not complicate real-time processing in either native or non-native speakers. There was no difference in the reaction to mismatching adverbials between the speakers of standard L1 varieties.

The eye-tracking data have shown a simple (main) effect of adverbial frequency and verb form frequency on processing the adverbial. The tendency is that a common verb in the Simple Past facilitates the processing of the adverbial, unlike a rare verb in the Present Perfect. Most importantly, there is a two-way interaction of verb form frequency and adverbial match. Generally, mismatching adverbials do not slow down the reading times, but their difference from the matching adverbials becomes significant only in relation to verb form frequency. The processing difficulty is caused by mismatching adverbials used with highly frequent verb forms.

Summing up the results of this part of the eye-tracking experiment, I have to admit that no effect of distance has been found. The tendency for normal reading is to speed up if the sentence is long. This is what happened to my stimuli in the long condition. There is no indication that distance does not matter for error perception. Rather, it turned out that my error example, in general, did not create any processing difficulty and was not sensitive to the manipulations with distance. The perception of mismatching temporal adverbials corresponded to normal reading.

5. Factors Influencing Error Processing Cost

This chapter aims at discovering effects of **various factors** that may increase the **error processing cost**. The reading material for this part of the **eye-tracking experiment** consists of a mix of L2 **learner errors** and L1 **non-standard features** in English. It is investigated how the presence of an error in isolated sentences affects the **reading behavior** of **native** and **non-native speakers** over the course of the experiment. Attention is paid to the effect of the **L1 variety** and the level of **proficiency in L2**. The error processing cost is analyzed in relation to the formal characteristics of errors (e.g. **size**, **position** in the sentence, **sentence length**, **distance** between the mismatching parts, the **homophone effect**) and linguistic properties of errors (the **typical producer**, the **part of speech**, and the **level of construction complexity**). Both quantitative statistical results for the variables and qualitative results for concrete sentences are discussed.

5.1 Method and experimental design

With such heterogeneous material, you run the risk of not getting anything. But if, through all this noise, you get a significant effect, it's worth a lot.

Lars Konieczny,
my supervisor in psycholinguistics

Error gravity (discussed in Chapter 2) was defined as a subjective impression of one error being more serious than another based on off-line error evaluations. The error processing cost investigated in this chapter is operationalized as the contrast in the reading times between the 'wrong' and the 'correct' sentence (in the area of interest). The purpose of this study is to explain the 'overtime' caused by deviations from Standard written English based on the data from real-time processing.

To investigate different kinds of errors, I built a small corpus of sentences with learner errors and non-standard features (most examples of the latter are authentic and extracted from COCA). Each sentence was paired with its corrected version and coded for a number of hypothesized factors (e.g. producer of the sentence, part of speech, operation, acoustic similarity, graphic novelty, distance in number of words, etc.), assigning the same tag to the

correct and to the wrong item (Appendix 5).

In an experiment with a homogeneous structure of the stimuli (as in the case of mismatching adverbials), results tend to get significant even with a modest number of observations. However, to explore naturalistic materials with a lot of variation, the diversity should be outbalanced by a rich amount of data. For this purpose, I analyzed over 14000 observations (each of the 56 subjects read 281 items[18]), compared to the 2200 observations for the mismatching adverbials. The stimuli for the error-mix part of the experiment were presented together with the items for the Distance Hypothesis. For details on the procedure see section 4.3.

The sample of the stimulus sentences consisted of:

a) examples of non-standard language of native speakers of English (Kortmann and Szmrecsanyi 2004),

(5.1) *a. She **must been** saving these egg cartons for months.* (Item 43a)
*b. A lot of that was, you know, I **gone** on a mission trip with my church.* (Item 55a)

b) typical errors made by learners of English as a foreign language based on negative transfer, e.g. German or Russian (Swan and Smith 1987),

(5.2) *a. I've known her **since** three years.* (Item 118a)
*b. He **very likes** Chicago.* (Item 172a)

c) other learner errors not necessarily caused by interference (Hultfors 1986),

(5.3) *a. They **fighted** bravely in the war*. (Item 240a)
*b. We have a great **deal** of problems.* (Item 276a)

d) typical spelling errors.

(5.4) *a. I had also seen **proffessional** and college-level performances by several musical ensembles and drama companies.* (Item 281a)

*b. The company will enjoy a wealth of free advertising from this **Wendesday's** press coverage.* (Item 285a)

The critical region in this part of the experiment is the particular word or constituent in which the error occurs and can be discovered (in the examples below indicated in bold).

(5.5) *a. Some of the top models in the world aren't **the most pretty** girls, but they have the extra something.* (Item 19a)
*b. Some of the top models in the world aren't **the prettiest** girls, but they have the extra something.* (Item 19b)

*c. I probably wouldn't let him go just because I **wants** him here.* (Item 24a)
*d. I probably wouldn't let him go just because I **want** him here.* (Item 24b)

[18] Initially, there were 290 items, but in the course of the analysis 9 items were removed.

*e. He **became finally** President on third attempt.* (Item 104a)
*f. He **finally became** President on third attempt.* (Item 104b)

*g. "None of it was written down." **"Please?"** "There's no document that says it."* (Item 145a)
*h. "None of it was written down." **"Pardon me?"** "There's no document that says it."* (Item 145b)

Before presenting the analysis of individual predictors and their interaction with other factors, I will start out with an overview of the general results for this part of the experiment shedding light upon the perception of more severe errors (than mismatching adverbials) and the reading behavior in the native and non-native language.

5.2 Error processing cost

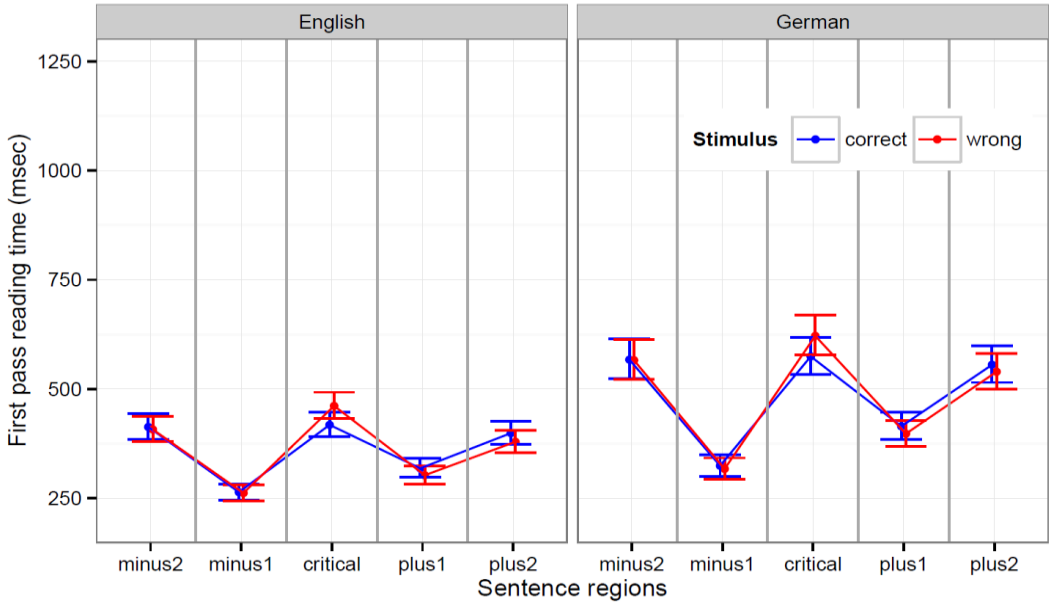
In the error-mix sample, each item was divided into five regions. The critical region containing a deviation from Standard written English (marked with an asterisk below) usually appeared in the middle of the sentence. Depending on the size of the item and the position of the error in the sentence, there were one or two regions prior to the critical region ("minus2", "minus1") and one or two regions following it ("plus1", "plus2"). The size of the regions was statistically controlled based on the number of characters.

	minus2	*minus1*	*critical*	*plus1*	*plus2*
1	*She*	*wouldn't have*	** went*	*back in the house*	*if she was hurt.*
2	*She*	*wouldn't have*	*gone*	*back in the house*	*if she was hurt.*
3	*If you*	*let them*	** to do*	*whatever*	*they want, then I guess they are your friends.*
4	*If you*	*let them*	*do*	*whatever*	*they want, then I guess they are your friends.*

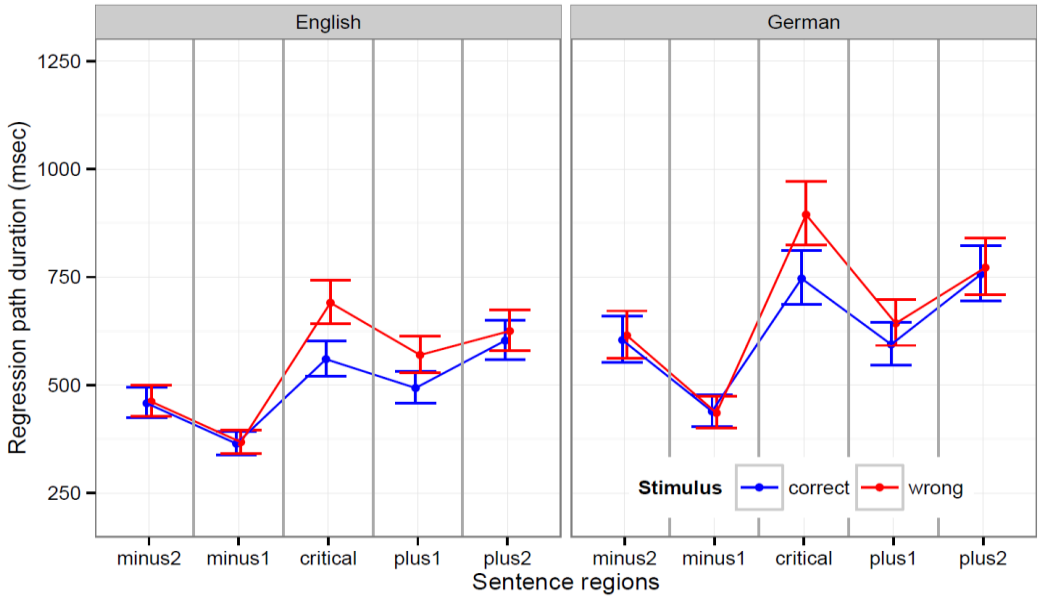
Table 5.1. *Division of items into regions*

Below are the results of mixed-effects models showing the estimates of the reading times for the sentence regions in three measures (FPRT, RPD and TRT). The sample for this part of the experiment was organized differently from the balanced structure of the Distance Hypothesis with mismatching adverbials, but the tendencies in the results remain the same in that native speakers show fastest reading times in processing correct sentences and German learners are generally slower. The fact that errors indeed cause a latency in the processing times is confirmed in Figures 5.1 *a, b* and *c*.

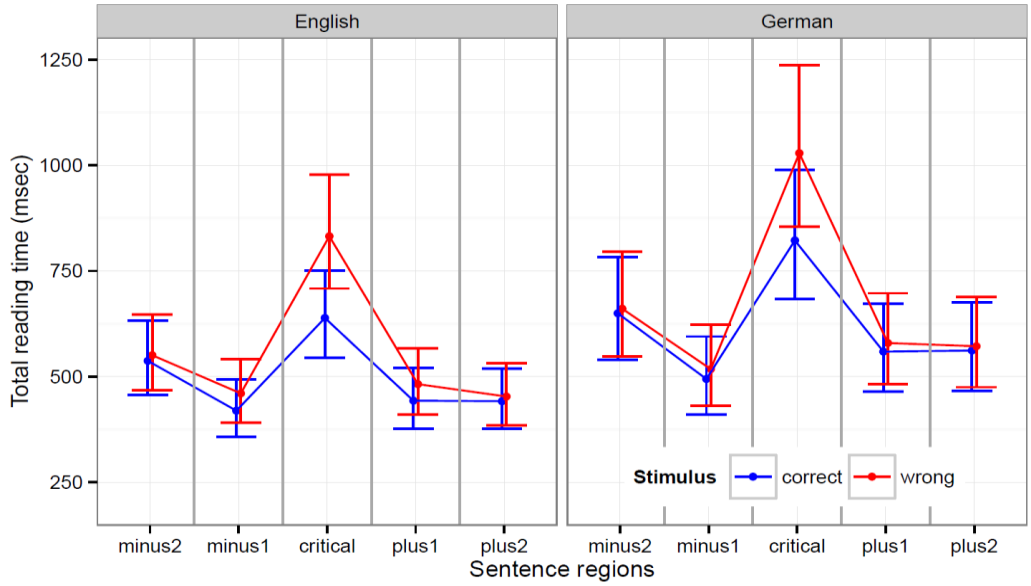
a.



b.



c.



Figures 5.1. *Time measures for the whole sentences*

The correct and wrong items stand significantly apart in the critical region where the target error or non-standard feature was discovered. Readers spend significantly more time on the deviations from the norm already in the first pass. The analysis of the regression path duration shows that wrong items make readers go back and reinspect the earlier parts of the sentence. The critical region has a massive spillover effect, too, so that the words following the error are processed significantly longer than those in the correct sentence.

The list of stimuli included different error types. They can be roughly divided into four categories according to the 'module' (although sometimes it is difficult to draw a clear line between them):

a) grammar – 211 items (10076 observations),

(5.6) *a. And it is good to be back with your friends and get **catched** up with school again.*
*b. He has an **own** company.*

b) lexical domain – 45 items (2091 observations),

(5.7) *a. Borodin is one of the Russian **compositors** well-known around the world.*
*b. Can I **become** a glass of beer, please?*

c) pragmatics – 7 items (335 observations),

(5.8) *a. *"Here's a book for you." "Thank you." "**Please**." "When do you need it back?"*
*b. "Here's a book for you." "Thank you." "**You're welcome**." "When do you need it back?"*

*c. *"Pass me the salt, please." "**Please**." "Thank you."*
*d. "Pass me the salt, please." "**Here you are**." "Thank you."*

d) spelling – 18 items (919 observations).

(5.9) *a. He is **an** useful member of the team.*
*b. Stores that sell junk food, soda and **bier** should offer healthy options or pay a junk-food tax.*

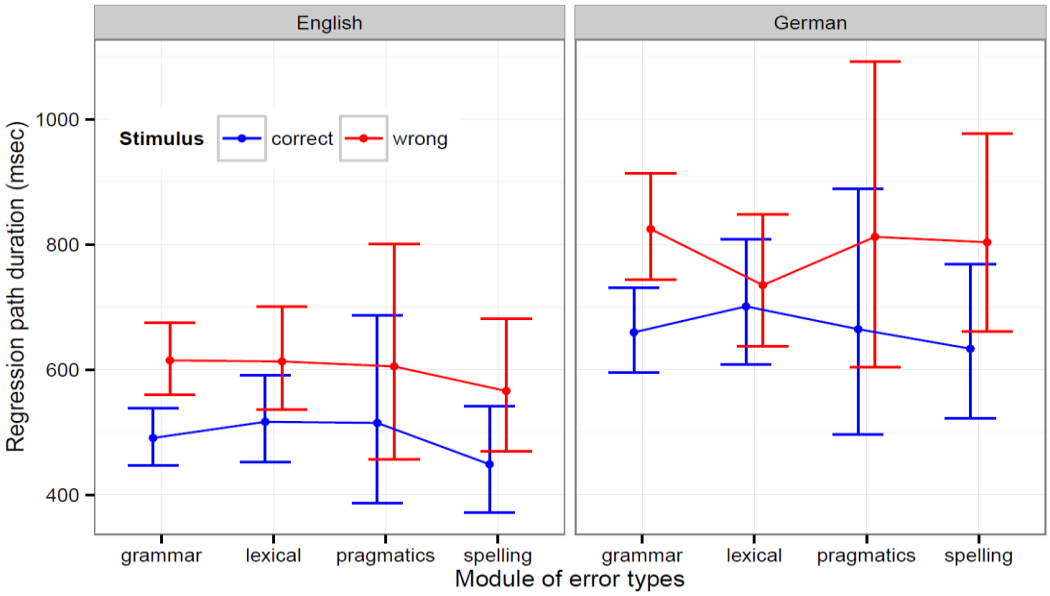


Figure 5.2. *The processing cost of the modules*

Figure 5.2 shows the differences in the processing cost for the four modules demonstrating that grammar and spelling represent the most costly error types in the sample for both native and non-native speakers of English. Lexical errors were clearly detected only by the native speakers, whereas the German learners reacted to them only with a slight increase in the reading times and did not significantly distinguish them from the correct stimuli. The category of pragmatics appeared to be much smaller than the other categories, that is why the conclusion that pragmatics errors do not complicate processing at all should be drawn with caution. On the one hand, it could be for a technical reason of having very few stimuli that this error type produced an insignificant result (although, in other analyses, some errors show a significant effect even with 7 items). On the other hand, indeed, this kind of errors may not strike as particularly wrong but is, in fact, only sensed as odd, and, therefore, such inappropriate uses do not produce any processing difficulty.

Errors generally require additional reading time ('overtime'), compared to the correct sentences. They are processed at the place where they are discovered, cause regressive saccades and complicate the processing of the next few words. I say "generally", because this is not always the case, at least in the list of sentences used in this experiment. In Figure 5.3 below, the items were sorted according to the 'overtime' at the critical region, i.e. the mean difference between the reading times (RPD) for the wrong and correct sentence.

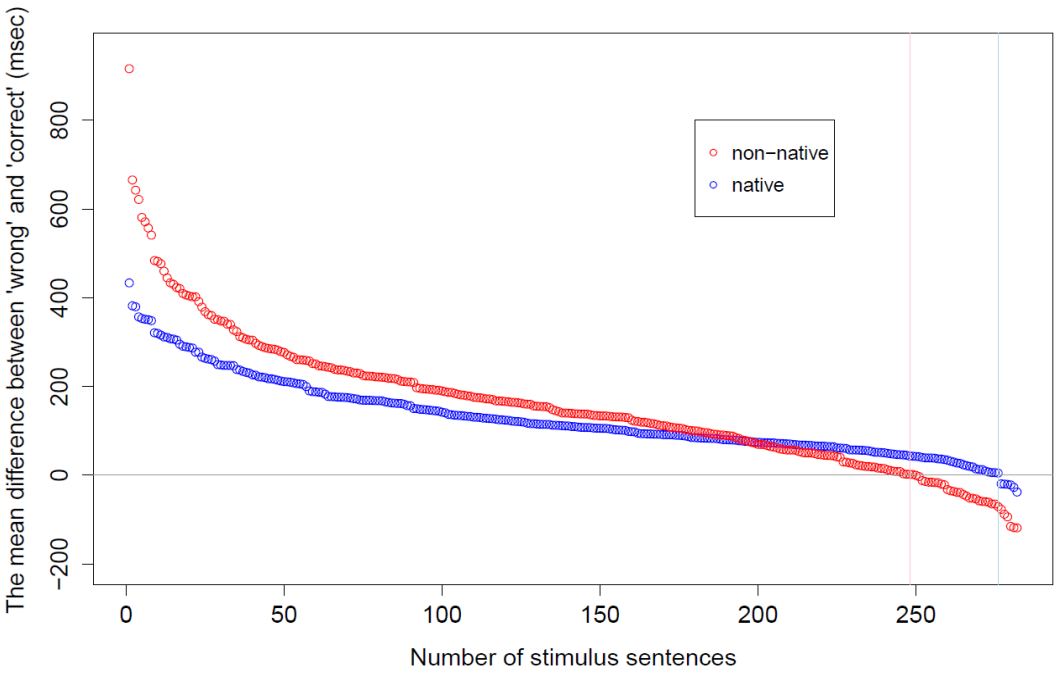


Figure 5.3. *The sorted overtime for the stimulus sentences*

The values on the y-axis are based on the best linear unbiased predictions (BLUPs) used in mixed models for the estimation of random effects, in our case,

of each item. The zero level on the y-axis indicates that the critical region in the wrong sentence takes the same amount of reading time on the average as the corresponding critical region in the correct sentence. The items with a positive value, located above zero, show how much additional processing time the error requires. Figure 5.3 illustrates that the majority of learner errors and non-standard features in this sample behaved as expected from errors, that is, caused a latency in the processing time. However, in the items below zero it is the correct sentence that imposed an additional processing cost, and the German learners showed such a reaction to more stimuli.

Below are the sentences that required the greatest effort on the part of the native speakers (see Appendix 6 for the full table).

Rank	Sentence	Overtime (msec)
1	*I heard **of his appointing** headmaster.*	*434*
2	*"Was that Eric Sears I hear **there, or?" "This** was Jim."*	*382*
3	*She was standing alone, **beside her** with rage.*	*380*
4	*She doesn't know what she **will** at all.*	*357*
5	*You know, he **useta get** this bad press all the time…*	*353*
6	*When I **five years ago visited London**, I could hardly speak English.*	*352*
7	*They fix refrigerators, washing machines, TV sets and other types of **technique** every day.*	*351*
8	*When **you went** there?*	*348*
9	*Did you tell her where **your going** tonight?*	*321*
10	*I am **sure of that he** will come.*	*320*

Table 5.2. *Items with the greatest processing cost for native speakers*

The non-native speakers had a hard time reading the following sentences.

Rank	Sentence	Overtime (msec)
1	*You know, he **useta get** this bad press all the time…*	*916*
2	*The man **I met's** girlfriend is a real beauty.*	*665*
3	*The **in Britain with excitement awaited festival** will take place next week.*	*643*
4	*He **needs not** come.*	*622*
5	*When **you went** there?*	*581*
6	*I don't know what improvements **that are being** planned.*	*571*
7	*This is my **brother's-in-law** cap.*	*557*
8	*"Pass me the salt, please." "**Please**." "Thank you."*	*541*
9	*I never **would of thought** that he'd behave like that.*	*484*
10	*Next morning we went and shook hands at the Ford factory **for to thank** them for their support.*	*482*

Table 5.3. *Items with the greatest processing cost for non-native speakers*

The ten sentences where the critical regions (marked in bold) caused the greatest processing cost are different for native and non-native speakers, except for one sentence with a failure to produce a question (*when you went*

there?) and another sentence with wrong spelling based on phonetic similarity (*he useta get*). However, both language groups have other sentences of the similar error type (*where your going* and *would of thought*). Besides that, both L1 and L2 readers were delayed at massive constituent reordering (*when I five years ago visited London*, *the in Britain with excitement awaited festival*) and insertions (*sure of that he will come*, *what improvements that are being planned* and *for to thank them*). At the same time, only the native speakers were extremely puzzled by inappropriate lexical uses (*what she will at all* and *technique* instead of *equipment*), while the learners paused more over the wrong attachment of the possessive *'s* (*the man I met's girlfriend, my brother's-in-law cap*). Generally, the native speakers paid more attention to the wrong usage of particular grammatical constructions and other error types resulting in unclear meaning (in Table 5.2: *appointing headmaster, or?, beside her with rage, what she will at all*). Meanwhile, the non-native speakers were more preoccupied with the morphosyntactic reordering in sentences where the meaning was relatively clear (in Table 5.3: *the man I met's girlfriend, the in Britain with excitement awaited festival, he needs not come*).

A number of sentences in Figure 5.3 have a negative value, which means that their correct version took longer to read on the average than the error. These were 6 sentences for the L1 speakers and 33 sentences for the L2 speakers. The 'bottom' items for the native speakers are presented below.

Rank	Sentence	Overtime
277	*Do you **have brother** or sister?*	*-19*
278	*Patients **are wanting** a lot of personal attention.*	*-20*
279	*Everybody is invited to come here at **18 o'clock** today and then again tomorrow.*	*-21*
280	*Though the book **has finished**, he continues to find inspiration for it in the most unlikely places.*	*-22*
281	***My eyes are running in different directions**, there's so much of everything.*	*-27*
282	*"Can you speak German?" "**Yes, of course.**" "Great!"*	*-38*

Table 5.4. *Wrong items processed faster than their correct versions in L1*

The German learners preferred the following sentences to their correct versions.

Rank	Sentence	Overtime
273	*She had her radio **beside herself** and it played her favorite song.*	*-61*
274	*She has very limited **knowledges** of German.*	*-65*
275	*When we go to the party on Saturday, let's **bring** a bottle of wine.*	*-65*
276	*They don't know that **tommorrow** may be a better day.*	*-71*
277	*We have to live in **the society** after all.*	*-77*
278	*It is **a hard work** to write a book.*	*-88*

279	*She **is gone out** and will come back in an hour.*	*-94*
280	*I **have seen Mary yesterday** and she looks good.*	*-115*
281	*"Can you speak German?" "**Yes, of course.**" "Great!"*	*-118*
282	***My eyes are running in different directions**, there's so much of everything.*	*-119*

Table 5.5. *Wrong items processed faster than their correct versions in L2*

The qualitative analysis of the items in Tables 5.4 and 5.5 demonstrates that there was also some agreement between the native and non-native speakers as to which errors had a negative processing cost. The subject groups share two sentences at the bottom of the table. *My eyes are running in different directions* is a calque translation of a Russian idiom that was analyzed against the control sentence *I can't concentrate on one thing, there's so much of everything*. Some typical errors of Russian learners of English were included in the sample to put the native English-speaking and the German participants in equal position and test their reaction to errors of 'third parties'. As previously mentioned, the sample contained multiple examples of non-standard features, and presumably, the L1 participants would have an advantage processing them. Another group of stimuli consisted of typical errors of German learners, and the non-native speaking participants would have an advantage there. Deviations from Standard English that could not be attributed to L1 or explained through transfer from German is a perfect condition to test error processing when neither of the groups 'have a clue'. The meaning of the *eyes running in different directions* is quite transparent, yet this particular word combination is not part of the English language. For example, a Google search of the exact phrase "my eyes are running in different directions" returned no result, whereas "I can't concentrate on one thing" enjoyed over 45 000 uses. And it is the latter that required more processing time from the participants.

The other stimulus considered inappropriate from the point of view of pragmatics is the answer *Yes, of course* analyzed against the neutral *Yes, I can*. It is also believed to be a typical feature of Russian learners due to transfer (Swan and Smith 1987: 159). In Russian, the word *конечно* [kaneshna], widely recognized as equivalent to "of course", does not have the connotation of "as you know" or "it goes without saying", but rather serves only to strengthen the "yes". James (1998: 223) discusses Russians' overindulgence in *of course* based on his own experience as carrying the "irritation potential". However, the 'inappropriate' stimulus required less processing time than the correct and neutral variant. In the cases where both participant groups had no advantage of understanding the background of the grammatically correct but inappropriate item, native and non-native speakers showed the same pattern.

It is surprising to find another inappropriate use with a negative value: *18 o'clock* was processed faster than *6 p.m.* by native speakers while they should not have expected the use of "military time". Based on such examples, it is hard to escape a conclusion that pragmatics violations do not necessarily increase reading times. However, the reason for this conclusion in the present experiment may be the lack of context: it is not possible to evaluate the appropriateness of the utterance in just one or two sentences without any reference to a particular situation or a picture.

In non-native speakers, among the errors processed faster than their corrected versions are also indications of German transfer (*knowledges*, *she is gone out*, *I have seen Mary yesterday*), which confirms that familiarity with particular constructions may facilitate processing. The perception of errors with transfer from German is elaborated on in section 5.5.1.

Although it is curious to see which wrong items were processed easier than their controls, and we may be wondering about the reasons for each particular case, the results should not be interpreted too far. The variable of stimulus has a strong significant effect, which proves that whatever is defined as an error from the point of view of prescriptive grammar tends to be processed longer. The items whose processing cost turned out to be negative, globally viewed, are not more than noise.

In the whole sample of 282 sentences, the processing costs for native versus non-native speakers based on the RPD time measure correlate moderately, $r = 0.48$. It demonstrates that the reading effort for the same sentences is to a large extent different in L1 and L2.

5.3 Reading behavior in L1 and L2

5.3.1 Perception of errors over the course of the experiment

This model checks the effect of trial in the mixed sample to see how learner errors and non-standard features are perceived in the course of the experiment.

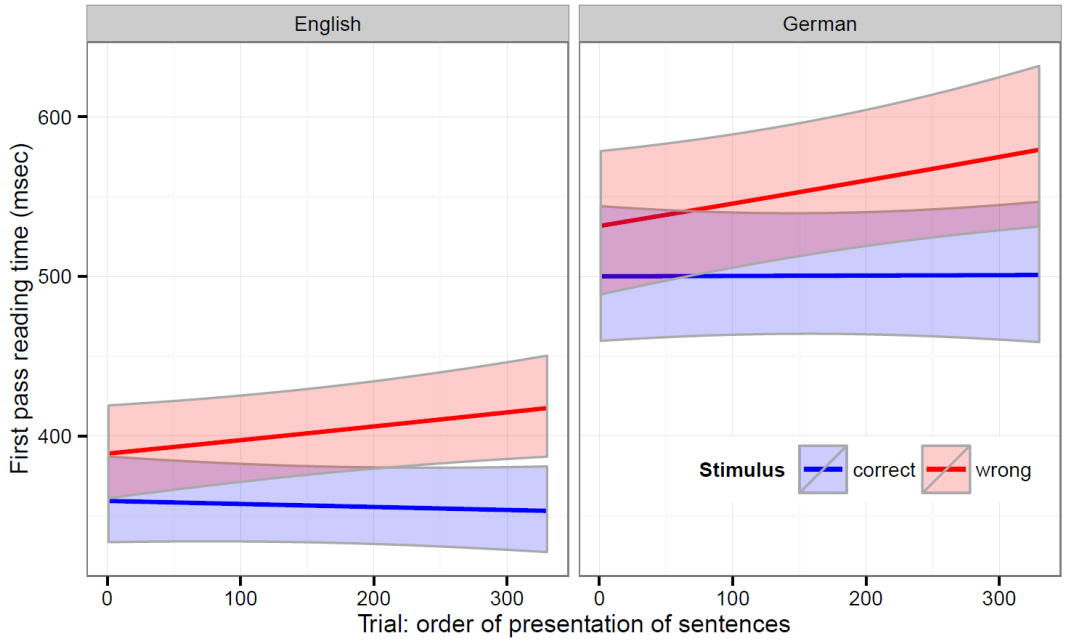


Figure 5.4. *Reaction to the stimuli over the course of the experiment*

Figure 5.4 shows that processing of correct sentences over the course of the experiment either becomes slightly faster (native speakers) or remains at the same level (non-native speakers), whereas both speaker groups demonstrated longer reading times for the erroneous sentences towards the end of the experiment. It is interesting to notice that the native speakers made a clear distinction between right and wrong from the start, unlike the non-native speakers. Generally, the difference between the perception of correct and wrong stimuli over the course of the reading session is significant (the interaction of trial by stimulus, t = -2.236). This effect appears only if we analyze one time measure – the first pass reading time. In other time measures, both correct and wrong stimuli have a simple familiarization effect and no interaction with the stimulus (trial, sum contrast, t = -7.731 (RPD), t = -14.861 (TRT); trial : stimulus, t = 0.345 (RPD), t = 1.358 (TRT)). Despite the familiarization effect, errors are still processed significantly slower than their corrected versions (stimulus, RPD: t = -10.540, TRT: t = -14.896). All of this means that, over the course of the reading session, errors increasingly add processing cost only at the first encounter of the deviance in the sentence, but do not increase regressive movements either to the previous words or from the next words back to the error.

5.3.2 Regressive movements

When do readers return to the previous parts of the sentence? To answer this question, I extracted the regressive paths (RPD minus FPRT >0) and found that the participants went back to the previous parts of the sentence in 25% of the observations. Generally, in reading English, right-to-left movements along the line or movements back to previously read lines are found in approximately 10-15% of the saccades (Rayner 1998: 375).

In the logistic regression presented in Figure 5.5, the probability of regressive saccades decreases over the course of the experiment for both subject groups (trial, $p < 0.001$), but more rapidly for the native speakers (trial : native language, $p < 0.05$). Interestingly, native speakers made more regressions at the beginning of the reading session than non-native speakers did. The L1 readers consistently moved backward from errors more than from the corrected versions, whereas the L2 readers made no distinction between the two categories towards the end of the experiment (beginning to return in more correct sentences and less wrong ones). On the whole, the effect of error remains highly significant ($p < 0.001$) but its interaction with native language is not.

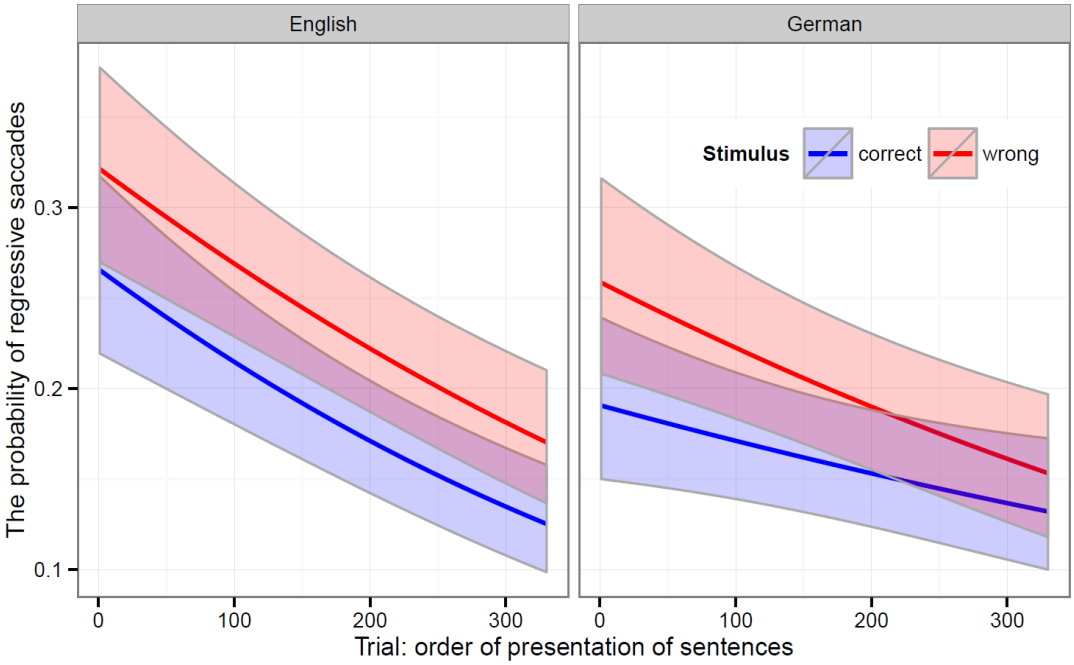


Figure 5.5. *Regressive paths over the course of the experiment*

As for the duration of regressive paths, Figure 5.6 demonstrates that both subject groups spend significantly more time to repair the sentences containing an error (t = 5.551). Native speakers prove to get adjusted to errors better than non-native speakers as the former needed less and less time to recover from errors over the course of the experiment (trial : native language, t = 1.829).

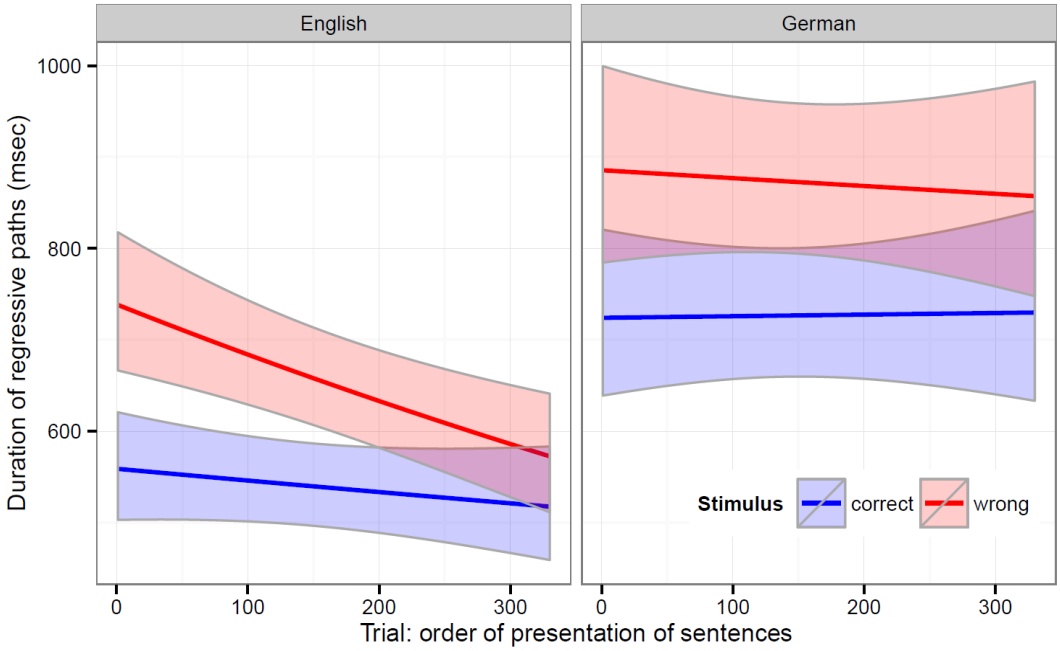


Figure 5.6. *Duration of regressive paths over the course of the experiment*

Although it has been confirmed that deviations from the norm cause more frequent and extensive regressions, it is surprising that the number and duration of regressive movements are so large in correct sentences. It is not in line with previous findings that longer regressions of more than 10 letter spaces (which were most probable in this experiment) "occur because the reader did not understand the text" (Rayner 1998: 375). In cases of long regressions, good readers were very accurate in going back exactly to the part of the text that caused them difficulty (e.g. Frazier and Rayner 1982), whereas poor readers showed more backtracking through the text (Murray and Kennedy 1988).

I developed the idea of individual differences and hypothesized that it might be due to the person's reading style that some people go back more readily than others in all circumstances. I calculated the average duration of regressive paths for each participant and colored it in relation to the stimulus (Figure 5.7).

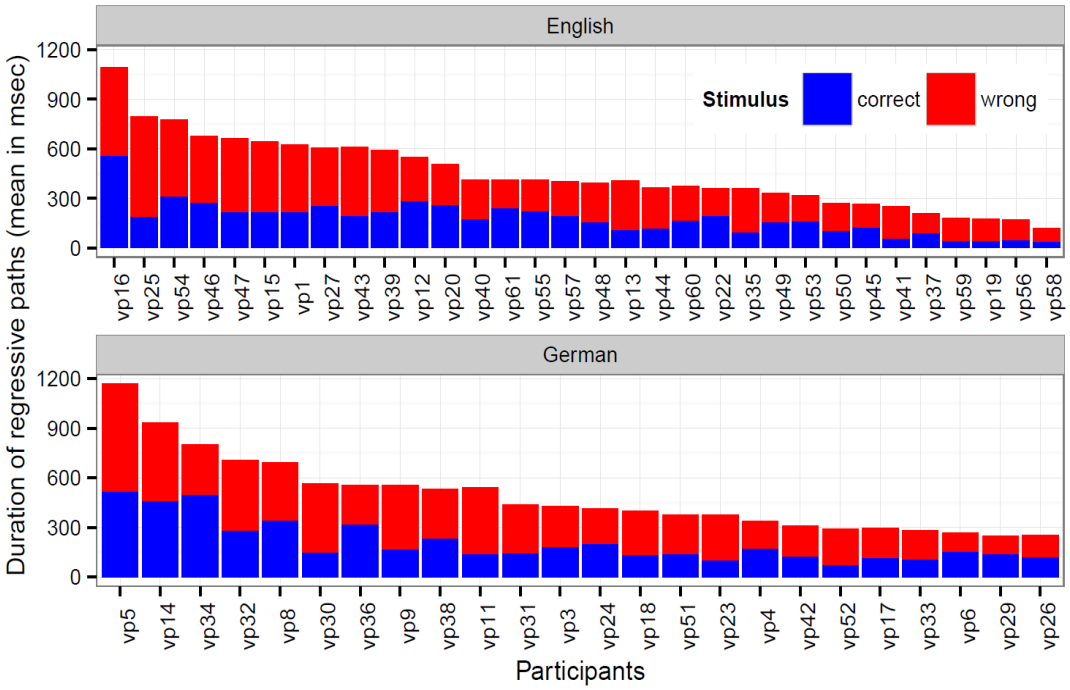


Figure 5.7. *Mean duration of regressive paths per subject*

Now there is no doubt that some readers are greater 'backjumpers' than others, and it is relevant for both native speakers of English and German. Whoever has this 'backjumping' feature in their reading style goes back almost equally to the correct and to the wrong stimuli (for the exception of vp25). Some participants made even more regressions in correct sentences, e.g. vp61 and vp22 among the native speakers and vp34 and vp36 among the non-native speakers. It is not in line with the claim that long regressions occur because readers did not understand the sentence.

5.3.3 Skipping

Another interesting question is how much skipping is done during reading, how much chance there is for errors to remain unnoticed, or whether errors attract more attention and reduce skipping. Previously, word skipping was thoroughly investigated as part of the eye movement research in the 1970s-1980s. It was found that there are systematic differences in what kinds of words get skipped.

> Short words (3 or fewer letters) are much more likely to be skipped than longer words (6 or more letters). Words that are 8 letters or more are rarely skipped, and words that are 6 letters long are fixated most of the time (Rayner & McConkie, 1976). Carpenter and Just (1983) reported that content words were fixated 83% of the time, whereas function words (which are much shorter) were fixated only 38% of the time (Rayner et al. 1989: 24).

Moreover, compared to the content words of 5-9 letters which were fixated 84% of the time, the definite article *the*, i.e. a function word preceding content words, was fixated only in 19% of cases (Rayner and Duffy 1988). Besides that, high-frequency words are skipped more often than low-frequency words when they are not longer than six letters (O'Regan 1979, Rayner et al. 1996).

To answer my research questions, I made a logistic regression taking as the dependent variable all fixations of 0 milliseconds in the first pass reading time. Absence of fixations was found in approximately 5% of the data.

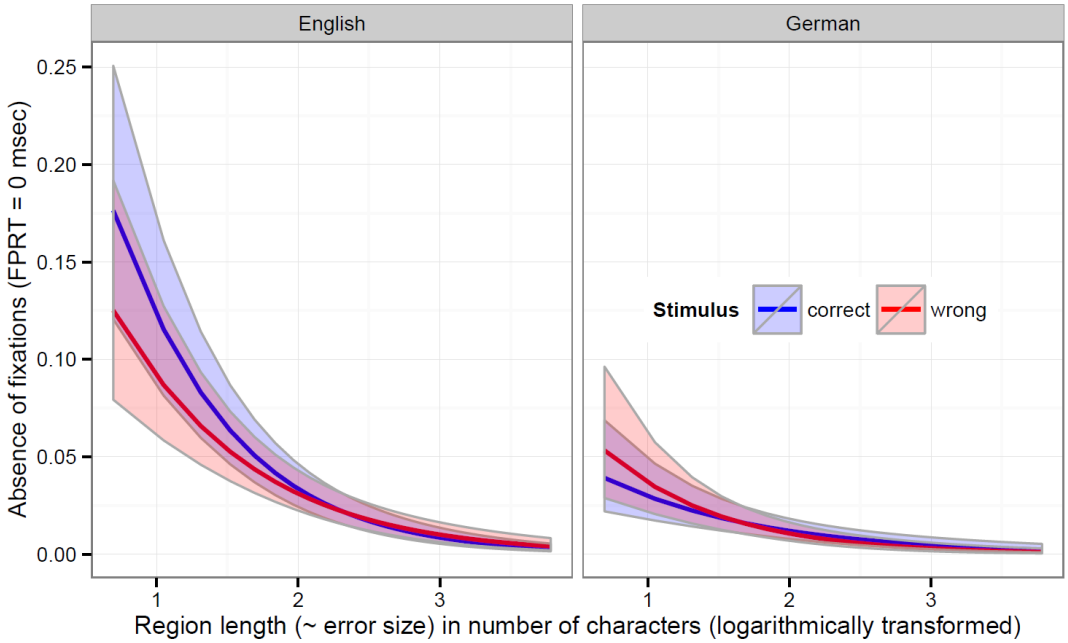


Figure 5.8. *Probability of error skipping depending on region length*

Region length has a significant effect on whether the region will be skipped or not ($p < 0.001$). In my sample, there were regions from 2 to 44 characters, and in the same analysis without logarithmical transformation, the probability for native speakers to saccade over a region of 5-6 characters is less than 5%. Deviations from the linguistic norm reduce the skipping rate in L1 (the three-way interaction of region length, stimulus and native language is significant, $p < 0.05$). The implication of this is that native speakers may overlook small errors located within 2-5 characters in approximately 5-13% of sentences.

Native speakers do more skipping in correct sentences, which is in line with observations that they fill in when it 'reads smoothly'. Studies confirmed that a predictable target word is more likely to be skipped than an unpredictable one (Ehrlich and Rayner 1981, O'Regan 1979). Similarly, high-constraint target words are skipped more than medium- or low-constraint target words (Rayner and Well 1996). Thus, native speakers can automatically repair a sentence

without even noticing the error, and this explains why proofreaders sometimes overlook typographical mistakes.

Non-native speakers read so much slower and more attentively that the chance of their skipping is negligible. They could be perfect proofreaders, but the problem is that non-native speakers may have a different idea of what is correct and what is wrong, which is also reflected in Figure 5.8. As another indication of this, in an evaluation experiment where the participants had to detect errors and rate their gravity, non-native speakers failed to identify some errors but suggested miscorrections of well-formed sentences (Kobayashi 1992). Of course, the higher proficiency level of the judges provided better accuracy in error correction.

5.3.4 L1 variety

The reaction to the mix of errors was also analyzed across the varieties of English and compared to the German learners. Unlike in the case of mismatching adverbials, all groups of participants showed an increase in the reading times for learner errors and non-standard features, but only for the speakers from the US and the German learners this effect is significant (see Figure 5.9).

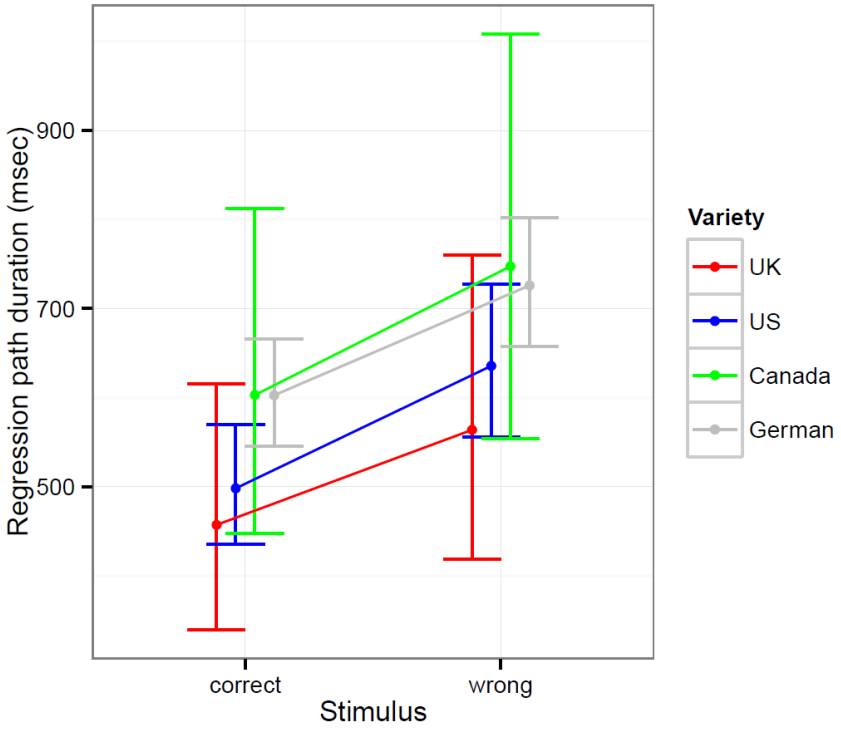


Figure 5.9. *Perception of errors in English L1 varieties*

Surprisingly, the overall reaction to errors in native speakers of British English is not significant. It is likely to be due to the very different results within this group (the confidence intervals are very large, even for 12 informants, compared to the 17 US participants). It probably indicates that some UK readers were either too skillful to be puzzled by the errors or too careless to pay attention. Also the Canadian participants did not make a clear distinction between the correct and the wrong stimuli, but the fact that there were just three of them explains the extra-large confidence intervals. Nevertheless, it is quite surprising that the native speakers from Canada showed mean reading times very close to those of the German learners. The latter proved to be slower than most of the native speakers, and they reacted to the errors with a significant raise in the reading times. Similar to the result with mismatching adverbials, the US participants seem to be most sensitive to errors.

5.3.5 L2 proficiency

It is intuitively expected that the level of proficiency in the foreign language has an impact on error perception. To gather data on language proficiency, all German learners were asked to complete a small English test where they scored from 8 to 20 (Appendix 2). The best possible result of 20 points was achieved only by one German learner. I asked three native speakers to try this test as well, and they all scored 20. The mean result of the German participants is 15.68. In the questionnaire before the experiment, the candidates were requested to assess their English skills on a 10-point scale, and the mean result turned out to be 7.4. On a 20-point scale, it amounts to 14.8, which is slightly under the confirmed result of 15.68. The Pearson correlation of English self-assessment and the result of the English test is $r = 0.636$, which is described as a strong positive relationship. The development of the reading times depending on the level of L2 proficiency is presented in Figure 5.10.

The result of the English test alone proved to be almost significant for the general reading speed of the German participants in that more advanced learners read faster ($t = -1.976$). There also appeared to be a significant difference in how readers at various levels of proficiency process learner errors and non-standard features (the two-way interaction of the English test result and the stimulus, $t = 2.389$).

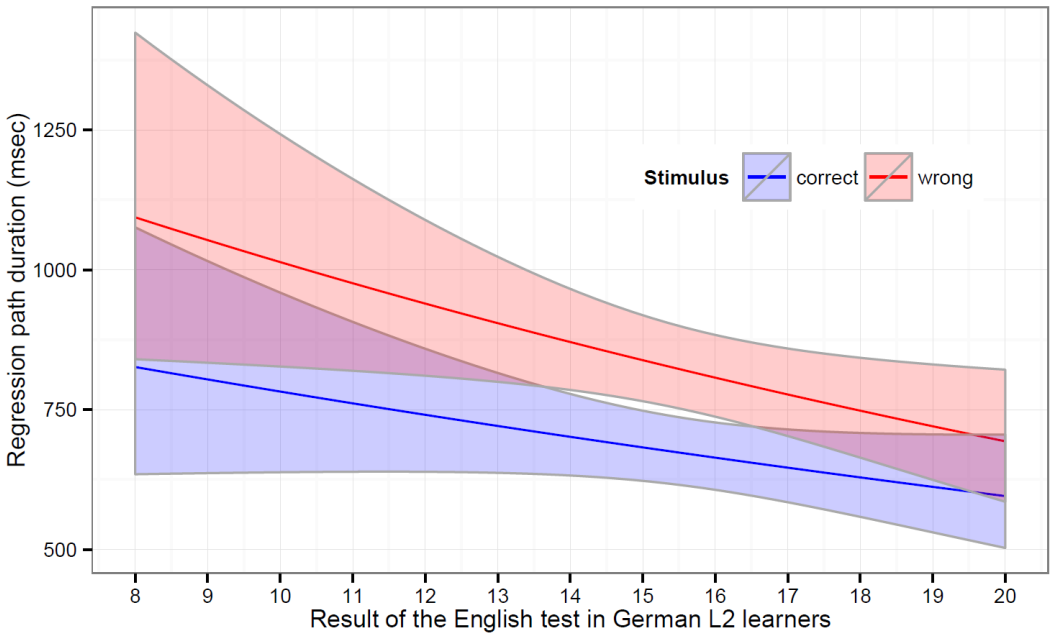


Figure 5.10. *Perception of stimuli depending on the L2 proficiency*

Thus, the higher the level of proficiency is, the faster the errors are processed. Advanced learners are sure to be more familiar with non-standard features and for that reason may need less time for processing them. The large confidence intervals at the result of 20 points are likely to be due to the fact that it was provided only by one person, and not likely that this person made little distinction between the correct and wrong stimuli.

5.4 The effects of formal characteristics of errors

5.4.1 Error size

In this section, I will investigate the reaction to errors depending on their physical, or formal, characteristics. One of them is error size. To operationalize this factor, each item was coded based on the difference between the correct and the wrong sentence, i.e. how much has to be repaired. It is expected that the smaller the error size is, the lower is the processing cost, given that the region length is statistically controlled.

Error size was analyzed in three categories:

1) small – involves repair of a few letters within one word (85 items, 4141 observations);

(5.10) *a. They've been promoting **theirself** as the best women's club in the world.*
*b. Herman was one of those singular individuals for **who** there is no adequate categorical description.*
*c. You remind me **on** your father.*

2) medium – involves change of a whole word or a verb form (115 items, 5415 observations);

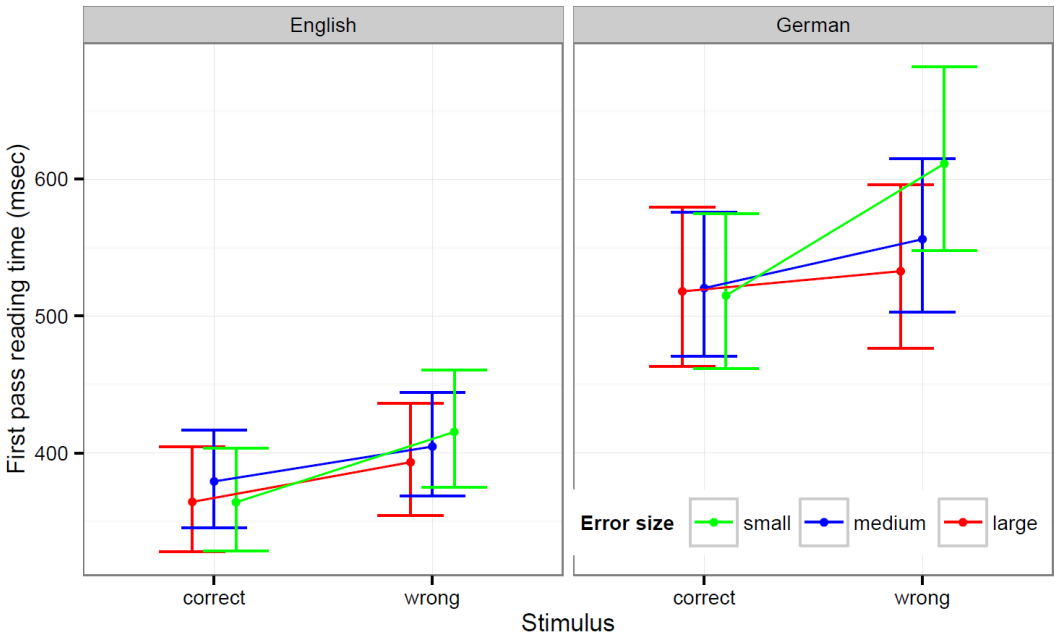
(5.11) *a. Patients **are wanting** a lot of personal attention.*
*b. I'll tell you what we might **should** do about it.*
*c. He was dressed **with** a dark suit.*

3) large – involves change of two or more words, often with reordering (82 items, 3865 observations).

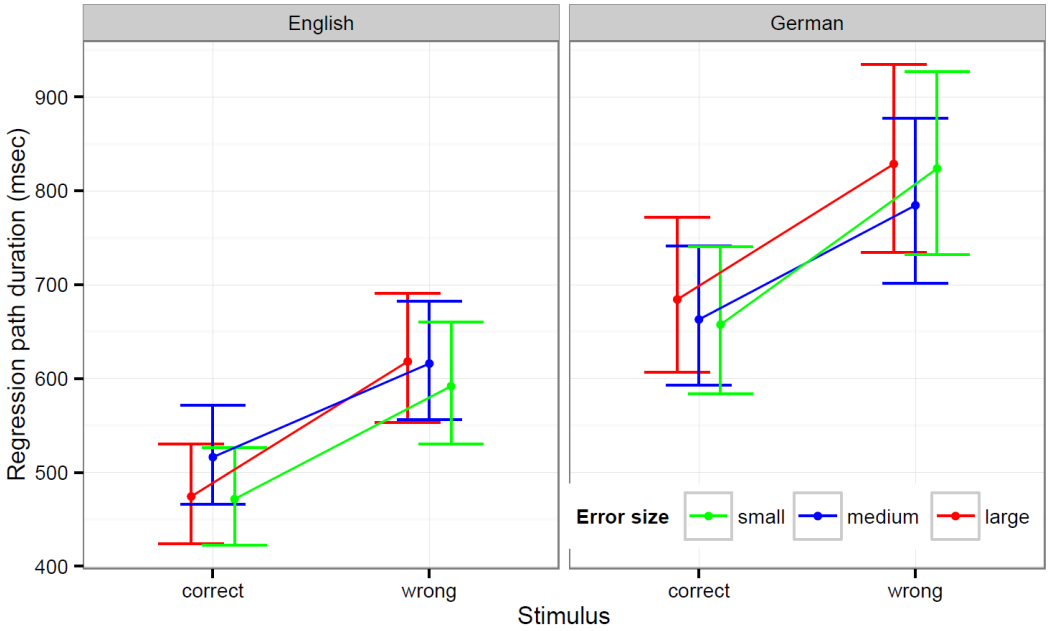
(5.12) *a. "**Were you ever** a smoker?"*
"No, I never smoked, except one misfortune when I was 8."
*b. I mean, he's so wrong in this thing, and **he's like**, oh, the deficit is out of control.*
*c. Did I tell you that my mother **English speaks**?*

Figures 5.11 *a* and *b* show how errors of different sizes are processed in the first pass and whether their processing involves regressive movements. The first diagram illustrates that only small errors involving difficulties in word recognition are clearly detected in early stages of sentence processing. By contrast, more complex errors do not seem to be processed in the same region, but rather immediately send the eyes back to the previous parts of the text, so that the error processing cost can really be captured only in late measures. It appears here and will be confirmed more than once in the following sections that non-native speakers are more sensitive to unusual word images than native speakers are. In this analysis, the German learners were particularly impressed by the small errors which included wrong spelling and non-standard forms.

a.



b.



Figures 5.11. *The effect of error size in two time measures*

The results from the regression path duration (Figure 5.11 *b*) indicate that, in the end, the complexity of repair necessary to correct the sentence does not have any significant effect on the processing cost. In other words, any error causes a significant increase in the reading times, regardless of its size, for both native and non-native speakers of English. Even if it is a matter of repairing just a couple of letters, the error is detected and in the end may cost the same effort as solving more complex grammatical problems. However, this should not be taken as an argument that all errors are equally costly. This analysis is merely an indication that differences in the processing cost are probably due to other factors rather than error size.

5.4.2 Position in the sentence

Not all places in the text receive equal attention from the readers. Therefore, it can be expected that errors located in 'unfavorable' positions in the line may remain unnoticed.

> …it should not be assumed that readers place their fixation to correspond to the beginning of a line. Rather, the first and last fixations on a line are generally 5-7 letter spaces from the ends of a line. Thus, about 80% of the text typically falls between the extreme fixations (Rayner 1998: 375).

For this reason, it is common in psycholinguistic experiments to place the target words in the middle of a line and in no case at the beginning of a line. However,

the purpose of the present study is not only to measure the reaction to errors but also to find out the circumstances under which no reaction occurs. In my sample, there were quite a few items with errors at the beginning of the sentence. Some examples are listed below.

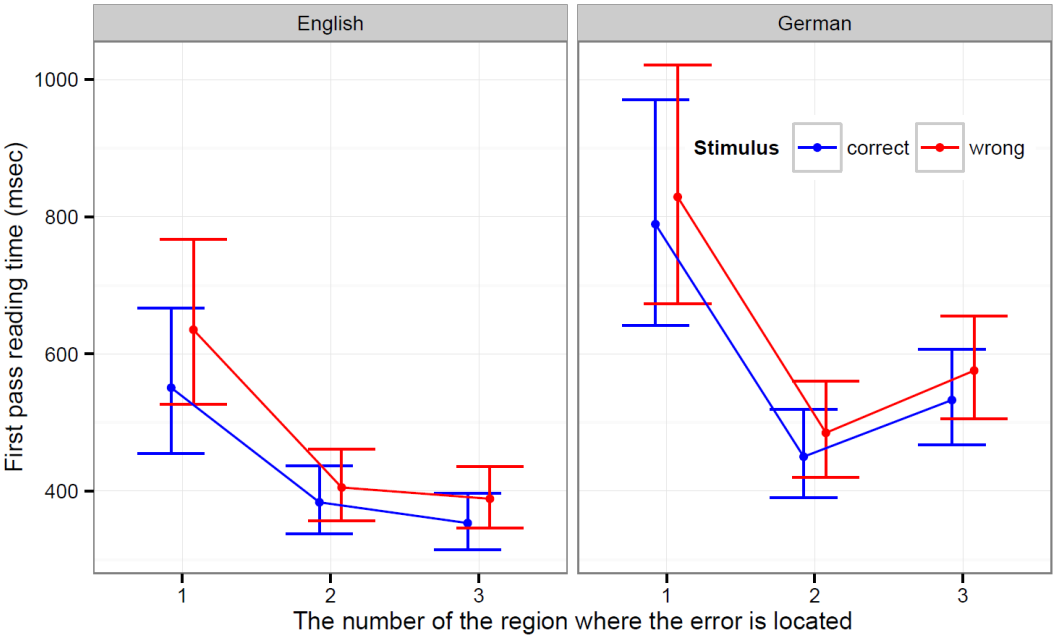
(5.13) *a.* ***Came he*** *to the party yesterday?*
b. ***Became he*** *a dentist?*
c. ***He a*** *good boy, and his eagerness makes him seem younger than his 21 years.*
d. ***How do you call*** *one thousand lawyers at the bottom of the ocean? A good start.*
e. ***It was little else*** *to do.*
f. ***New house is*** *being built near the cinema.*
g. ***No people*** *are here.*
h. ***The most people*** *would agree with you.*
i. ***They're not left*** *school yet.*
j. ***Who of you*** *can speak English?*

As previously mentioned in section 5.2, the stimuli consisted of up to five regions depending on the length of the sentence. Based on the position of the error, the critical region appeared:

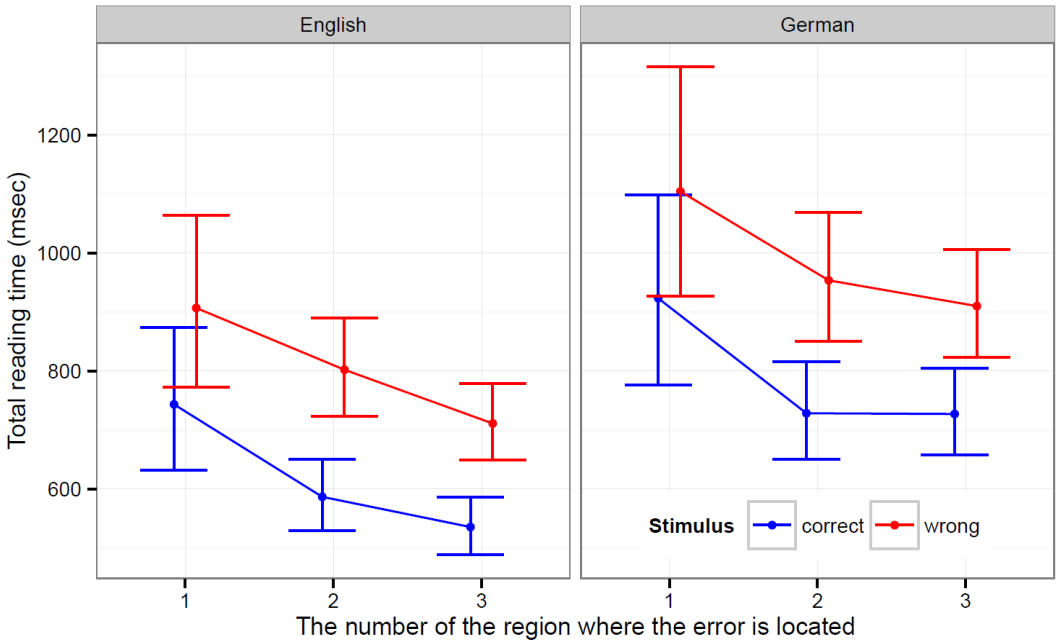
(1) at the beginning of the sentence (Region 1, as in Examples 5.13 – the data included 21 item, 594 observations),
(2) shortly after the beginning of the sentence (Region 2 – 82 items, 3828 observations), or
(3) in the middle of the sentence (Region 3 – 180 items, 8999 observations).

By analyzing the reading times for these three regions, it can be determined whether or not the reaction to errors depends on their position in the sentence. Particularly interesting is Region 1 because, at the beginning of the sentence, one would expect no clear distinction between correct and wrong stimuli. In this analysis, the dependent variable is not the regression path duration, as there is no place to regress from Region 1. In Figures 5.12 *a* and *b* I contrast the first pass reading time and the total reading time because the 'disambiguating item', the mismatching part, if not present in the same region, appears later in the sentence. If the error is not explicitly reacted to in the first pass and is not directly fixated, the increase in the total reading time will indicate that regressions were made back to the region containing the error, and the deviation from the norm was in fact noticed.

a.



b.



Figures 5.12. *The effect of the position in the sentence in two time measures*

Region 1 truly stands out in both Figures 5.12 *a* and *b* – it involves the longest gaze durations. The fact that the first fixation on a line is longer than other fixations is a general tendency (Heller 1982, Rayner 1977), and as the present analysis confirms, it works for both correct and wrong stimuli and for both subject groups. It also demonstrates that, generally, readers speed up towards the middle of the sentence. It is true that wrong stimuli are best detected already in the first pass when they are located in the middle of the sentence (in Region 3), but only by native speakers. In Regions 1 and 2, no clear distinction between targets and controls is made in the first pass. Non-native speakers poorly identify errors in the first pass altogether. Judging by Figure 5.12*a*, one

could have an impression that errors do not particularly create processing difficulty, only the position in the middle of the sentence is somewhat 'dangerous'. But the analysis of the total reading time in Figure 5.12*b* demonstrates that errors cause returns to the regions and clearly impose additional processing costs in ALL positions in the sentence in both native and non-native speakers. Thus, the idea that errors at the beginning of a sentence could slip away from readers' attention for technical reasons (the 'margins' for fixations in a line) has not been justified.

5.4.3 Sentence length

In the mixed sample, errors appeared in sentences of different length, and that can be used as an opportunity to test whether an error in a short sentence has a greater processing cost than that in a long sentence. A practical example could be if an error in the subject line of an e-mail stood out more than the same error in the body of the mail.

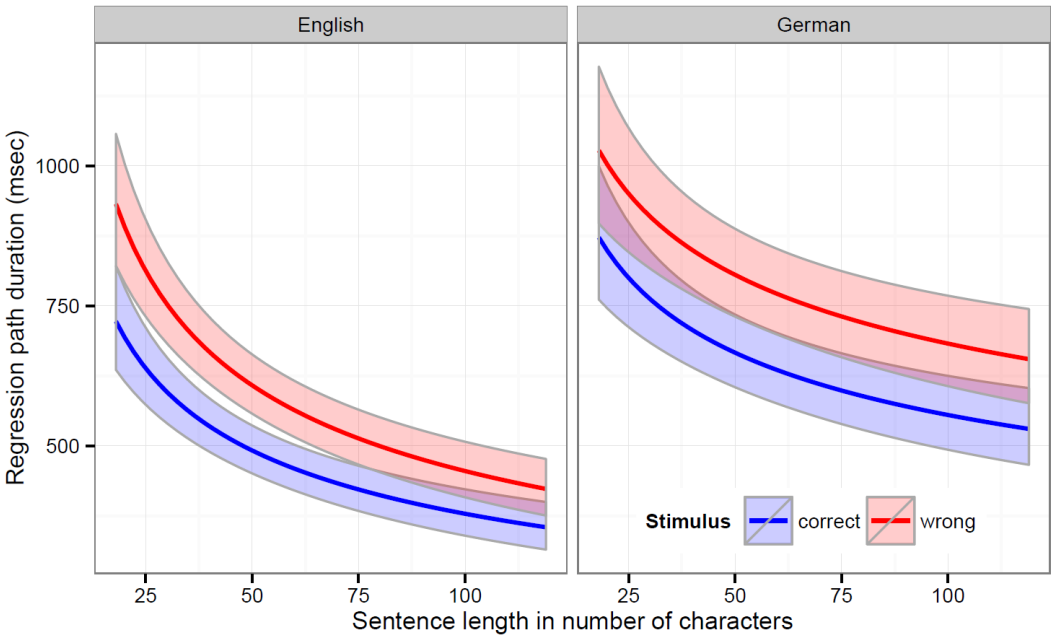


Figure 5.13. *The effect of sentence length on error processing cost*

The main effect of sentence length proved to be highly significant (log(sent_length), t = -7.349, sum contrasts). This means that the larger the sentence is, the faster is the processing of the critical region. The effect of error remains significant at every point of sentence length. It is clearly illustrated in Figure 5.13 that the additional processing cost caused by the error is almost the same for all sentences, which means that the overtime does not progressively increase or decrease with sentence length (log(sent_length):

stimulus, t = 0.384). But the number of characters in the stimulus significantly interacts with native language (t = -7.125). It indicates that native and non-native speakers differently react to sentence length. In Figure 5.13, native speakers of English started almost as high as German learners in short sentences but ended up much lower in long sentences. The steeper slope shows that native speakers speed up more when they read a longer sentence. This fact has no effect on error processing whatsoever.

Thus, the same word string with an error in the subject line of an e-mail will in fact consume more time than the same word string with an error in the body of the mail. But to be fair, a correct subject line will also be fixated longer than the same words embedded in a larger context, just because seeing a large text readers tend to move on faster. The difference is that, in case of the wrong item, processing of both will take a little longer due to the error.

5.4.4 Distance in number of words

Revisiting the Distance Hypothesis, I tested whether distance between two mismatching elements (this time not for verb-adverbial inconsistency but for other types of errors) influences the reading times at the critical region. To operationalize this variable, I counted the distance between the two words affected by the error (or to be exact, the spaces between them) with the idea that the first element 'sets the scene' for certain anticipations for grammar or lexical meaning, and it is after having processed the second element that the reader discovers the inconsistency. Here are some examples of how the category of "distance in number of words" was counted for our purposes.[19]

Sentence	Distance index
*She has very limited **knowledges** of German.*	0
*There were **less people** at that concert than last time.*	1
***They're** not **left** school yet.*	2
*When **ø** you went there**?***	3
*He was **ø** in a road accident **killed**, about two years ago.*	4
***When** he did that, I **would** be really mad!*	5
*My **watch** is broken. Can you fix **her** please?*	6

Table 5.6. *Calculation of distance*

[19] Here and below I mainly discuss grammatical errors; however, the analyses are based on the whole sample. For accuracy, I ran the same models on the subset tagged for the module "grammar" (about 10 000 observations), and there was no significant change in the result.

The operationalization of distance in number of words is similar to the 'error size' discussed in section 5.4.1, but only for small indices. In bigger distances, these two variables are very different. Consider the following example of a non-standard feature illustrating an extreme case of discrepancy.

(5.14) *As I **was** pushing to make the pile I almost **run** him over and I saw him, so we called the police.*

To repair the sentence, the reader has to change only one letter in the word *run* to turn it into the preterite form *ran* (the error size is "small"). And it has to be changed because eight words ahead of this verb there is a past tense marker that does not match the present tense (the standard meaning of *run*) in the main clause. According to the system introduced above, the distance index for this sentence would be "8", but as it was the only representative of such a large distance, it was removed from this particular analysis.

There are two ways to analyze the effect of distance: as a numeric variable or as a categorical variable. In the first case, the model returns a continuous line where the contribution of each distance index is smoothed down for the benefit of the whole regression. In case of a factor, the observations for every distance index are analyzed separately from each other. Such a method produces a more accurate account for each level (distance index), but the effect of the variable as a whole becomes blurred. For this reason, I will discuss both analyses below.

Table 5.7 lists the effects of the mixed model regarding distance between two mismatching elements as a numeric variable. As previously demonstrated in section 5.3.1, the reading times for errors evolve significantly over the course of the experiment, therefore, the effect of trial was statistically controlled in its interaction with the stimulus (correct or wrong).

	Estimate	Std. Error	t value
(Intercept)	6.365	0.244	26.115
distance.num	0.014	0.018	0.816
stimulus1	-0.123	0.01	-12.205
Native_Language1	-0.159	0.03	-5.323
rpd.minus1	0.001	0.001	-9.763
rpd.minus2	0.001	0.001	-2.707
reg.length	0.04	0.002	20.32
ia.index	-0.108	0.033	-3.268
Age	0.003	0.009	0.35
sex1	0.023	0.04	0.576
Language_major1	0.071	0.035	2.021
distance.num:stimulus1	0.014	0.004	3.888
distance.num:Native_Language1	0.008	0.004	2.214

stimulus1:Native_Language1	0.013	0.007	1.938
stimuluscorrect:trial	0.001	0.001	-5.13
stimuluswrong:trial	-0.001	0.001	-5.501
distance.num:stimulus1:Native_Language1	-0.013	0.004	-3.451

Table 5.7. *The effects of the numeric distance model (sum contrasts)*

What stands out in the model details is that distance between words as such does not have any effect on the processing times (t = 0.816), but its meaning becomes obvious only in interaction with other factors. The larger the distance is in case of an error or in case of L2, the greater is the processing cost (t = 3.888 and t = 2.214 correspondingly). Of particular interest is the significant three-way interaction of distance, presence of an error and native language (t = -3.451). It means that perception of errors in growing distance is different in native and non-native speakers.

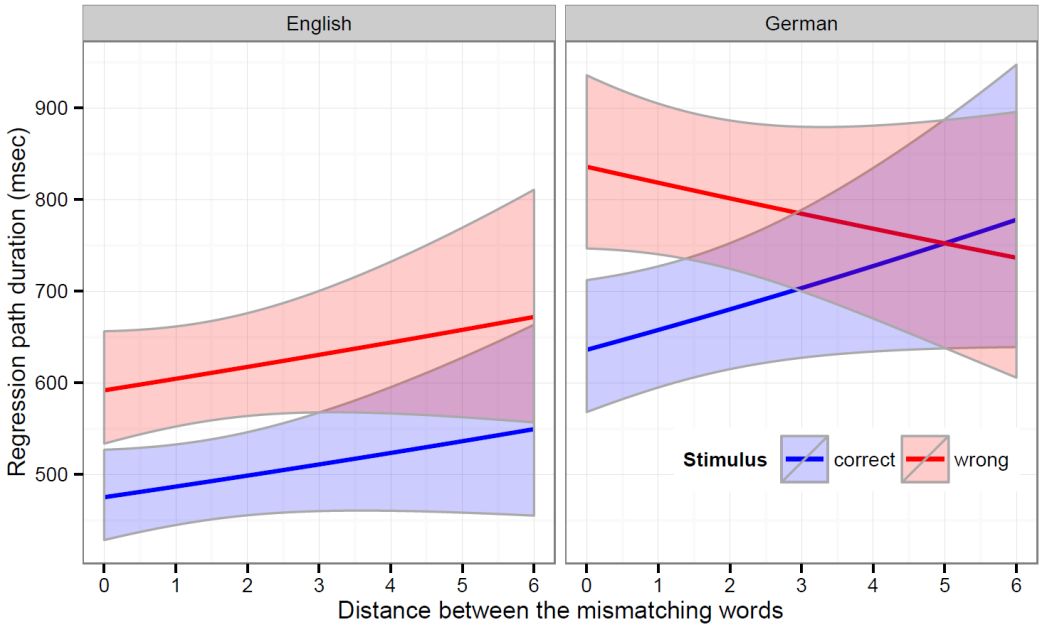


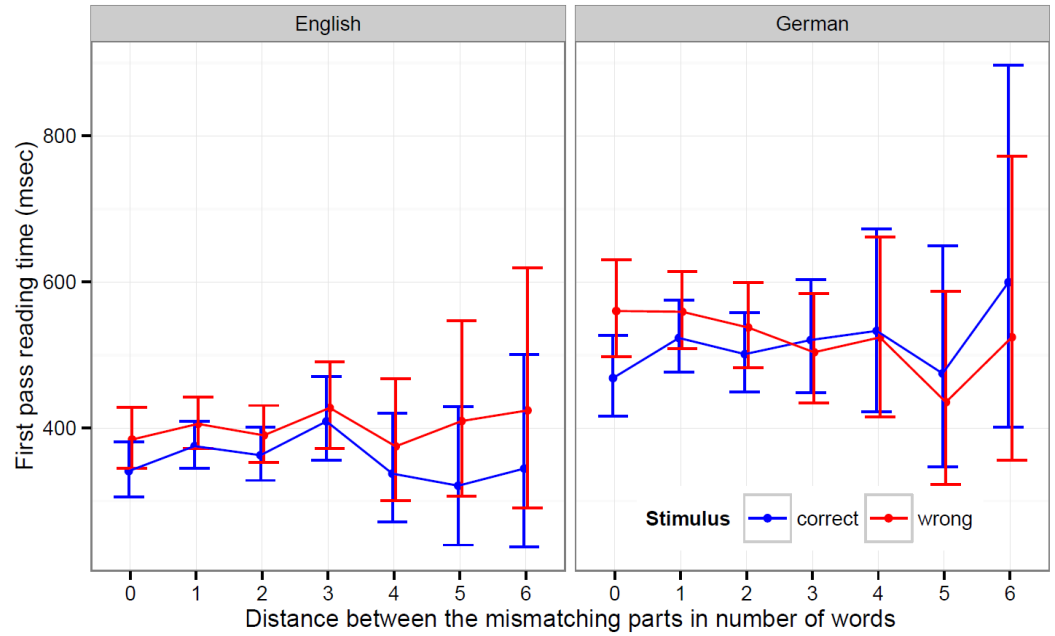
Figure 5.14. *The effect of distance as a numeric variable*

In correct sentences, L1 and L2 speakers react to growing distance in the same manner. The increasing times implicate that every new word separating interdependent items adds processing cost (also when the region length is statistically controlled), and non-native speakers continuously need more time than native speakers. However, when an error is present, the L1 and L2 patterns diverge. In native speakers, errors cause longer reading times at all distances; whereas non-native speakers need strikingly more time to process an error within the small distance (e.g. spelling, non-standard forms, agreement). As the distance between the mismatching elements grows, the error processing cost subsides. This result contradicts the findings in Chapter 4 based on mismatching adverbials (Figure 4.1). In section 4.4.1, it was shown

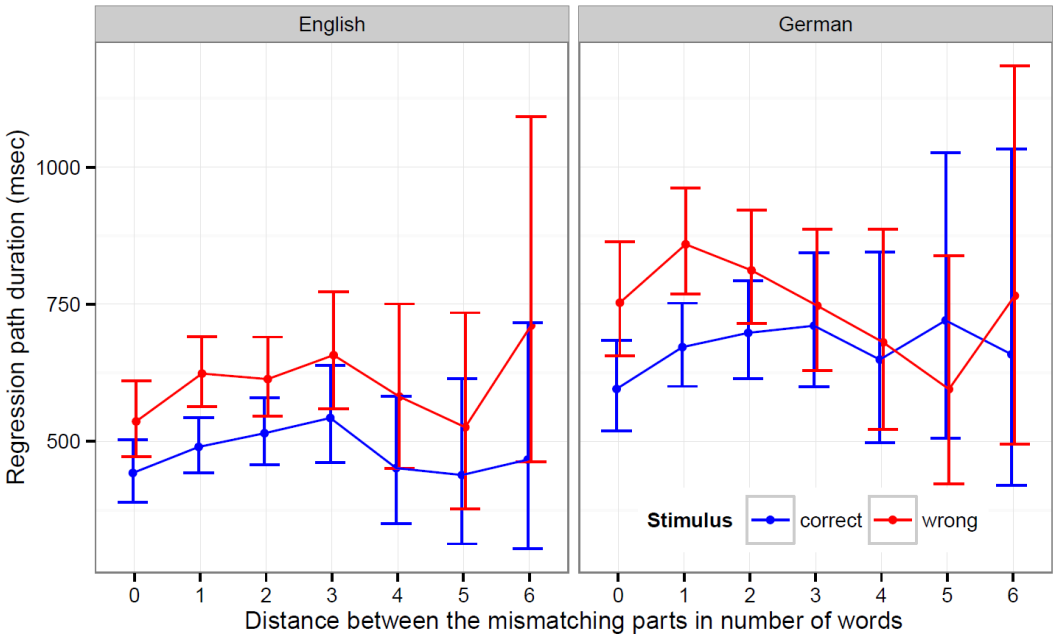
that participants read faster when the verb and the adverbial were separated by a longer object. Moreover, the ascending slopes in Figure 5.14 are not consistent with the observation that readers speed up if sentences get longer.

The tendencies described above follow from the analysis of distance as a numeric variable. If we look at the number of words between the affected parts as a factor, the picture is refined and offers a different interpretation. Figures 5.15 *a* and *b* present the reaction to errors in first pass reading time and regression path duration. By contrasting these time measures, it is possible to conclude that, for both subject groups, only errors within one word immediately add a significant processing cost (in Figure 5.15*a* the difference between "correct" and "wrong" is significant only for distance "0").

a.



b.



Figures 5.15. *The effect of distance as a factor*

The majority of errors within 2-5 words in native speakers and within 2-3 words in non-native speakers are not solved on the spot but require reanalysis, making the readers go back to the previous parts of the sentence (in Figure 5.15*b* the difference between "correct" and "wrong" is significant for distances from "0" to "4" in native language "English" and in distances from "0" to "2" in native language "German"). This result is similar to the effect of error size discussed in section 5.4.1.

Since errors at distance 0 behaved differently from the others in that they were recognized already in the first pass, it would be interesting to see whether they also cause regressions to the same extent as, e.g., distance indices 1-4 in native speakers. The thing is that the measure of regression path duration comprises the first pass reading time, and it is not clear whether the significant difference between the correct and wrong stimuli at distance 0 in Figure 5.15*b* is fully determined by the FPRT. To find out the scope of regressive paths, I extracted the FPRT from the RPD, which resulted in 3344 observations (25% of the data). Running the same model with the new time measure produced Figure 5.16.

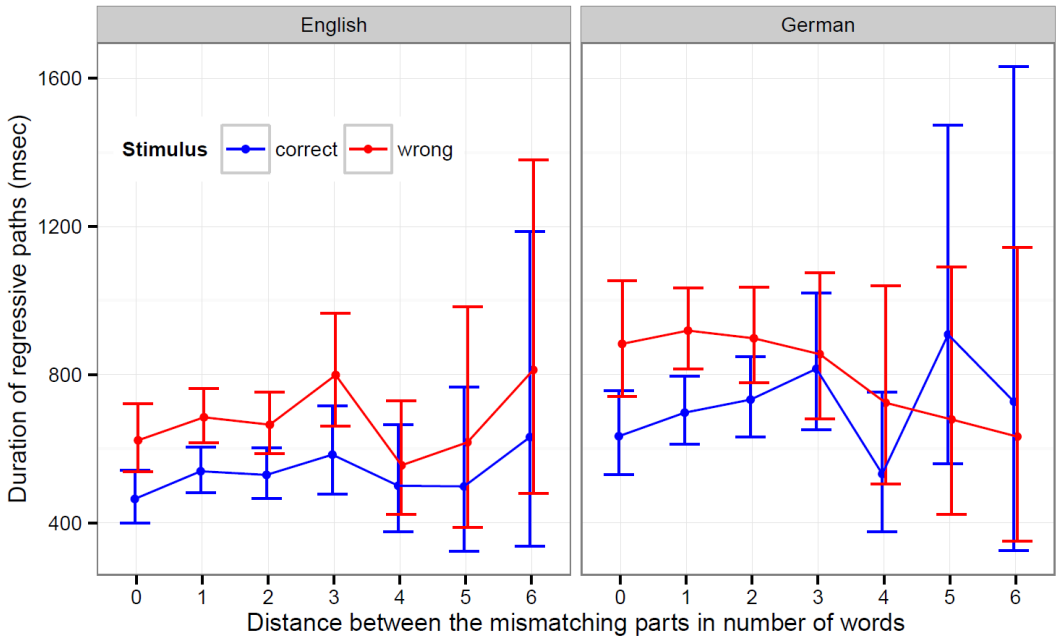


Figure 5.16. *The effect of distance on the regressions duration*

It is now clear that errors with distance 0 also cause regressions at the same extent as other distance indices with a significant result. Thus, errors within one word that not only complicate recognition but also require a second pass and errors within two adjacent words tend to be most costly. It is interesting to notice that in small distances, non-native speakers show almost the same pattern as native speakers.

It has already been mentioned that native speakers are sensitive to errors if the mismatching elements stand up to five words apart, whereas non-native speakers lose sensitivity after just three words. On the one hand, it may be due to working memory constraints. As described in Chapter 3, non-native speakers also employ other brain areas providing control functions, and while native speakers direct all their capacity to the meaning retrieval, learners also have to engage in a metalinguistic activity, gather information from multiple resources and for that reason may not be able to retain as many words in memory as native speakers can.

On the other hand, the distance of three or four words away is not "basic grammar" involving agreement of adjacent forms but rather such intricate grammatical congruence as sequence of tenses, aspect marking, etc. Such errors appear due to the use of other grammatical forms which also exist (e.g. competing forms) but are inappropriate in the particular context. Usually, they are difficult for learners to acquire, and such errors persist even in advanced students. Some examples of distance index 3 and 4, where the major differences between L1 and L2 speakers were discovered, are listed below.

(5.15) *a. *I **realized** that somebody **came** slowly up the stairs.*
b. I realized that somebody was slowly coming up the stairs.

*c. ***How** long are you **in** England?*
d. How long have you been in England?

*e. *I **didn't** know if she **is** at home.*
f. I didn't know if she was at home.

*g. *I really must stop **to** smoke, it's too **bad** for my lungs.*
h. I really must stop smoking, it's too bad for my lungs.

The sentences from the Distance sample testing the reaction to the verb-adverbial mismatch (in Chapter 4) were not used in the models reported above but were also tagged in the same manner for the distance between the tense/aspect marker and the final word of the temporal adverbial. It appears that only one item in the short condition had distance index 2:

(5.16) **If you **went** shopping **lately** you must have come back with an empty wallet.*

Ten items had distance index 3, and most items in the short condition were assigned "4" and "5". According to the findings based on the error-mix sample (Figure 5.15*b*), distance index 4 marks the decline of sensitivity for errors in native speakers; in non-native speakers, almost all sentences with adverbial mismatch passed beyond the border-line of two words away. In the long

condition, the distance spread was between "6" and "15". Here is an example of the latter:

(5.17) **Those who **didn't** visit the Sleeping Beauty Castle at Disneyland in California in the last couple of **years** are in for a big surprise.*

Extrapolating the results of the mixed sample to the sentences with adverbial mismatch discussed in Chapter 4, we notice that, in this error type, the conflicting parts generally stand further apart than in other errors with significant processing loads. However, native speakers had a better chance to detect a mismatching adverbial, and the fact that even they did not show a distinct reaction to it proves that tense and aspect marking is a grammatical category that is not clearly defined and where 'tastes' may differ. Moreover, this error type does not involve any processing difficulty of ambiguity because the cue for the lexical meaning of the temporal adverbial outweighs the grammatical constraints of the verb form.

By comparing Figures 5.14 and 5.15*b* (from the same time measure), it can be noticed that, when analyzed as a factor, the reaction to distance does not resemble straight lines but has rises and falls. The most striking inconsistency is in the reaction of German learners to errors with distance index 0. According to Figure 5.14, errors within one word have the greatest processing cost, whereas a more precise analysis in Figure 5.15*b* demonstrates that it is agreement failure between two adjacent words that is most costly. A reconciliation of these two analyses is offered in Figure 5.17. The polynomial model proves that the main (simple) effect of distance as a numeric variable is not linear, but quadratic (t = -2.006).

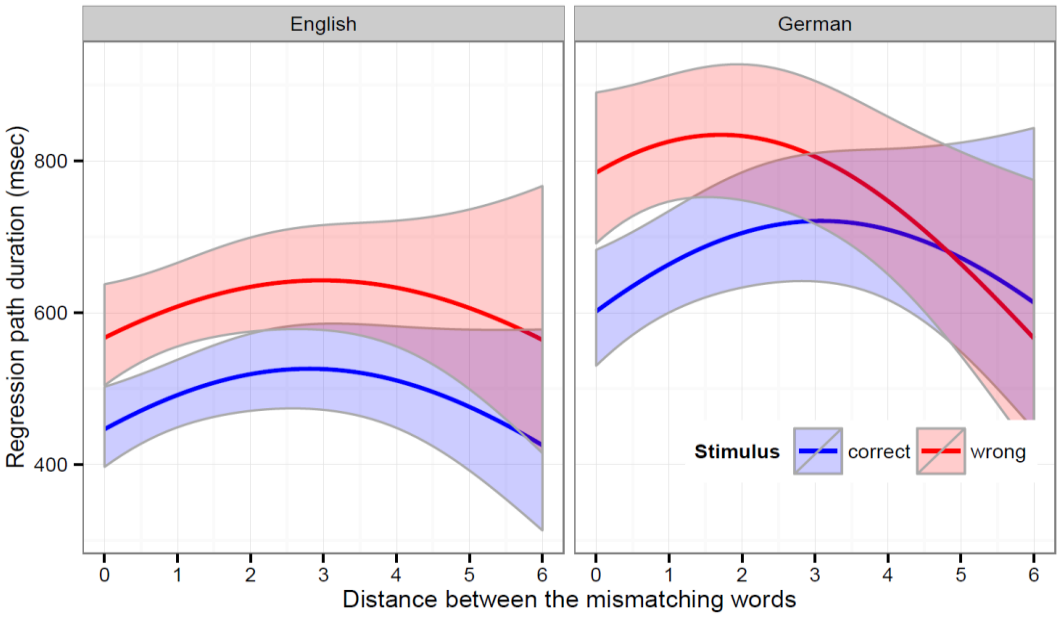


Figure 5.17. *The effect of distance as a numeric variable (polynomial)*

In correct sentences, the reading times increase up to the third word away and then descend in the same manner in L1 and L2 speakers, the latter being consistently about 200 milliseconds slower. When an error is present, both native and non-native speakers require additional times for processing the mismatching parts standing closer to each other. The difference between the subject groups in processing errors is that native speakers notice them and react with additional times at all distances, whereas non-native speakers make such a distinction only for errors detected within three words.

Although the main effect of distance has the form of a parabola, its interactions with other factors are significant only in the linear form.

	Estimate	Std. Error	t value
poly(distance.num, 2)1:stimulus1	1.952	0.542	3.599
poly(distance.num, 2)2:stimulus 1	-0.376	0.658	-0.572
poly(distance.num, 2)1:Native_Language1	1.279	0.543	2.356
poly(distance.num, 2)2:Native_Language1	0.416	0.651	0.639
stimulus1:Native_Language1	-0.006	0.004	-1.364
stimuluscorrect:trial	0.001	0.001	-5.183
stimuluswrong:trial	-0.001	0.001	-5.508
poly(distance.num, 2)1:stimulus1: Native_Language1	-1.924	0.548	-3.51
poly(distance.num, 2)2:stimulus1: Native_Language1	-0.713	0.659	-1.083

Table 5.8. *The effects of the distance model (polynomial)*

On the whole, there seems to be a relation of dependence between words, no matter if the parts are matching or not. Roughly speaking, the process of reading consists of word recognition followed by meaning integration - shifting forward as the eyes move on. The general effect of distance between two interdependent words may reflect the integration costs and working memory constraints: the processing times increase as the second, third and fourth words are added and held in memory. When the memory is 'full', the processing cannot be activated to higher extent, so the times gradually decline. It might be that the fourth word is the point when the global integration is completed after which readers do not recall the exact words of the utterance but only retain the meaning. Consequently, readers may lose sensitivity for grammatical inconsistency discovered beyond the fourth or fifth word. On the other hand, grammar as a system is built in such a way that the most strictly defined rules involve the arrangement of closely standing words, while the constraints for distant relations are loosened. It may be that the perception of errors proceeds correspondingly: we react more strongly to the violations of strictly defined rules within a small distance and are more open for variation at larger distances.

5.4.5 Acoustic similarity

Some errors are associated with incorrect spelling, but when the sentence is read out loud, the difference between the incorrect and correct forms is hardly audible. As known from psycholinguistic research (see, e.g. Leinenger 2014 for a review), in the process of visual word recognition, the signal also goes through the mental phonetic store, and the acoustic image of the word is also retrieved. In other words, when reading silently to ourselves we as if hear the sentence spoken.

In error judgment studies, the violation of rules allowed in the spoken German, e.g. leaving out case endings (*Das kann ich ohne dein Mann fertig machen [ohne deinen Mann]*), was rated least severe, compared to all other error types in an auditory experiment (Politzer 1979). And it was rated second least severe in a study with written materials (Delisle 1982), suggesting that "assimilation patterns from the spoken language are carried over to the written one" (op. cit.: 43). Wrong case endings were only more serious than the following spelling errors: *ahnlich [ähnlich], Büro Vorsteher [Bürovorsteher], Französischer Wein [französischer Wein]*.

Newman and Connolly (2004) report an event-related brain potential (ERP) study where they manipulated orthographic, phonological and semantic congruency of the final word in the sentence. One of the conditions was 'pseudohomophones' – items orthographically incongruent, but phonologically congruent to the anticipated word (e.g. *The ship disappeared into the thick phog* [fog]). As a result, an N400[20] was observed in response to violations of semantic expectations but was significantly reduced when phonological expectations were met regardless of the orthographic appropriateness. Besides that, the N270 appeared as a reaction to the violation of orthographic expectations, which was the case in the pseudohomophone condition. It indicates that participants registered the orthographic inappropriateness of the sentence completion before integrating the meaning, and the relative absence of the N400 in the pseudohomophones is evidence that some prior mechanism influenced the final semantic interpretation. These findings support "the existence of a phonologically mediated pathway that is responsible for semantic integration" (op. cit.: 102).

Another neurolinguistic study confirms phonological mediation based on fMRI data:

[20] The N400 effect is usually associated with access to the word meaning; N270 is an index to identification of conflicting information, and N430 reflects processing of complex conflicts following memory retrieval (Zhang et al. 2003).

> We propose that lexical decisions to pseudohomophones involve phonology-driven lexico-semantic activation of their basewords and that this is converging neuroimaging evidence for automatically activated phonological representations during silent reading in experienced readers (Braun et al. 2015).

The application of these findings to the present study suggests the following hypothesis: if phonological congruency makes access to the word meaning easier, then we can expect that errors which do not violate the acoustic image of the sentence will be processed faster than other errors which also strike with a different pronunciation. In this section, I will test whether acoustic similarity between the wrong and correct stimuli 'neutralizes' the effect of the error so that it does not add any processing load. Some examples of sentences coded for having the acoustic similarity ("yes" in Figure 5.18) are listed below.

(5.18) *a. Did you tell her where **your/you're** going tonight?*
*b. I am sure he **could of/could've** done it better.*
*c. They don't know that **tommorrow/tomorrow** may be a better day.*
*d. You know, he **useta/used to** get this bad press all the time, but he was a great guy.*
*e. Arizona State is **definately/definitely** playing good basketball, having won four games in a row.*
*f. You better not start all that **cryin/crying**, we can go back to the house right now.*
*g. The disaffected nobles **haff/have** little grasp of sophisticated conspiratorial politics.*
*h. I never **would of/would've** thought that he'd behave like that.*
*i. **Its/It's** going to be sunny tomorrow.*
*j. What's that? I can't remember **it's/its** name.*

Similar to my stimuli, 90% of the pseudohomophones in Newman and Connolly's (2004) study were not orthographic neighbors of the target words (same-length words different only in one letter). Their sentences included, e.g., *New York is a very big sitee* [city] and *The old milk tasted very sower* [sour]. Examples from language use that can be viewed as pseudohomophones include some common spelling errors (characteristic of both native speakers and learners) and the use of 'eye dialect' (more common in L1). The latter appeared as a deliberate misspelling in fiction to indicate that the speaker uses a non-standard or dialectal speech (Walpole 1974). However, multiple examples of 'writing as it sounds' are known from modern informal discourse, especially in computer-mediated communication. Vivian Cook gives a number of instances on his webpage (Cook 2015), e.g.:

(5.19) *a.* **'me' for 'my':** *Me name's Dave, me mum's at the top of the hill*
b. **'bin' for 'been':** *I've bin down Romany lane, where you bin?*
c. **'outta' for 'out of':** *get him outta here*
d. **'kinda' for 'kind of':** *you kinda lift your legs kinda up…*
e. **'sort've/sorta' for 'sort of':** *We had this sort've a gang*

f. **'n'' for 'ing':** *that's somethin' you don't often see*
g. **'woz' for 'was':** *It woz nuffink*
h. **'wot' for 'what', ''' for 'h'**: *Wot's 'appenin'?*

The products of grammaticalization, *gonna* and *wanna*, are also on that list.

Does this orthographic 'simplification' in fact make processing easier? In my sample, 30 items (1539 observations) with a deviant spelling met the criteria for acoustic similarity; the rest of the sentences (251 item) were categorized as having no acoustic similarity.

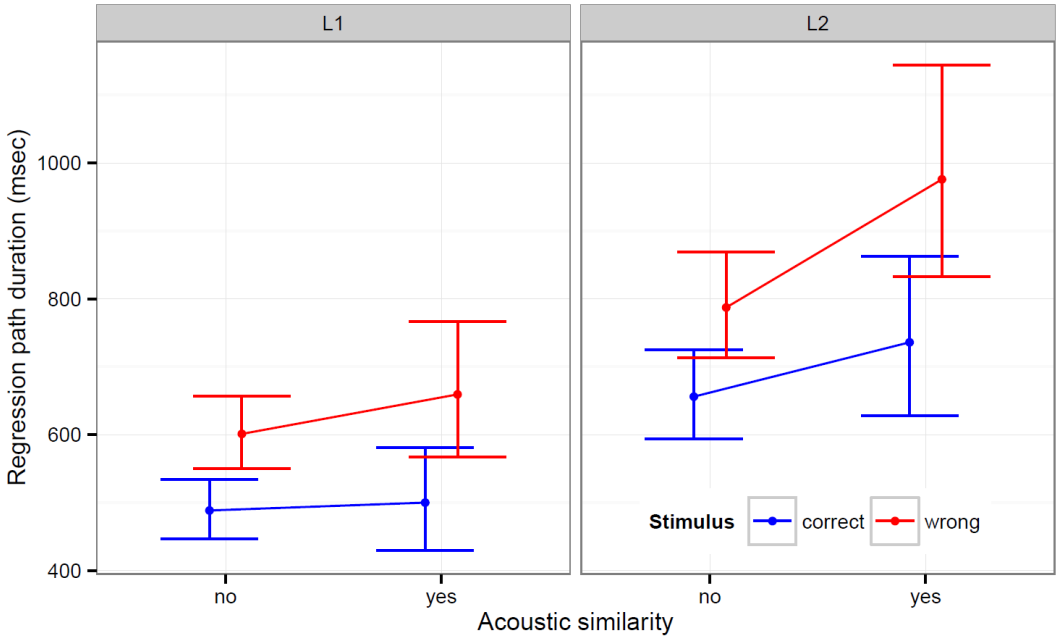


Figure 5.18. *The effect of acoustic similarity on error processing*

If the homophone effect of the acoustically similar errors made processing easier, the red lines in Figure 5.18 would have descended from "no" to "yes", making the distinction between the correct and wrong stimuli not significant. However, this is not the case for both language groups. The acoustically similar errors are not only significantly costlier than the correct stimuli (t = 3.126), but their processing is more time-consuming than that of other errors (that also sound differently).

In native speakers, the difference in the reading times for errors with and without the homophone effect is not significant, unlike in non-native speakers who seemed to be particularly attracted by the deviant orthography. In this respect, L2 learners are similar to children learning to read in their L1. In Blythe et al. (2015) children (aged 7-9), compared to L1 adults, showed lexical processing of pseudohomophones and pseudowords that was more disrupted by the presence of spelling errors, "suggesting a developmental change in the relative dependence upon phonological and orthographic

processing in lexical identification during silent sentence reading" (op. cit.). Learners' higher sensitivity to spelling can also be explained by the fact that it is an important part of foreign language training, or that learners are not familiar with instances of non-standard spelling in the 'eye dialect' and thus spend more time looking at it. In contrast, native speakers do not show particular interest in non-standard spelling, probably because they are more familiar with it.

	Estimate	Std. Error	t value
(Intercept)	6.475	0.243	26.689
homophone1	-0.056	0.034	-1.654
stimulus1	-0.119	0.01	-11.836
Language1	-0.168	0.03	-5.63
rpd.minus1	0.001	0.001	-9.624
rpd.minus2	0.001	0.001	-2.404
reg.length	0.046	0.002	21.544
ia.index	-0.04	0.034	-1.183
sent_length	-0.005	0.001	-5.523
age	0.004	0.009	0.387
sex1	0.023	0.04	0.579
Language_major1	0.069	0.035	1.987
homophone1:stimulus1	0.021	0.007	3.126
homophone1:Language1	0.027	0.007	3.976
stimulus1:Language1	-0.002	0.007	-0.363
stimuluscorrect:trial	0.001	0.001	-4.937
stimuluswrong:trial	-0.001	0.001	-5.339
homophone1:stimulus1:Language1	-0.004	0.007	-0.568

Table 5.9. *The effects of the acoustic similarity model*

Thus, my data do not support the hypothesis that readers, particularly in L1, are 'carried on' by the phonological congruency and are not disturbed by the error. The model shows for both language groups that even if the error 'sounds right', its processing will cost even more time than that of other errors. What could be the reason? Eye-tracking data only give us information about delayed reactions but no details on the quality of the underlying processes. For these purposes, ERP studies are more insightful. Newman and Connolly (2004) report that besides registering the orthographic inappropriateness (with the N270), there was another process involved in reading phonologically consistent stimuli.

> The large P300[21] elicited in the CN [congruent non-word = pseudohomophone] condition likely reflects the amount of attentional resources or the depth of processing required for

[21] The P300 indicates the processes of decision making, stimulus evaluation and categorization; the P600 is elicited by hearing or reading grammatical errors or other syntactic anomalies.

> evaluating nonwords and integrating them within the sentential context (op. cit.: 102).

It indicates that, during the same reading time, there are a few competing processes involved in the decoding of acoustically similar items with a deviant spelling, compared to the baseword. On the one hand, the phonological cue makes lexical access easier (absence of N400), which would allow to save the reading time. On the other hand, the orthographic anomaly not only requires time for its identification and processing (N270), which is also the case for many other errors, but the mismatch between the phonological consistency and the orthographic inappropriateness also engages additional resources to deal with such unusual words (P300). Obviously, the latter processes outweigh the quick access to meaning through matching phonology, and the processing of non-standard features imitating the spoken forms results in even greater cost.

5.4.6 Irregular graphics

There is a category of errors in which the wrong form graphically stands out because such a word does not exist in the (standard) language, by contrast with a misuse of an existing form. "In English orthography, *word* is easily defined as a unit of language that is written contiguously, with a space on each end" (Murphy 2011: 924). In my set of stimulus sentences, 37 items contained such a combination of letters written together that does not belong to the recognized words of the English language (e.g. would be underlined by spellcheckers). This category partly overlaps with the type of errors discussed in the previous section, therefore, the homophone effect is statistically controlled in the present model. The difference is that here I did not include examples where the acoustic similarity is transmitted with existing words, e.g. *where your going tonight, could of done, would of thought, its going to be sunny, remember it's name*, etc. But such instances of non-standard language as *useta* for *used to*, *kinda* for *kind of*, *haff* for *have*, etc. carry acoustic similarity and stand out as words not allowed in the formal language use, which implies that they are not likely to be presented to learners in the classroom. On the other hand, learners themselves can 'invent' forms in English by overgeneralizing the rules or transferring phenomena from their L1. Here are some examples from my sample of sentences with irregular graphics[22] ("yes" in Figure 5.20).

[22] The general term "graphics" is intentionally preferred because violations of orthography or spelling cover only a small part of the sample. For most non-standard features, the spelling is conventional (although for the purpose of our analysis marked as "wrong"), and for learner errors,

(5.20) *a. The house was full of **mouses/mice**, they were all over the place.*
*b. It's nice of all **youse/you** to have me here tonight.*
*c. He drives **badlier/worse** than his brother.*
*d. I've always fancied **meself/myself** married to a lordship and swanning up and down the stairs as lady.*
*e. The man I **met's** girlfriend [the girlfriend **of** the man I **met**] is a real beauty.*
*f. And it is good to be back with your friends and get **catched/caught** up with school again.*
*g. She has very limited **knowledges/knowledge** of German.*
*h. The river was a mile or so inland from the **fishervillage/fishing village** called Marietta.*

As previously mentioned in Chapter 3, the phonological mediation model was contrasted to the hypothesis of direct access to the word meaning from the orthography, especially in experienced readers and in high-frequency words. Both approaches are effectively combined in the dual-route model (Patterson et al. 1985, Colheart et al. 2001) arguing that both holistic word retrieval and letter-by-letter decoding may well operate in parallel, however the former route should be faster. "With visual access, phonology is *addressed*; with phonological access, phonology is *assembled*" (Carello and Lukatela 2011: 924). Let me once again show the hypothesized processing cost for different modes of word recognition and discuss it in relation to errors.

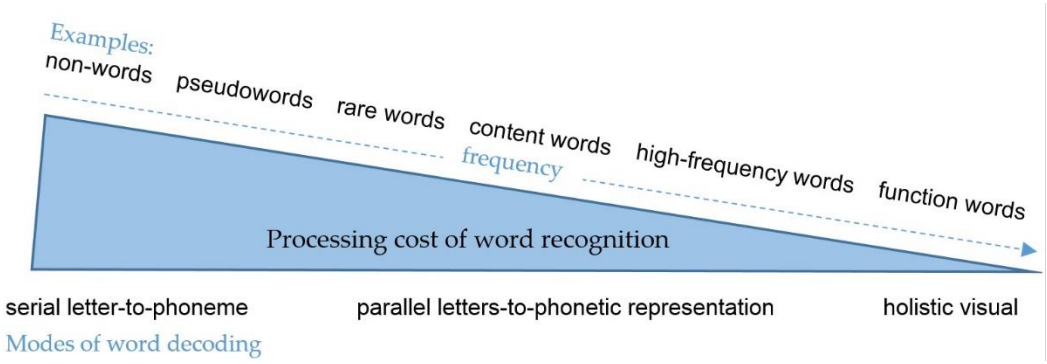


Figure 5.19. *The word recognition continuum*

Where do my stimuli belong in this continuum? Learner errors and non-standard features are different from pseudowords and non-words because their meaning is quite transparent. Typographical mistakes, deliberate misspellings and incorrect grammatical forms are contextually constrained, and in most cases it is clear how to repair the error. An exception might be unfamiliar words for the English learners, and such items would probably function as 'pseudowords'. I suggest that processing of real errors (with irregular graphics) is similar to recognition of compound words.

the problem is rather in the morphology than orthography, but, in the end, they all violate the graphic image of the words.

> Psycholinguistic and electrophysiological studies of lexical processing show convergent evidence for morpheme-based lexical access for morphologically complex words that involves early decomposition into their constituent morphemes followed by some combinatorial operation (Brooks and Cid de Garcia 2015).

Most studies on compounds naturally involve nouns (e.g. Rastle et al. 2004, Fiorentino et al. 2014), semantically transparent, e.g. *sailboat, roadside*, or opaque, e.g. *bootleg, butterfly*. I had only two items with compound nouns in my sample (*fishervillage* and *nurseryschoolteacher*). However, important is the principle of morpheme-based analysis during lexical access. In many items with an irregular word image, the violation is restricted to the grammatical morpheme (see Examples 5.20, also *I don know, there's a storm a-coming, she gots eyes like a cat*). Therefore, it can be expected that the processing cost for such kind of errors is going to be less than for pseudowords because, despite the orthographic violation, the meaning is easily retrieved from the base morpheme that remains intact. At the same time, the graphic deviation in the stimuli affects rather frequent content words, also large in size, so there is no chance for holistic visual access. Only in a few cases they could be considered orthographic neighbors and match the basewords in overall shape (e.g., *comitted, embarass, salat, Wendesday* – for the analysis of this category see section 5.5.1). For most other errors, the violation of the word image is sure to be detected during careful letter shape analysis. Thus, in the continuum of Figure 5.19, I would place the misspellings and the non-standard features around "rare words" and expect the corresponding mechanisms involved in their decoding.

My set of naturalistic sentences does not make it possible to check the error processing cost against the cost of meaningless pseudowords or non-words. Instead, it can be tested whether an irregular graphic image increases or decreases the reading times compared to other errors that only consist of existing words. If a word looks different from what readers are used to, it is reasonable to expect additional processing costs for the serial letter-to-phoneme decoding. On the other hand, although the correct orthography in errors involving existing words is processed faster than misspellings or wrong morphology, the grammatical or semantic inconsistency of the familiar words may be even costlier.

Figure 5.20 shows that errors with an irregular word image are significantly costlier than their correct controls (graphics : stimulus, $t = 2.559$), but they do not distinguish themselves from other errors involving existing words. For native speakers, unusually spelt errors do not make any difference from other errors at all, whereas non-native speakers process them slightly

longer, but the increase in the reading times is not as dramatic as in case of the acoustic similarity.

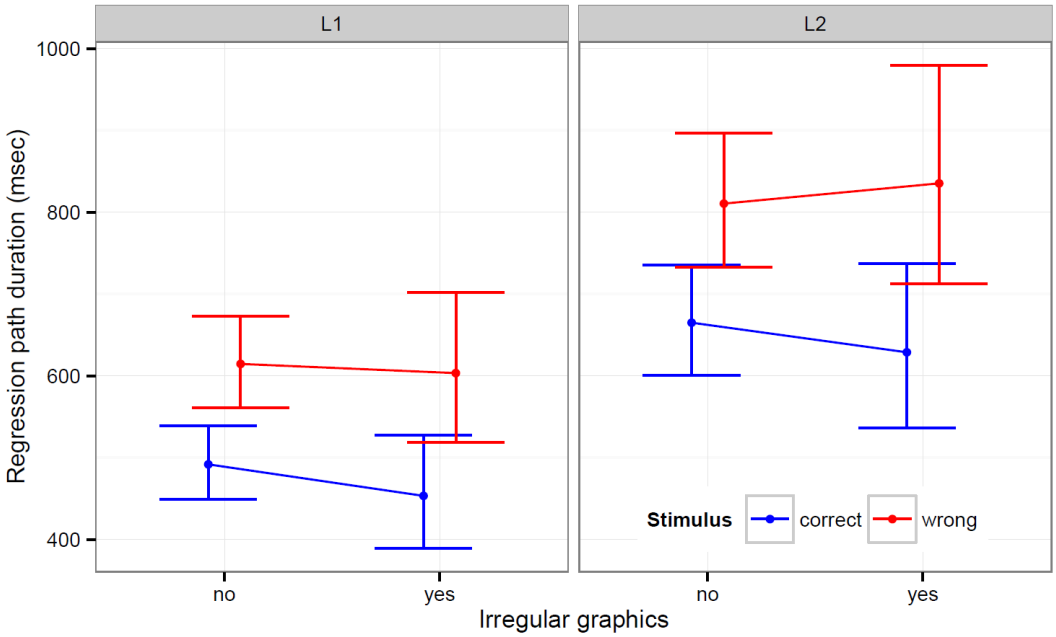


Figure 5.20. *The effect of an irregular word image on error processing*

	Estimate	Std. Error	t value
(Intercept)	6.473	0.243	26.605
graphics1	0.016	0.035	0.446
stimulus1	-0.124	0.01	-12.157
Language1	-0.154	0.03	-5.181
rpd.minus1	0.001	0.001	-9.648
rpd.minus2	0.001	0.001	-2.383
reg.length	0.045	0.002	21.087
ia.index	-0.042	0.034	-1.226
sent_length	-0.005	0.001	-5.246
age	0.003	0.009	0.377
sex1	0.024	0.04	0.607
Language_major1	0.07	0.035	2.02
graphics1:stimulus1	0.019	0.007	2.559
graphics1:Language1	0.009	0.006	1.494
stimulus1:Language1	-0.003	0.006	-0.539
stimuluscorrect:trial	0.001	0.001	-4.993
stimuluswrong:trial	0.001	0.001	-5.28
stimuluscorrect:homophone1	-0.05	0.039	-1.295
stimuluswrong:homophone1	-0.073	0.039	-1.871
graphics1:stimulus1:Language1	-0.003	0.006	-0.459

Table 5.10. *The effects of the irregular graphics model*

The power of the homophone effect proved to be almost significant for this model as well (the interaction of the homophone effect by wrong stimulus,

t = -1.871). Thus, the influence of the acoustic similarity between the error and the correct control is stronger than the impact of the irregular graphic image of the word.

5.4.7 Operation (deletion, substitution, insertion, reordering)

The assumption that silent reading involves somewhat interior articulation and 'hearing' of the text can explain native readers' reaction to certain errors and attributed gravity. For instance, Hultfors (1986) in his judgment study gives a possible explanation why an article error in the sentence *He is an useful member of the team* was perceived as more foreign and more difficult to understand than the wrong form of the article in *It took me a hour to get there:*

> It is probably easier to spot a letter which should not be there than it is to notice that a letter which one expects to find actually is not there (Hultfors 1986: 47).

In other words, for a reader it is easier to automatically insert a letter (sound) that is supposed to be there, than to stumble over a letter that should not be there. The extra *n* in the form of the article in *He is an useful member of the team* does not irritate the eyes; on the contrary, it looks like following the rule because the next word starts with a vowel. However, after decoding the sentence into a sound sequence, the reader immediately feels the distortion. Presumably, when reading *It took me a hour to get there,* most native speakers and advanced learners would automatically pronounce *an* and may not even notice that *n* was not there.

My hypothesis for an increasing processing cost is the following cline:

deletion -> substitution -> insertion,

based on the idea that it takes less effort to insert (sometimes automatically) what is missing (one edit) than to delete the wrong element and insert the correct one (two edits). In the process of coding the stimuli, I also added a fourth operation – reordering, which involves transposition of elements.

This classification is similar to Dulay, Burt and Krashen's (1982) "Surface Structure Taxonomy", or as James (1998) suggests renaming it, the "Target Modification Taxonomy". It is based on the 'behavioral' criterion of what the learner did wrong. Stemberger (1982) counted native speakers' slips and found that misselections (substitutions) are most frequent and additions (insertions) are least frequent. In an error gravity study by Garrett and Austin (1993), the participants attributed least importance to cases where apostrophes were

simply omitted and were most concerned with the insertions of apostrophes to mark the plural (e.g. *we sell car radio's*).

The *logical operation* variable consists of four levels:

1) deletion – 42 items, 1969 observations,

(5.21) *a. Do you have **brother** or sister?*
*b. He **say** he'll take care of everything. Just like always.*
*c. How **you** like the color of these walls?*

2) substitution – 180 items, 8750 observations,

(5.22) *a. Anybody knows that there'**s** all kinds of ways you get delayed as you get ready for flight.*
*b. Can I have a piece **from** that cake please?*
*c. He **said** me that you were ill.*

3) insertion – 36 items, 1696 observations,

(5.23) *a. Are you coming to the party? **The** Sarah is also going to be there.*
*b. He had to pay five hundred**s of** pounds.*
*c. Last month Osborne was elected **to the** chairman of the board of The Associated Press.*

4) reordering – 23 items, 1006 observations.

(5.24) *a. Actually, I **smoke not** and I never tried.*
*b. He was in a road accident **killed**, about two years ago.*
*c. I must go to Susie to **cut my hair**.*

Figure 5.21 shows the reading times for the four operations underlying the error. They refer to the erroneous sentence, i.e. what the producer of the error has done (inserted or deleted), and not what the reader has to do to repair the sentence.

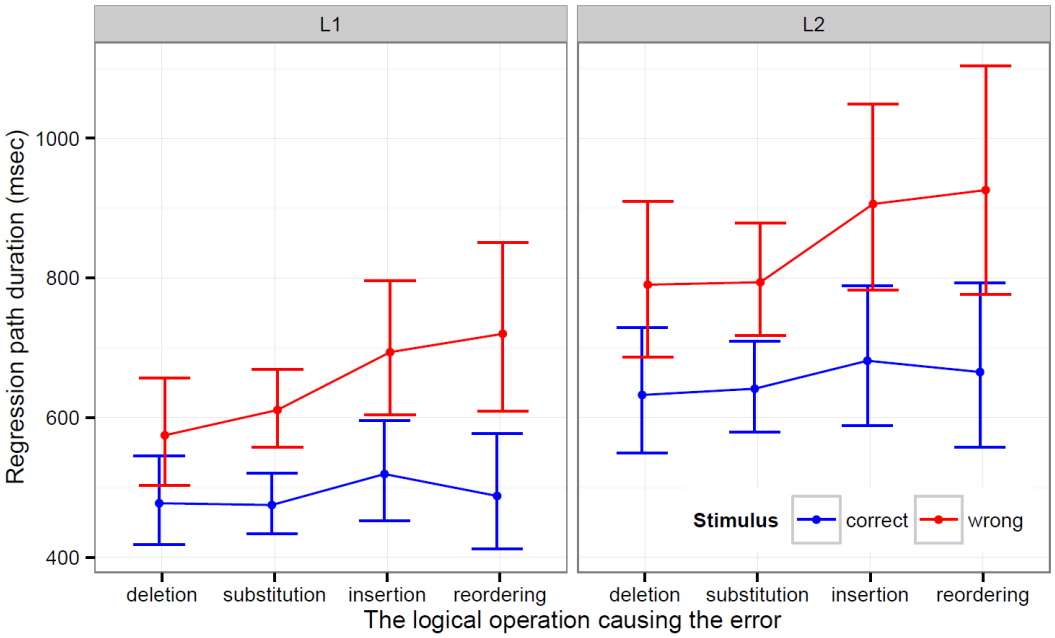


Figure 5.21. *Reading times depending on the logical operation*

The trajectories of the native and non-native speakers are very similar, which means that this cline might be a universal cognitive feature, independent of the language. As expected, deletions (also called omissions) are for native speakers the easiest error type to handle, followed by substitutions that requires somewhat more time, but the difference is not significant. In non-native speakers, the processing cost of deletions and substitutions is practically the same. By contrast, insertions and reordering are not only significantly costlier than their correct controls but also distinguish themselves from deletions and substitutions. The difference is significant in native speakers and close to being significant in non-native speakers.

Interestingly, repairing substitutions requires two edits (remove and add), while repairing insertions involves only one edit (remove) but is so much costlier. It is probably because insertions also break the flow in reading, these extra elements lead to false predictions, and the region has to be largely reanalyzed, similar to what happens in case of reordering. It is logical that reordering takes the longest time. It mostly refers to syntactic errors and involves moving whole constituents.

5.5 The effects of linguistic properties of errors

5.5.1 Typical error producer

As already mentioned in Chapter 1, some deviations from the norm can be attributed to native speakers (L1), others most likely come from foreign learners (EFL), some are characteristic of users of English as a second language (ESL), and particular errors can happen to anyone.

The Mouton World Atlas of Variation in English, WAVE (Kortmann and Lunkenheimer 2012), and its electronic version, eWAVE (Kortmann and Lunkenheimer 2013), give a comprehensive account of the spread of non-standard features in Englishes around the world. The presence/absence and pervasiveness of 235 features are reported for 30 mother-tongue varieties, 18 "indigenized L2 varieties" and 26 English-based Pidgins and Creoles. The results of the survey have demonstrated that geographic distance "turns out to be a weak predictor of overall morphosyntactic similarities between L1 varieties of English, accounting for no more than 5% of the linguistic variance" (Szmrecsanyi 2012: 841). The distribution of the features supports Trudgill's (2009, 2011) hypothesis that L1 varieties can be split into traditional (or low-contact) and high-contact varieties. The former (10 varieties) are characterized by complexification (more irregularity, less transparency) and such features as,

for example, agreement sensitive to subject type, *he/him* used for inanimate referents, or distinction between *gotten* and *got* for the dynamic vs. static meaning. By contrast, 20 high-contact varieties are characterized by simplification (less irregularity, more transparency) due to adult language acquisition, e.g. "null and deletion phenomena that do away with overt contrasts and markers (often inflectional) that are obligatory in Standard English" (Szmrecsanyi 2012: 831).

As for the L2 varieties, also known as "the Outer circle" or "New Englishes", they are spoken in Africa, Asia and other territories where English is not the native language of the majority of the population but plays an important role. "As a result of the extensive use of English for local purposes, these varieties come to develop phonological, lexical and grammatical characteristics that distinguish them from other varieties" (Lunkenheimer 2012: 843). Generally, L2 varieties use a smaller number of non-standard features than L1 varieties or pidgins and creoles. For the features investigated in WAVE, most L2 varieties cluster together in the NeighborNet diagram, "distinct from most varieties classified as L1, but united with most of them in the split that separates them from pidgins and creoles" (op. cit.: 857). On the one hand, second language varieties are acquired via the education system and are highly oriented towards Standard British English (therefore, less non-standard profile). "Deviations, particularly at the level of morphosyntax, tend to be perceived as 'errors' rather than features of a local variety, and official language policy often actively discourages speakers from using them" (op. cit.: 849). On the other hand, deviations from Standard English that L2 varieties exhibit resemble characteristics of learner language at the intermediate level and for that reason result in 'error types' different from mother-tongue features.

> Many of the structural differences between New Englishes and standard L1 varieties lie at the interface between lexis and grammar, in the area of "the co-occurrence potential of certain words with other words or specific structures" (cf. Schneider 2007: 86). For instance, many New Englishes differ from (standard) L1 varieties in their use of phrasal and prepositional verbs, e.g. creating new ones like *discuss about* or using StE phrasal or prepositional verbs without the particle or preposition, as in *pick someone* for 'pick someone up'… (Lunkenheimer 2012: 848).

In my sample, all sentences were coded for their possible producer, i.e. what kind of speaker (or, to be more accurate, writer) is likely to make such an error or use such a feature[23]:

[23] Cordial thanks to Kerstin Lunkenheimer for her help in coding the stimuli for this variable.

1) L1, ESL, EFL – errors and non-standard features characteristic of native speakers, learners of English as a second language and learners of English as a foreign language (29 items, 1419 observations),

(5.25) *a. Hopefully people **are liking**/like my voice and my music more than they are thinking about my age.*
*b. She **is gone**/has gone out and will come back in an hour.*
*c. I came **for**/ø to help you.*

2) L1, ESL – features used by speakers of traditional L1 varieties, high-contact L1 varieties and learners in an environment where English is the dominant language (69 items, 3475 observations),

(5.26) *a. As for Tant, he **weren't**/wasn't so scared as he pretended.*
*b. But Daddy goes to First Baptist most of the time, he **be**/comes over here for Sunday dinner.*
*c. But he **was after taking**/has taken four pitchers of draft with him, so I thought he'd not be needing more.*
*d. My daughter, **what**/who lives in London, is currently a student.*

3) ESL, EFL – features and errors characteristic of learner varieties, including indigenized L2 varieties (8 items, 350 observations),

(5.27) *a. I **no**/don't like the way people think that special education is an easy job.*
*b. Take them to **market**/the market, please.*
*c. They said they **will**/would do the work by 5 p.m.*

4) EFL – errors typically made by learners of English as a foreign language (168 items, 7806 observations),

(5.28) *a. **Became he**/Did he become a dentist?*
*b. Excuse me, **what is the clock, please**/what time is it?*
*c. He is a very **alone**/lonely man.*

5) and, additionally, spelling mistakes that can be attributed to all groups of speakers and learners (7 items, 371 observations).

(5.29) *a. Many men emphasized that their illegal acts were **comitted**/committed to allow them to engage in an honorable task.*
*b. His scenario is based on the pattern of fox remains found at sites **thoroughout**/throughout the island chain.*

Figure 5.22 shows the difference in the processing of the deviations from Standard written English produced by different categories of speakers. Essentially, the native speaking participants were delayed most at learner errors, whereas the non-native speakers spent more time processing L1, ESL features.

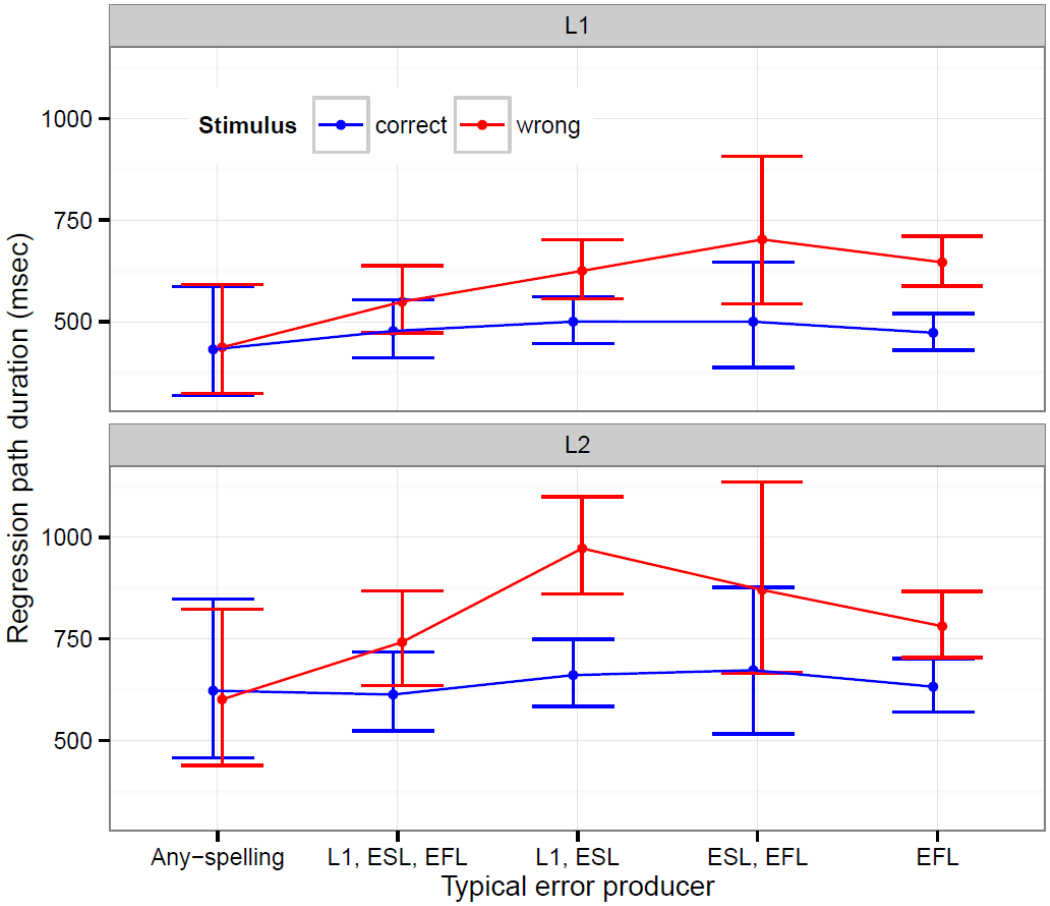


Figure 5.22. *RPD depending on the typical error producer*

The category that clearly stands apart in this analysis is spelling mistakes that can be produced by all kinds of English users. For this error type, the estimated means and the confidence intervals fully overlap each other in both native and non-native speakers demonstrating that small misspellings do not add any processing cost. This result could not be determined by such a small number of items (n = 7). With just one more item (n = 8), learner errors (ESL, EFL) are significantly more time-consuming than their corrected versions. Generally, all error types classified according to the typical producer, except for spelling mistakes, are significantly costlier than the correct stimuli, but to a different degree for the two language groups. Native speakers obviously spent more time on learner errors (ESL, EFL and EFL), whereas non-native speakers found non-standard features (L1, ESL) most challenging. This difference can be explained by the amount of experience each language group has with the corresponding error types. Native speakers are usually more familiar with typical vernacular and dialect features and, consequently, process them faster. Non-native speakers would have less contact with common deviations from Standard English, especially those who did not study a language as a major subject (38% of the German participants, and out of those who majored in language studies, only a part had expertise particularly in English linguistics).

While processing L1, ESL errors clearly differs in native and non-native readers, the estimated cost for ESL, EFL and EFL errors seems to be the same for both groups of participants. Also in this situation familiarity can be an important factor. Not only German readers are familiar with learner errors which they often explicitly discuss in the classroom, but also our L1 informants could have heard them many times. Most of the native speakers who participated in the eye-tracking experiment had lived in Germany for months or years, or traveled worldwide, and they learned German. To check whether transfer from German could have an effect, the wrong stimuli sharing features with the German language were singled out from their categories.

(5.30) EFL-German:

*a. "None of it was written down." "**Please?**" "There's no document that says it."*
*b. Actually, I **smoke not** and I never tried.*
*c. Can I **become** a glass of beer, please?*

ESL, EFL-German:

*d. I promise I **bring** it back tomorrow.*
*e. I knew she **is** in town.*

L1, ESL, EFL-German:

*f. How long **are you** in England?*
*g. I was **terrible** impressed with their new song.*

My expectation was that the errors with a German 'flavor' would be processed faster than the rest of the category.

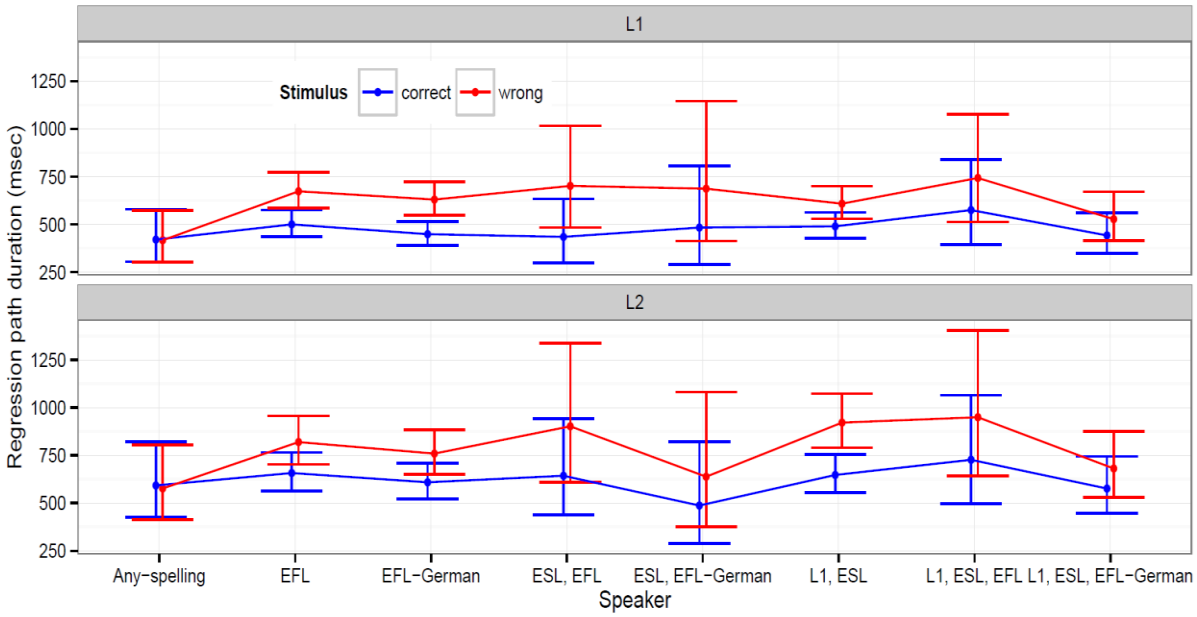


Figure 5.23. *Typical error producer with emphasis on German transfer*

The figure demonstrates no effect of transfer from German. Most notably, the largest categories – EFL and EFL-German – do not show any difference (where

shorter reading times for EFL-German were expected). In the diagram, the processing times are reduced for ESL, EFL-German, but with so few items the result is not significant. There is a similar situation with errors common to all categories of producer – L1, ESL, EFL(-German). On the one hand, the cost is smaller for the errors with transfer, but the difference between processing the correct and wrong stimuli is not significant, also not in the case without German transfer. Thus, I find no evidence that knowing a certain language facilitates the processing of errors due to negative transfer from this language. It is likely that the participants registered the presence of transfer and did not have difficulties with accessing the meaning but spent time evaluating the stimulus (similar to the homophone effect). It seems to be more important that an error goes against the structure of the target language, and understanding the background of the error does not make perception easier. At the same time, familiarity with particular surface structures (e.g. non-standard features) may expedite processing.

Another interesting question is how much the reaction to non-standard features differs across L1-varieties. I made a subset of the L1, ESL category and checked the reading times for the correct and 'wrong' stimuli in the native speaker groups, comparing them to the German learners.

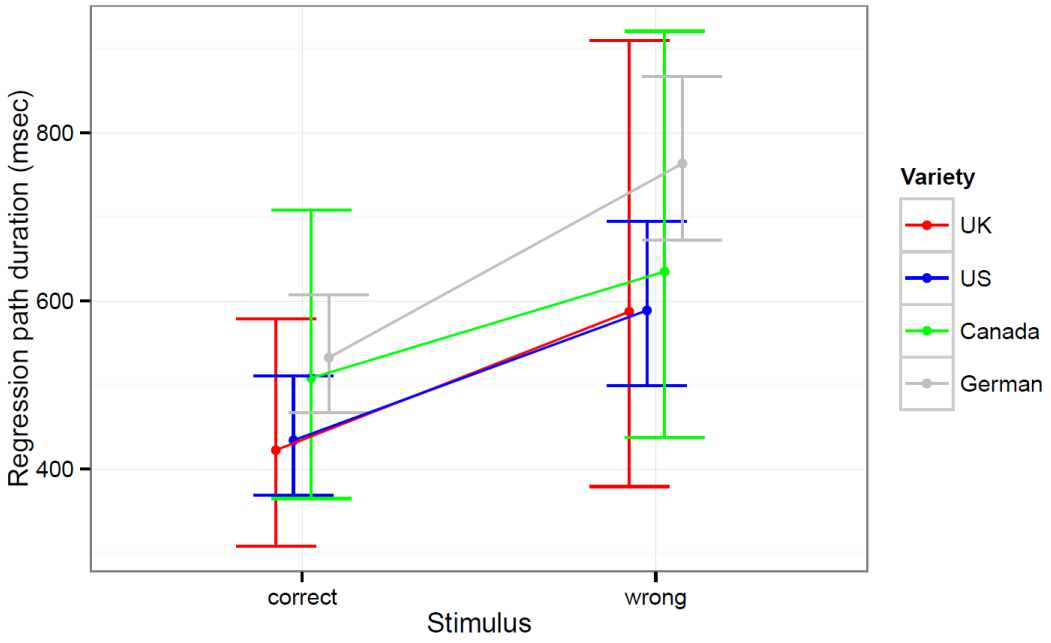


Figure 5.24. *Reaction to L1,ESL features in English L1-varieties (and German learners)*

In Figure 5.24, the estimated reading times for the UK and US speakers turned out to be remarkably close, the means being almost the same. However, the confidence intervals for the British participants are much larger than those for the Americans. In the latter group, the distinction between processing the norm and non-standard features is significant, which proves them to be more

sensitive readers. As for the UK participants, they seem to have a large divergence in the reading times within the group. As they showed themselves fastest readers in general, I suppose that some participants have developed good skills in suppressing the effort and, consequently, the processing costs for deviant sentences. Being either highly experienced readers or extremely motivated to be finished with the experiment in the shortest possible time, they might have adjusted themselves to dealing with any sorts of stimuli with equal pace.

5.5.2 Part of speech

The categorization of words into word classes goes back to Greek and Roman grammarians. The ten well-known parts of speech (noun, verb, adjective, adverb, pronoun, preposition, conjunction, numeral, article, interjection) encompass the similarities of behavior in terms of morphology, semantics and syntactic roles. In the structuralist approach, each word has a label for the class it belongs to, and the word can be used as a building block to form a phrase or a sentence it should fit in by definition. According to Construction Grammar, for example, it is the other way round: "the constructions are basic, and the parts of speech come into being as generalisations across different types of construction" (Hilpert 2014: 69). In the functionalist approach, to form a phrase or a sentence, one has to know the construction, and the part of speech information is akin to metalinguistic knowledge. Indeed, the grammatical categorization of words has been an essential part of linguistics, it prospered out to typology studies, and this is where it reached its limits. "Although most of the traditional word class distinctions can be made in most languages, the cross-linguistic applicability of these notions is often problematic" (Haspelmath 2001: 16538). It appeared that not all languages have the same categories. Besides that, the problem with the categorization is that some words can belong to two classes (e.g. the English *there* can be a pronoun or an adverb). Hence, it has been suggested to set pronouns and numerals apart from the other word classes or use intermediate categories, e.g. adverbial pronouns or adjectival numerals (cf. Haspelmath 2001).

While the traditional classification remains imperfect, the assignment of words to parts of speech proved to be practical for linguistic analyses. Dulay, Burt and Krashen (1982) proposed their linguistic category classification covering a few dimensions, one of which involves word classes. Each sentence in my sample was also coded for the part of speech of the word affected by the error. There was no particular hypothesis about the order of the grammatical categories in which the processing cost should increase. Therefore, the order of

the word classes in this analysis is arbitrary and formally relies on the order of the corresponding chapters in the grammar book by Quirk et al. (1985). Note that, compared to the tagging criterion in the next section (level of construction complexity), the approach taken here is in fact that of 'each word is a building block' which is 'damaged' or missing, and its relation with other words is not taken into account.

There appeared nine categories:

1) verb (83 items, 3981 observations),

(5.31) *a. Who **learnt you** Spanish?*
*b. What do you **watch? It looks** interesting!*
*c. So you be kind to him, **he be a** good man.*
*d. What **do you got** in Hawaii that you don't have here?*

2) auxiliary (38 items, 1861 observations),

(5.32) *a. When **started you** to play the piano?*
*b. **They're not left** school yet.*

3) noun (24 items, 1161 observations),

(5.33) *a. The house was full of **mouses**, they were all over the place.*
*b. Stores that sell junk food, soda and **bier** should offer healthy options or pay a junk-food tax.*

4) article (19 items, 862 observations),

(5.34) *a. You can't get there **without car**, there's no public transportation there.*
*b. **The most people would** agree with you.*

5) pronoun (39 items, 1811 observations),

(5.35) *a. This is the man **what** painted my house.*
*b. I haven't heard **something** from him for a long time.*

6) adjective (13 items, 681 observations),

(5.36) *a. This is the **goodest** cake I have ever tasted!*
*b. He is a very **alone** man.*

7) adverb (15 items, 768 observations),

(5.37) *a. You better come **quick**, I'm not sure how long it'll last.*
*b. There is one chair too **much**. Shall I put it away?*

8) preposition (27 items, 1238 observations),

(5.38) *a. You remind me **on** your father.*
*b. She called **during** you were out on your lunch break.*

9) and, additionally, phrase (23 items, 1058 observations) - errors involving larger units, such as lexical idioms and wrong pragmatic uses.

(5.39) *a. I tried a few times and it **brought nothing** in the end.*
*b. I must go to Susie to **cut my hair.***

*c. Excuse me, what **is the clock**, please?*

Figure 5.23 illustrates the reading times for the linguistic categories affected by the error as represented in the sample, contrasted for native and non-native speakers of English. The first striking conclusion is that, in native speakers, errors involving **all** parts of speech require significantly more processing time than the correct controls. The most costly categories are verbs, nouns, pronouns and prepositions. The easiest to process are errors involving articles.

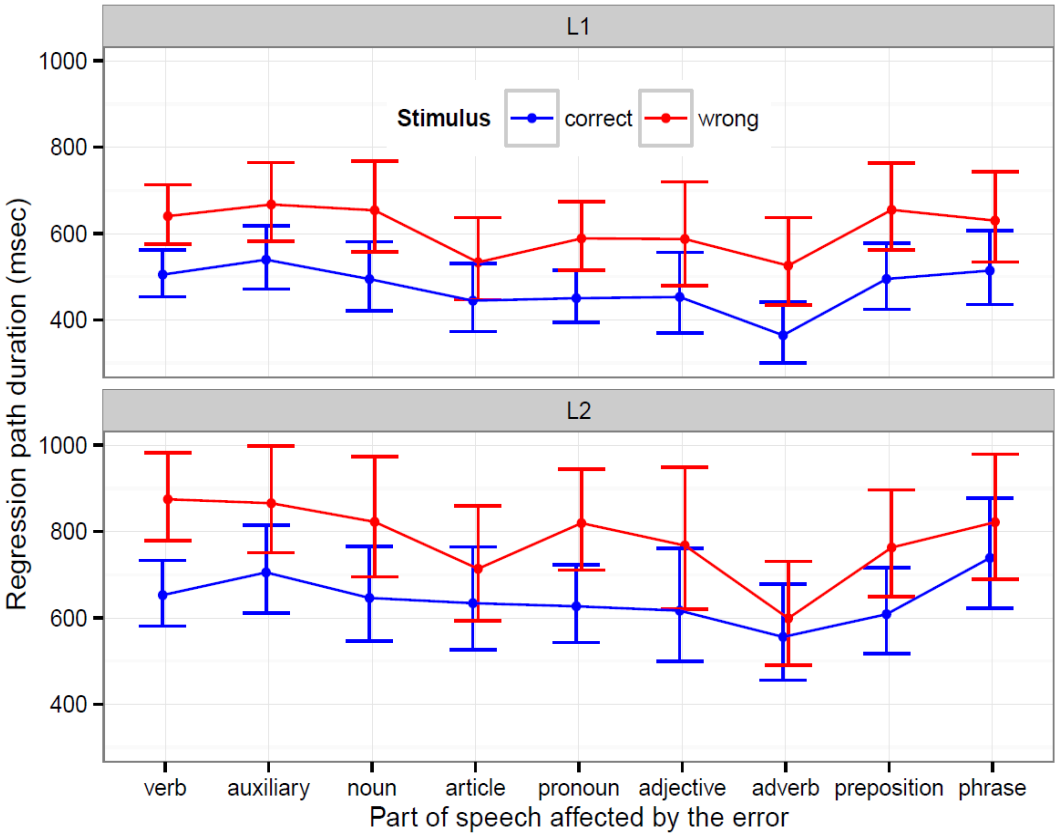


Figure 5.23. *RPD depending on the part of speech*

Non-native speakers showed no significant distinction between correct and wrong stimuli involving articles, adverbs and phrases, and close to that are also adjectives. The types of errors involving adjectives and adverbs included forms of comparison (*more easier*, *most pretty, badlier*) and using an adjective for an adverb (*doing good, come quick*). This type of deviation from Standard English occurs not only in learners but also in non-standard use of native speakers. In spite of it being so common, native speakers have drawn a clear line between grammatical and ungrammatical forms.

The category of pronoun contained a large number of non-standard forms as well (*yall, youse, meself, theirselves*, etc.) which should be more familiar to native speakers. However, it did not make processing easier and was particularly challenging for non-native speakers. The learners were most

puzzled by errors involving verbs. It was the largest category including not only wrong verbs as lexical errors but also non-standard tense and aspect marking and wrong verb forms.

As a result of this analysis, Russian learners may feel relieved that wrong articles are the easiest errors for processing (James (1998: 222) describes Soviet teachers of English who spoke "an article-less English"). Other than that, part of speech has little predictive power because it does not show a clear distinction between the error processing cost across the word classes and, furthermore, it lacks directionality (unlike the factor discussed in the following section).

5.5.3 Level of Construction Complexity

Construction Grammar as a theory was introduced in Chapter 1. At a closer look, what exactly is a construction? What kinds of constructions are there?

> Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency (Goldberg 2006: 5).

Hence, the main diagnostic features of constructions are non-compositionality (non-predictability) and conventionalization due to high frequency. The repository of constructions, the construct-i-con, is described as a highly structured, hierarchical network of interlinked constructions (cf. Hilpert 2014). They vary in size: they can be small (e.g. words) and large (e.g. word strings). At the same time, constructions differ in their level of abstractness from concrete to abstract. "For example, the concrete *dog* and the abstract N are both smaller constructions; likewise, the concrete imperative *Ask not what your country can do for you…* and the abstract passive Subj Aux VP (PP_{by}) are both larger constructions" (Boyd and Goldberg 2009: 418).

In the next analysis, I explore to what extent a violation of constructions at different levels complicates processing. The hypothesis is that the error processing cost (the reading time) grows proportionally to the size and complexity of the construction affected by the error. The factor is presented in the form of a cline.

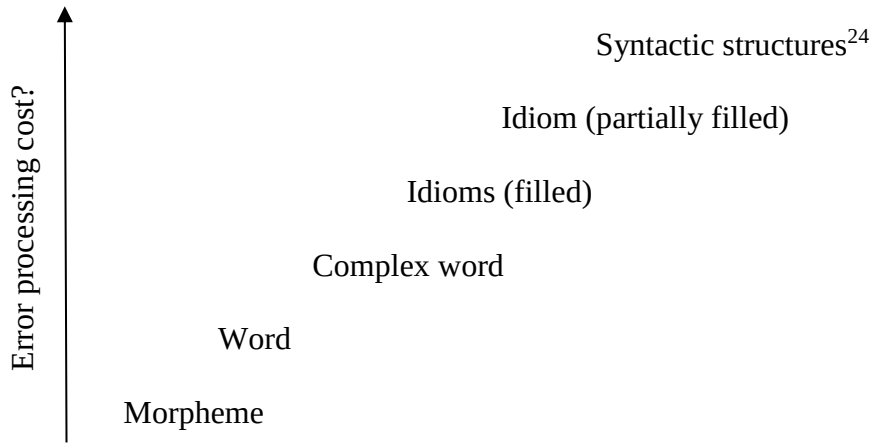


Figure 5.24. *Illustration of the hypothesis*

According to the hypothesis, spelling (or typographical mistakes) should be at the bottom of the processing difficulty scale as not involving any 'misconstruction' of the unit. The easiest construction violations should be observed at the level of morphemes, followed by words and idioms, and topped by the most complex syntactic structures. Driven by the stimuli used in the eye-tracking experiment, I had to 'customize' the cline of construction complexity by adding several other levels. For this factor, I made a selection of 164 items (7820 observations) and classified them into the following kinds of deviations.

- Level 1 – **spelling**. It includes deviations from the correct orthographic image of the word that could be attributed to a 'slip of a hand', a typographical mistake, or careless writing. It does not include examples of 'eye dialect' (*useta, nuthin, all that cryin, haff)* as intentional violations of orthography and does not include substitutions with homophonous existing words (*their* for *there're, your* for *you're, could of done* for *could've done*, or similarly *should of seen* and *must of been*).

(5.43)

1) *I think it's **wery** important when we talk about these issues to be specific…*
2) ***Its** going to be sunny tomorrow.*
3) *Stores that sell junk food, soda and **bier** should offer healthy options or pay a junk-food tax.*
4) *The newspaper is moving away from the news business and toward **backgrount** stories with specific profiles.*
5) *Allow the **salat** to marinate at room temperature for 10 to 15 minutes, tossing occasionally.*
6) *He is **an** useful member of the team.*
7) *It took me **a** hour to get there.*
8) *His scenario is based on the pattern of fox remains found at sites **thoroughout** the island chain.*
9) *They don't know that **tommorrow** may be a better day.*

[24] The hierarchy is based on Table 1 "Examples of constructions varying in size and complexity…" in Goldberg (2003: 220). What I summarized under "syntactic structures" was exemplified by the Covariational-Conditional construction, Ditransitive construction and the Passive.

10) *I had also seen **proffessional** and college-level performances by several musical ensembles and drama companies.*
11) *Many men emphasized that their illegal acts were **comitted** to allow them to engage in an honorable task.*
12) *Arizona State is **definately** playing good basketball, having won four games in a row.*
13) *It's just one of their many schemes to **embarass** me and to just drag me through mud.*
14) *The company will enjoy a wealth of free advertising from this **Wendesday's** press coverage.*

- Level 2 – **morpheme**. There are several inflectional morphological constructions in English that mark grammatical distinctions, and the most important of them are (cf. Hilpert 2014: 75): plural, present tense 3rd person singular, s-genitive, past tense as well as past participle and present participle marking, the comparative and superlative constructions for adjectives. From the derivational morphological constructions, my stimuli include only the adverbial suffix *–ly* (or its deletion, to be exact).

(5.44)

1) *That President has two **Secretary of States**.*
2) *Her **hairs** are very beautiful.*
3) *She has very limited **knowledges** of German.*
4) *The man **I met's** girlfriend is a real beauty.*
5) *This is my brother**'s**-in-law cap.*
6) *Some of the top models in the world aren't the **most pretty** girls, but they have the extra something.*
7) *He's making the battle a lot **more easier** for me, personally, because he's so strong.*
8) *He **say** he'll take care of everything. Just like always.*
9) *"Jack!" I **says** putting my elbows on that cushion. "How are you doing?"*
10) *I probably wouldn't let him go just because I **wants** him here.*
11) *He works very **hardly** and he likes his job.*
12) *You better come **quick**, I'm not sure how long it'll last.*
13) *My father said that was the only way you could get a **real** good look at the faces, and he was right.*
14) *I was **terrible** impressed with their new song.*

- Level 3 – **suppletion**, a level at the interface of morphology and autonomous words. Such functions as plural or tense marking can be fulfilled not only by inflectional morphological constructions but, alternatively, by means of different words, i.e. suppletive forms. One of the most common errors in children's L1 and learners' L2 is the regularization of the historically fixed forms. Besides those errors, my stimuli also include double marking and other violations.

(5.45)

1) *And it is good to be back with your friends and get **catched** up with school again.*
2) *They **fighted** bravely in the war.*
3) *I may have **gave** you a little too much for your first time.*
4) *So he's betrayed the program - the platform that he had **ran** against for the last 15 years.*

5) *I **seen** him a while ago, but I don't know where he got to now.*
6) *A lot of that was, you know, I **gone** on a mission trip with my church.*
7) *I didn't **saw** him today.*
8) *Things have **went** too far.*
9) *She wouldn't have **went** back in the house if she was hurt.*
10) *As I was pushing to make the pile I almost **run** him over and I saw him, so we called the police.*
11) *He drives **badlier** than his brother.*
12) *This is the **goodest** cake I have ever tasted!*
13) *The situation turned out to be **more bad** than I thought.*
14) *The house was full of **mouses**, they were all over the place.*
15) *As for Tant, he **weren't** so scared as he pretended.*
16) *Ron said there last night that you **was** his ex-wife and his friend only, that he had a girlfriend.*
17) *My brother **have** worked for this company for three years now.*

- Level 4 – **word**, a **lexical** error. In Construction Grammar, word is a multi-faceted construction with its specifications for phonology, morphology, semantics, syntax and pragmatics. Each of the 'facets' can be misused in the context. The following selection of stimuli instantiate the violation of the lexical meaning of the word.

(5.46)

1) *Yesterday's accident **depended on** the bad weather.*
2) *Who **learnt** you Spanish?*
3) *Can I **become** a glass of beer, please?*
4) *Please **bring** those books to the library for me.*
5) *They fix refrigerators, washing machines, TV sets and other types of **technique** every day.*
6) *The bus stop is right **before** our house.*
7) *She called **during** you were out on your lunch break.*
8) *She doesn't know what she **will** at all.*
9) ***When** he did that, I would be really mad!*
10) *These young leaders are decisive, **self-conscious**, quick-thinking, and excellent public speakers when addressing peers.*
11) *I always tried to **make** my best at school.*
12) *What shall we **make** today? Do you want to go somewhere?*
13) *You **do** good coffee!*
14) *The car was badly **injured** in the accident.*
15) *Borodin is one of the Russian **compositors** well-known around the world.*

- Level 5 – **word**, a **grammatical** error. It is difficult to disentangle semantics from grammar, and it is often that grammatical features follow from the meaning of the word (e.g. a mass noun is preceded by the zero article). At the same time, the examples below are qualitatively different from the purely lexical errors in Level 4. The following errors are caused by a violation of the grammatical specifications of the word, e.g. agreement in number (5.47: 1-3) or gender (5.47: 4-6), being countable (5.47: 7-10) or compatible with other constructions (5.47: 11-14).

(5.47)

1) *I've just heard **one** interesting **news** on TV.*
2) *Where **are** the **money** I gave you yesterday?*
3) *Here **are** the **money** I owe you.*
4) *My watch is broken. Can you fix **her** please?*
5) *The girl was lost, **it** didn't know what was going on.*
6) *I brought a new book yesterday and put **her** on the table.*
7) *It is **a** hard work to write a book.*
8) *What **a** dreadful weather!*
9) *There is one chair too **much**. Shall I put it away?*
10) *I don't know **much** people in this town.*
11) *He **said me** that you were ill.*
12) *He has **an own** company.*
13) *He is a very **alone** man.*
14) *Are you coming to the party? **The Sarah** is also going to be there.*

- Level 6 – **idiom** in its traditional form of a fixed string of words with non-compositional **lexical** meaning (in this case, a violation of it).

(5.48)

1) *I am convinced that he will **hold his word**, but he faces problems…*
2) *I tried a few times and it **brought nothing** in the end.*
3) *Statistics sometimes has to compare apples **with pears** to make the math work.*
4) *It sounds plausible, but I wouldn't **lay my hand in the fire** for it.*
5) *The global missile is a propaganda weapon. Let the Americans **break their head** over what I said.*
6) *My **eyes are running in different directions**, there's so much of everything.*
7) *It's **written with a pitchfork on the water**, so don't rely much on it.*
8) *Oh, stop **hanging noodles on my ears**, I wasn't born yesterday.*
9) *Excuse me, **what is the clock**, please?*

- Level 7 – **idiom**, a fixed string of words with a particular **grammatical** behavior. Such word combinations can be assigned to idioms because the use of their parts, or in this case mostly one part, is not transparent, i.e. cannot be predicted, unless one knows the construction. Several examples below (5.49: 1-10) include the wrong use of a preposition following a verb (phrasal verbs), an adjective or a participle. Among other sentences, there is an idiomatic expression with the zero article (5.49: 11), uses of the modal verb and causative verbs with a bare infinitive (5.49: 12-14), and other individual constructions.

(5.49)

1) *I **listen** different kinds of music.*
2) *It's **typical for** him to leave without saying a word.*
3) *Ladies and gentlemen, we are shortly **arriving** the Freiburg Train Station.*
4) *You **remind** me **on** your father.*
5) *I was **operated** last week.*
6) *She is **married with** a German.*
7) *He was **dressed with** a dark suit.*
8) *This room **smells** food, doesn't it?*
9) ***Smell on** these flowers!*

10) *I shouted to him to **look up** for the car.*
11) *Our son is 17 and he still **goes to the school**.*
12) *If you **let** them **to** do whatever they want, then I guess they are your friends.*
13) *You **must to** work hard, if you want better things.*
14) *They **made** me **to** do it.*
15) ***How do you call** one thousand lawyers at the bottom of the ocean? A good start.*
16) ***Can you** French?*
17) ***It** was little **else** to do.*
18) *My brother **and me** were late for school.*
19) *I'm **doing good** in school, I'm in the highest reading group.*
20) *Hey mom, **us kids** want to help with Christmas too.*

- Level 8 – **idioms, partially filled**. "…There is evidence to suggest that many idioms cannot be stored as fixed strings, which makes it necessary to think of idiomatic expressions as schemas with slots that can be filled with certain elements but not others" (Hilpert 2014: 6). This level is one step more schematic than the previous kind of idioms. It is reflected in the fact that the words in the sentences below can be replaced with other words that fit in the same construction, but the error affects the essential (irreplaceable) part of the construction. For example, in the quantity expression with a numeral (5.50: 1) which should be corrected as *five hundred pounds*, the word *hundred* can be replaced with *thousand* or *million,* and instead of *pounds* one can use *dollars* or *euros*, but the first word should not be marked plural, and the preposition *of* should be deleted. In Examples 5.50: 2-5, the indefinite article is missing in constructions HAVE + *a* + NP or BE + *a* + NP which are generalized to Verb + *a* + NP with *have* and *be* as highly frequent verbs. Similarly, *without car* (5.50: 6) and *to market* (5.50: 7) are missing an indefinite and a definite article correspondingly. Sentences 5.50: 8-13 represent the constructions requiring the *-ing* form of the verb (instead of the infinitive). The other examples also include ill-formed constructions where one element can be replaced.

(5.50)

1) *He had to pay five **hundreds of pounds**.*
2) *I am in the mortal pain, I **have the toothache**, you know.*
3) *We've **had nice conversation** with people from all over the world.*
4) *My mother **is doctor**, a neurologist, to be exact.*
5) *My sister **is nurse** and she works at a hospital.*
6) *You can't get there **without car**, there's no public transportation there.*
7) *Take them **to market**, please.*
8) *Many cities have **stopped to expand**.*
9) *I really must **stop to smoke**, it's too bad for my lungs.*
10) *I **couldn't help but laugh** when I saw this.*
11) *I am **used to get up** early in the morning.*
12) *I **look forward to hear** from you.*
13) ***Instead of to fight**, they started to laugh.*
14) *This cake **smells well**! Did you make it?*

15) *Can I have a **piece from** that cake please?*
16) *This happened yesterday **on the party**.*
17) *I spent the evening **by John and Alice**.*
18) *I want to **explain you this** in more detail.*
19) *Last month Osborne was **elected to the chairman** of the board of The Associated Press.*

- Level 9 – **syntactic structures**. In our case, these sentences represent such abstract and schematic constructions where all content words affected by the error can be replaced. Basically, it boils down to syntactic slots that are incorrectly ordered or erroneously filled. For example, the unsuccessful attempts to form a question (5.51: 1-4) are missing the auxiliary verb *did* and the corresponding word order. Sentences 5.51: 5-11 violate the norms of expressing negation (by a mismatching pronoun or a missing auxiliary) employing different kinds of syntactic structures. Several sentences (5.12-18) illustrate the incorrect order of constituents in an affirmative clause from switching around two words to more complex rearrangements. The list is closed by deletions and an insertion of syntactic elements.

(5.51)

1) ***Came he to** the party yesterday?*
2) ***Became he** a dentist?*
3) *When **started you** to play the piano?*
4) ***When you went** there?*
5) *I **don't have no** reason to worry about it.*
6) *I **no like** the way people think that special education is an easy job.*
7) *I **didn't buy nothing** at the supermarket.*
8) *I **haven't heard something** from him for a long time.*
9) *Actually, I **smoke not** and I never tried.*
10) *He **thinks not** that they know what to do.*
11) ***No people are here**.*
12) *But it'll be **soon dark** and I won't be able to dig anything if I can't see.*
13) *On Tuesday **have we** a holiday and we'll make a small trip.*
14) *Did I tell you that my mother **English speaks**?*
15) *He **became finally** President on third attempt.*
16) *He **was in a road accident killed**, about two years ago.*
17) ***The in Britain with excitement awaited festival** will take place next week.*
18) *When I **five years ago visited London**, I could hardly speak English.*
19) ***He very likes** Chicago.*
20) *I **can Spanish**, because I learned it at school.*
21) ***He a good boy**, and his eagerness makes him seem younger than his 21 years.*
22) *I don't know what **improvements that are** being planned.*

- Level 10 – **tense and aspect**. This category goes beyond the conventional understanding of constructions, such as argument structure constructions (e.g. ditransitive or caused motion). However, they were selected here to illustrate the effect of errors similar to the mismatching adverbials discussed in Chapter 4 and to compare them to the

'misconstructions' at the other levels. Tense and aspect errors reflect a conceptual incongruity between the verb form and the meaning of this verb (5.52: 1-3), the situation backgrounded by another verb (5.52: 4-11) or an adverbial (5.52: 12-20) or other lexical means (5.52: 21-22). Figuratively speaking, the previously discussed errors affect the 'hardware' of the language, whereas the deviations in tense and aspect affect the 'software' (the verb form is correct on the surface, but does not fit the context)[25].

(5.52)

1) *He **is meaning** this book, not that one.*
2) *Hopefully people **are liking** my voice and my music more than they are thinking about my age.*
3) *Patients **are wanting** a lot of personal attention.*
4) *I realized that somebody **came** slowly up the stairs.*
5) *When he came I **slept** on the couch in the living room.*
6) *He said he already **finished** work.*
7) *They said they **will** do the work by 5 p.m.*
8) *He said he **lives** here for a long time.*
9) *I didn't know if she **is** at home.*
10) *What do you **watch**? It looks interesting!*
11) *I promise I **bring** it back tomorrow.*
12) *"**Were you** ever a smoker?" "No, I never smoked, except one misfortune…"*
13) *Let me just ask you, **Were you** ever in a car with a pair of binoculars near their house?*
14) *Recently she **became** a vegetarian.*
15) *This time tomorrow I will **lie** on the beach.*
16) *I **know** her since we were children.*
17) *Maybe they **didn't find** it yet, I'll go, and you wait here.*
18) *I **have seen** Mary yesterday and she looks good.*
19) *The German prison system **improved** a lot in recent years.*
20) *Where **do** you **go** now?*
21) *I'm sorry I'm late. Have you **waited** long?*
22) *How long **are** you in England?*

Judging by several sentences in 5.52, this error type is also found in L1. Although *are liking* and *are wanting* may be judged as somewhat strange or creative, these particular examples (5.52: 2 and 3) were taken from the corpus (COCA). Generally, the extension of progressive to stative verbs is attested in all continents (rate 63%) with the pervasiveness of 65% (eWave). The violation of the sequence of tenses[26] is exemplified in Sentences 5.52: 7-9. The latter is

[25] Here I did not include non-standard tense and aspect markers that affect the surface structure (*he be a good man, he be over here for dinner every Sunday, you done got your wish*, etc.)

[26] Sequence of tenses is agreement in tense between the verb phrase in a subordinate clause and the verb phrase in the main clause that accompanies it. "Most commonly it is a case of a past tense in a main clause being followed by a past tense in a subordinate clause. Compare:
(a) *I **assume** [you **are** going to be late]*. (present followed by present)
(b) *I **assumed** [you **were** going to be late]*. (past followed by past)
The interesting thing is that the past tense of the subordinate clause can easily refer to the present time, as in *Hello! I didn't know you **were** here*. In such cases, **sequence of tenses** overrules the normal meanings of past and present tenses" (Leech 2006: 106).

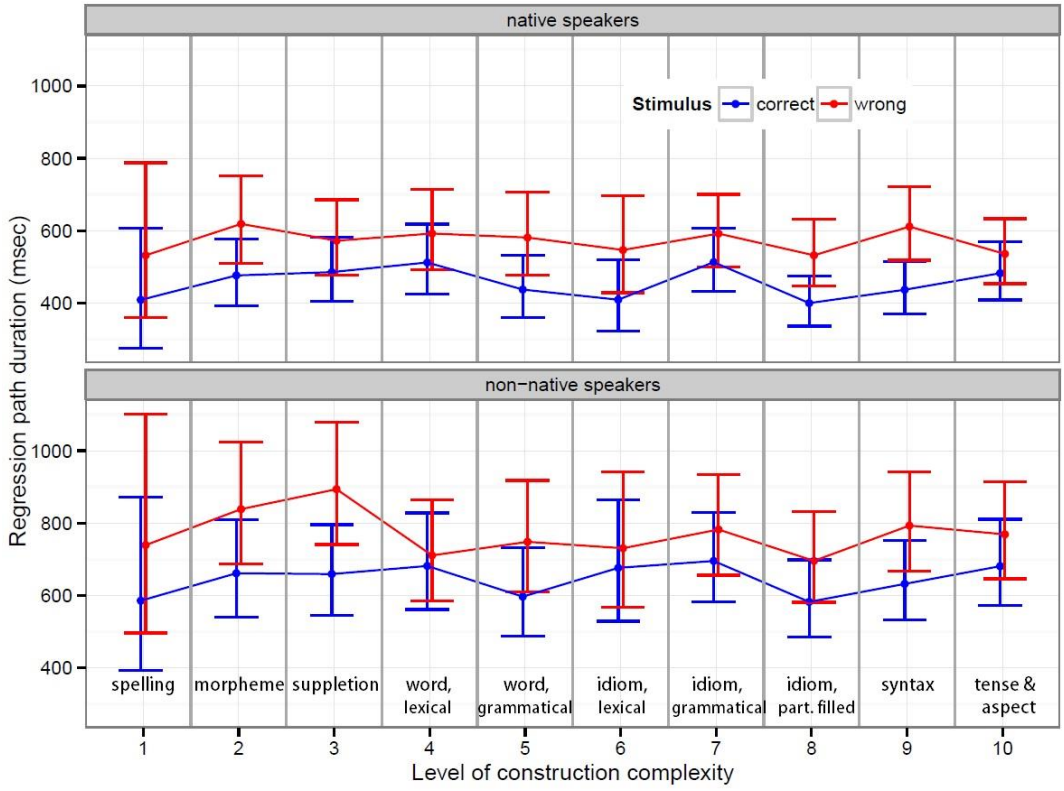
an example of a typical error of German learners (Swan and Smith 1987: 42). Sentences 5.52: 4-7, 10, 15, 16, 20-22 are known to be typical errors of Russian learners and partly German learners as well (ibid.). Russian does not have the corresponding progressive and perfect forms of the verb, so learners tend to use the simple tenses in English for all functions. German does not have progressive forms either, and the similarity of the simple past and the present perfect forms often works as a false friend. Several items above are examples of using the Simple Past with particular adverbials where the British English favors the Present Perfect (5.52: 12-14, 17, 19). This feature is spread around the world, from traditional L1-varieties to Pidgins and Creoles. The opposite case, using the Present Perfect for the Simple Past (5.52: 18), is somewhat less common (see eWave).

In Figures 5.25 *a* and *b*, the mixed effects models estimate the reading times for the above described levels as a factor, contrasting the correct and wrong stimuli. I present the same analysis for two time measures: the regression path duration (RPD) and the total reading time (TRT). The reason for this is the assumption that, for example, in an error type that does not affect the surface structure of the language (tense and aspect), the processing latency may not appear in the first pass, and the inconsistency may not be immediately clear enough to make the readers jump back to the previous parts of the sentence (that is captured by the RPD). Instead, subjects might want to revisit the conflicting part after they have processed the whole sentence, and any second, third, etc. pass over the critical region will appear in the TRT. In other words, a significant difference between the correct and wrong stimuli in Figure 5.25 *a* means that this type of construction violation is immediately noticed, processed within the same region, and probably triggers regressive movements. If the significant difference first appears in Figure 5.25 *b*, it signals that the error at this construction level is noticed later, and the reader goes back to revisit it. This is exactly the case of spelling mistakes in native speakers (Level 1) and wrong words (Level 4).

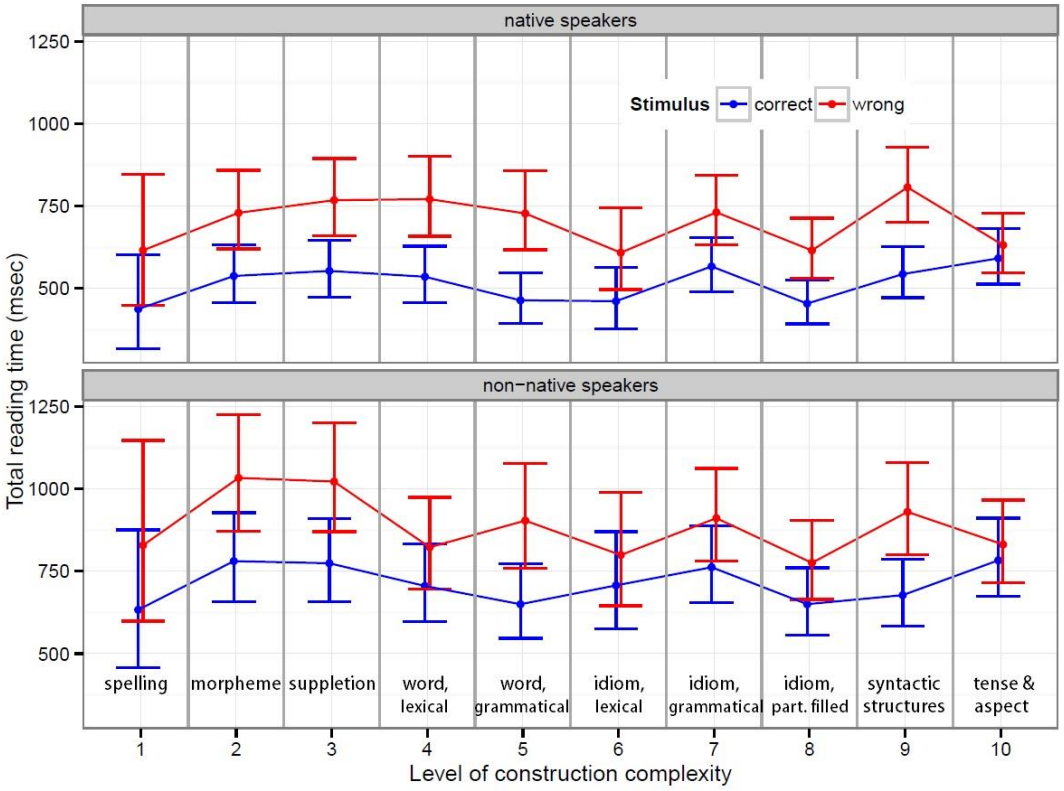
My expectation was that spelling errors not involving constructional violations in their acoustic form would not be significantly distinguished from the correct stimuli at all. But this is true only for early stages of processing. Obviously, orthographic mistakes do not interfere with visual word recognition; however, L1 readers go back to them and look at them significantly longer than at their controls. I had expected this effect for the tense and aspect inconsistency, but it did not appear. It turned out that this delicate error type does not entail any processing difficulty. It is another piece of evidence that the mismatching adverbials discussed in Chapter 4 were not appropriate material to test the effect of distance in error processing because they belong to **the**

only type of deviations from the norm readers pay less attention to (see Figure 5.25b). At the same time, they are discussed in terms of 'errors' and not 'strangeness', 'oddness', etc.

a.



b.



Figures 5.25. *RPD and TRT by level of construction complexity*

The present experiment demonstrates that tense and aspect 'errors' are different from the other error types. It seems that the reading times significantly increase only for 'hardware' violations and not for 'software' inconsistencies. It implies that either there is no clash at the conceptual level while processing tense and aspect 'errors', or the normal time is enough to solve it. An ERP study would be more informative here.

Generally, native and non-native speakers have a similar pattern in the perception of the levels. The largest differences are found in the reaction to Levels 3 and 6. In native speakers, errors involving suppletive forms are only close to be significantly distinguished from the controls in the RPD, while for the non-native speakers, they impose the greatest processing cost. In the TRT (Figure 5.25 *b*), suppletion remains one of the most time-consuming errors, and obviously, native speakers 'catch up' through regressions to the wrong forms. Unusual word strings (Level 6) did not attract any attention of the non-native speakers but required a significant processing effort from the native speakers. The reason for it could be that lexical idioms were slightly more transparent for the German learners or, rather, that non-native speakers have less experience with set phrases, and any word string, with a literal or figurative meaning, is potentially an idiom and is read at the same pace (learners also looked at the correct idioms longer). By contrast, native speakers accumulate more frequency information regarding idiomatic constructions in their language development, and they react more strongly if the word string is infrequent, so that in both RPD and TRT the difference from the controls is significant. Similarly, wrong words (lexical errors in Level 4) did not look suspicious for the German learners in the beginning, but the participants apparently returned to them more often than it would happen by chance. The native speakers were also less disturbed by the wrong meaning of the words in the beginning, but in the end, this error type turned out to be one of the most irritating. The errors at the other levels (mostly grammatical) had a significant difference from the correct stimuli in both time measures.

The most challenging errors for the native speakers in the RPD are at the level of morpheme and syntax, i.e. the most simple and the most complex levels of constructions. In the non-native speakers, the same tendency is outperformed by suppletion. In the TRT, the largest processing cost is produced by the violations of the grammatical properties of the word and syntax for the native speakers, and the same levels plus morpheme and suppletion for the non-native speakers. One puzzle remains: why are certain levels looked at significantly longer in correct sentences? For example, grammatically correct word strings that have to be learned as a unit (Level 7 included phrasal verbs and other uses of prepositions) required significantly more effort than the

neighboring categories of lexical and partially filled idioms from both groups of speakers in the RPD.

For the next analysis, only grammatical errors have been selected. After cutting off the spelling mistakes, tense and aspect, lexical words and idioms, the diagram in the RPD appears as a somewhat symmetrical figure with expansions at the extremes and a narrowing in the middle (Figure 5.26). This becomes possible if the assumed cline of construction complexity is treated as a numeric variable. In this case, the levels are not analyzed independently of each other but contribute to the overall tendency. The interaction of the stimulus and the level in the quadratic form has a significant effect, t = -2.68.

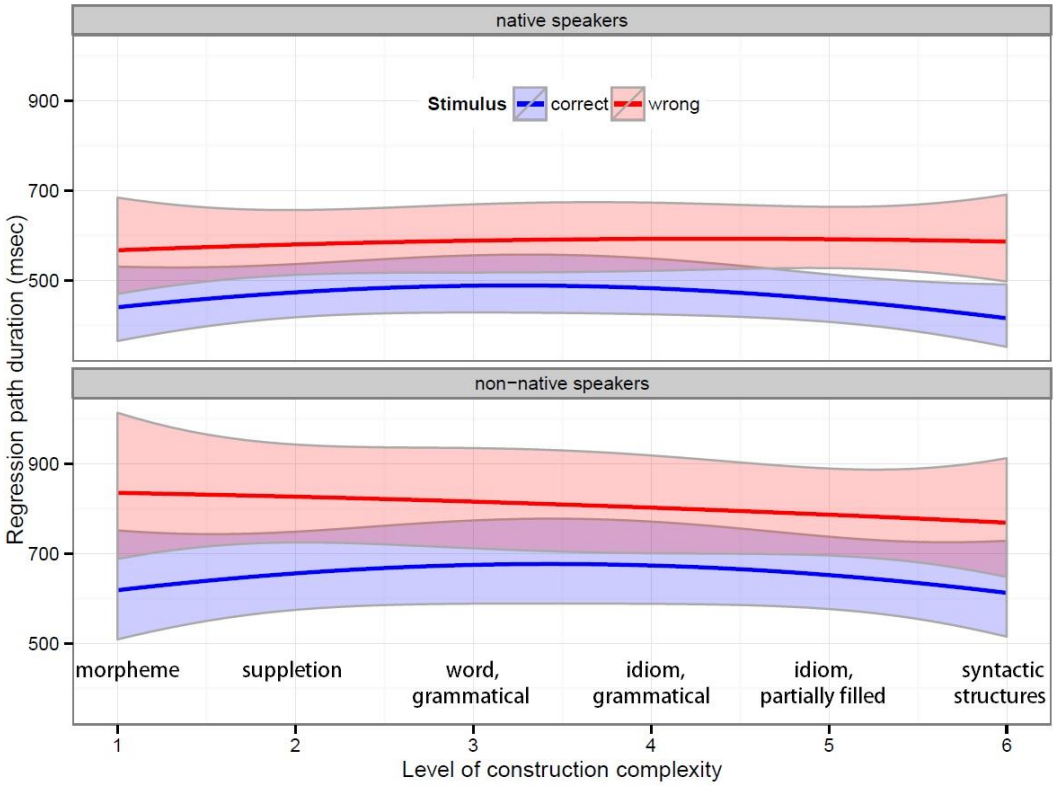


Figure 5.26. *RPD by level of construction complexity as a numeric variable*

	Estimate	Std. Error	t value
(Intercept)	6.496	0.049	133,789
poly(Level.num, 2)1	-0.862	2.541	-0.339
poly(Level.num, 2)2	-1.799	2.498	-0.72
stimulus[S.correct]	-0.116	0.007	-16.605
language[S.native speakers]	-0.166	0.027	-6.16
rpd.minus1	0.001	0.001	-5.519
rpd.minus2	0.001	0.001	-1.929
region.length	0.051	0.004	12.709
age	0.005	0.008	0.571
sex[S.f]	0.029	0.036	0.79
Language_major[S.no]	0.076	0.032	2.391
poly(Level.num, 2)1:stimulus[S.correct]	-0.223	0.491	-0.454

poly(Level.num, 2)2:stimulus[S.correct]	-1.322	0.493	**-2.68**
poly(Level.num, 2)1:language[S.native speakers]	0.379	0.485	0.781
poly(Level.num, 2)2:language[S.native speakers]	-0.355	0.485	-0.732
stimulus[S.correct]:language[S.native speakers]	-0.007	0.007	-1.065
stimuluscorrect:trial	0.001	0.001	-2.897
stimuluswrong:trial	0.001	0.001	-2.513
poly(Level.num, 2)1:stimulus[S.correct]: language[S.native speakers]	-0.995	0.488	**-2.036**
poly(Level.num, 2)2:stimulus[S.correct]: language[S.native speakers]	-0.15	0.492	-0.306

Table 5.6. *The effects of the levels model (RPD)*

The statistically significant interaction of the stimulus and the level of construction complexity in the quadratic form confirms that the largest distance between the correct and wrong stimuli appears towards the extremes, namely wrong morphemes and syntactic structures. While the main effect of the native language is always significant (in this model t = -6.16, sum contrasts), meaning that non-native speakers are generally slower, it usually affects the interaction of the main predictor with the language, implying that non-native speakers are also significantly slower in reaction to many levels of the variable on the X-axis, no matter whether the stimulus is correct or wrong. However, this time it is not the case. Instead, there is a significant three-way interaction of the levels in the linear form, the stimulus and the native language (t = -2.036). It means that there is an effect of the native language in perception of errors at particular levels. In Figure 5.26, it is observable that as the construction complexity increases, the reading times for wrong stimuli ascend for native speakers and descend for non-native speakers in the RPD. Thus, ignoring the native language, the general tendency is that the most simple and most complex errors require the greatest processing cost, but zooming in to the languages, we find that native speakers are more disturbed by wrong syntactic structures (the level with the largest scope), whereas non-native speakers spend more time solving local morpheme problems (get entangled in the low-level constructions). But this is the case only before the participants exit the critical region and continue reading the sentence.

Unlike RPD, the TRT (in Figure 5.27) includes the gazes entering the region from the right. In this analysis, the three-way interaction disappears, which implies that, in the end, native and non-native speakers perceive errors depending on the complexity level in a similar way. Obviously, both language groups outbalance by revisiting the other extreme: native speakers return to the low-level constructions, and non-native speakers attend to the high-level constructions.

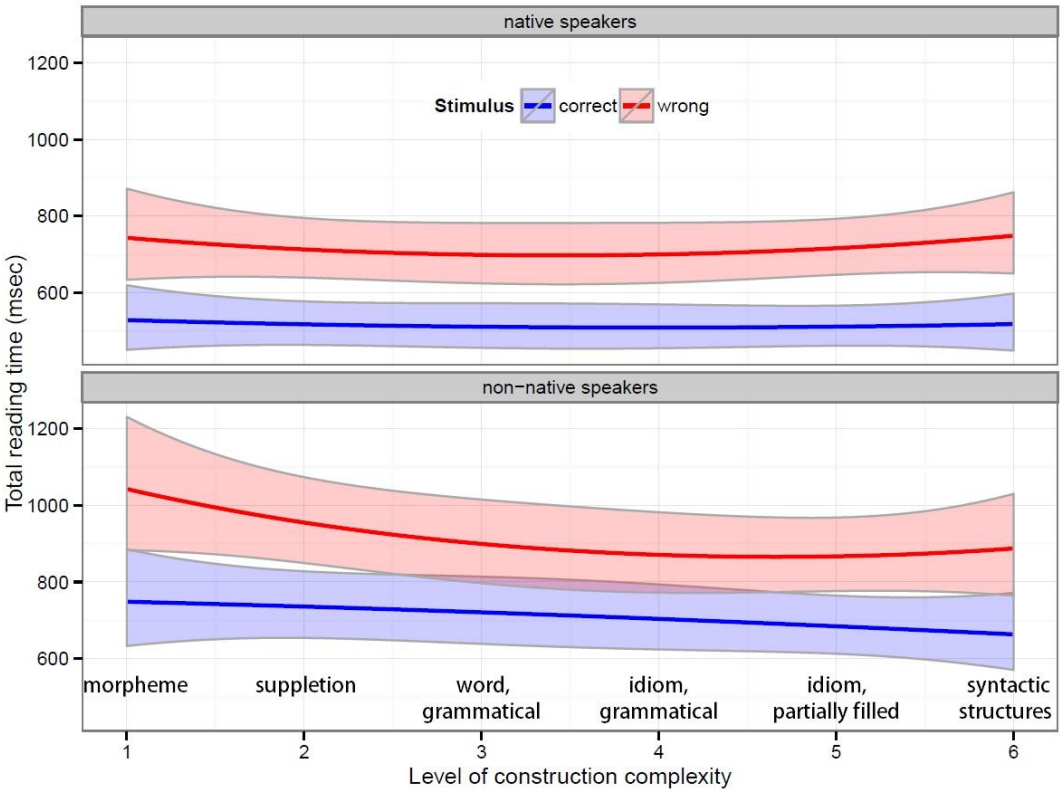


Figure 5.27. *TRT by level of construction complexity as a numeric variable*

	Estimate	Std. Error	t value
poly(Level.num, 2)1:stimulus[S.correct]	-0.011	0.473	-0.023
poly(Level.num, 2)2:stimulus[S.correct]	-0.957	0.475	**-2.013**
poly(Level.num, 2)1:language[S.native speakers]	1.632	0.468	3.487
poly(Level.num, 2)2:language[S.native speakers]	0.127	0.468	0.272
stimulus[S.correct]:language[S.native speakers]	-0.019	0.007	-2.82
poly(Level.num, 2)1:stimulus[S.correct]: language[S.native speakers]	-0.37	0.471	-0.785
poly(Level.num, 2)2:stimulus[S.correct]: language[S.native speakers]	0.396	0.475	0.833

Table 5.6. *The selected effects of the levels model (TRT)*

The fact that the most time-consuming errors are found at the beginning and at the end of the assumed construction complexity scale remains intact (t = -2.013), given that the reading times of all informants are analyzed together as one group. However, in Figure 5.27 most curves appear in the non-native speakers' diagram, which means that this effect is more relevant for them. Native speakers have two almost straight lines that are close to be parallel. It indicates that no matter which level of construction complexity is affected by the error, sooner or later the gaze will be attracted to it almost to the same degree. It is also remarkable and well expected that the reaction to correct stimuli in native speakers is a flat line. It is strengthened by the absence of the main effect of the complexity levels and means that, generally, constructions do not distinguish themselves by processing ease or processing difficulty; it only makes a difference when there is an error.

5.6 Summary

The analysis of the reading behavior shows that, generally, errors produce an extra processing cost and slow down the perception of the next few words. Readers learn to recover from errors already in the first pass (immediately after they have seen them), as they consistently avoid regressive movements towards the end of the task. There is little chance for errors to remain unnoticed because they reduce skipping in native speakers. Non-native speakers do very little skipping anyway, but as their L2 proficiency progresses, they learn to cope with errors more easily.

Almost all errors are detected by native speakers, if not in the earlier stages of processing, surely later on. The formal properties of errors such as error size, position in the sentence or sentence length do not have a particular effect on final error processing, although some error types may be parsed differently at intermediate stages.

Reading times increase primarily for surface-level errors. Errors in pragmatics or tense and aspect are not significantly distinguished from their controls. The reason for such a result in this experiment could be that a stimulus consisting of only one sentence provides too little context to trigger stronger reactions. However, the fact remains that, in equal conditions, tense/aspect and pragmatics inconsistencies are processed differently from other grammatical errors.

Error processing is complicated by the discovery of incongruence. Even if the meaning of the utterance is clear, as in the case of homophones, transfer phenomena, or use of 'eye dialect', it does not make reading faster. It is likely that participants are engaged in the stimulus evaluation and solving the conflicting information, which requires the same amount of time as other errors, or in case of the non-native speakers, even more.

The major difference between reading in L1 and in L2 is that native speakers incline towards top-down reading, while leaners tend to bottom-up decoding. This finding reconciles very well with the results of error evaluation studies which discovered that native speakers prioritize clarity, content and organization of the learner's text, while non-native speakers, primarily teachers, set rule infringement as crucial. In line with these observations, the native speaking participants in my eye movement experiment were able to spot errors at all distances (up to six words between the mismatching parts), indicating that they have a large sensitivity scope. The German learners of English did not react to errors at the distance beyond three words, which means they are only sensitive to local errors. This result supports the shallow structure hypothesis (Clahsen and Felser 2006a).

Also in line with this hypothesis, the non-native speakers in my eye-tracking experiment were mostly attracted by errors in smaller constructions with morphological functions. Native speakers, in contrast, ignored spelling mistakes and paid less attention to the violations of smaller constructions in the first pass but leaped over them, focusing more on the errors in large syntactic structures. Unlike non-native speakers who persisted in processing the simpler constructions, native speakers revisited them, so that, in the end, errors at all levels of construction complexity were equally attended.

Figuratively speaking, a native speaker looks for the object with a wider spot light and tries to see a bigger picture first. A non-native speaker searches with a narrow spot light and is so preoccupied with analyzing the details that the bigger picture escapes from his view. It is not surprising that teachers admit as unsuccessful their attempts to teach foreign language learners some L1 reading skills which presuppose skimming through the text. Learners have to be given enough time for the thorough text decoding before they can use it as a basis for developing their communicative competence.

Finally, there is a methodological conclusion. This part of the eye-tacking experiment based on a heterogeneous mix of errors was organized differently from the properly structured and balanced design of the part with mismatching adverbials discussed in Chapter 4. There was a risk that the noise produced by the enormous variety across the stimuli would stir all patterns and consume the effects. However, the presented results are most often in line with the general tendencies observed in eye movement studies (e.g. spillover and wrap-up effects, skipping and regressions, reaction to the position in the sentence – converged across completely different stimuli). Moreover, the results of this experiment are consistent with other psycholinguistic studies and theories of language processing. For example, they confirmed that the perception of non-native speakers is slower in speed, shorter in span (effect of distance) and shallower in depth (construction complexity). Besides, this was an eye-tracking experiment (as opposed to Gibson et al.'s off-line methods) providing evidence that errors involving deletions are processed significantly faster than those involving insertions and, therefore, might be more subject to internal repairs. In the end, there are all reasons to trust the conclusions drawn from this part of the experiment, in spite of the non-traditional design.

6. Processing Cost vs. Error Evaluation

This chapter investigates the **correlations** between **error evaluation** elicited in off-line ranking tasks and the **error processing cost** derived from the eye-tracking experiment. The analysis is carried out separately for a selection of **non-standard features** and another selection of **learner errors**. The correlations are computed **between the tasks** and, for non-standard features, also **between the participant groups**, i.e. native vs. non-native speakers. Within the group of native speakers, the evaluation data are analyzed across the **L1 varieties of English**. For non-native speakers, the effect of the **proficiency level** on evaluation is discussed. Besides **statistical information**, the results include an **examination of the sentences** for which the evaluation and processing data or the speaker groups mostly converge and diverge.

6.1 Hypothesis and Method

An overview of error evaluation studies carried out by means of questionnaires was given in Chapter 2. The method of calculating the error processing cost based on eye movement data was presented in section 5.2. Now it is possible to find out how much the reading times for errors correlate with the assessment of their severity.

The investigation of literature in Chapter 3 and the results of the eye-tracking experiment discussed in Chapters 4 and 5 brought to the revision of the initial hypothesis that error gravity is proportional to the processing cost. There is evidence, for example, that a mismatch between the tense form of the verb and the temporal adverbial is rated as not acceptable but does not involve any additional processing cost. The revised hypothesis holds that there is a difference between the declared understanding of correctness and the actual perception of incorrect utterances. For some errors, a latency in the processing times may not necessarily determine the low ranking, and vice versa, errors rated low on the acceptability scale may not involve any processing difficulty.

Ideally, the following investigations should compare the conscious and unconscious perception of errors. Although it is not correct to claim that an eye-tracking experiment entirely reflects unconscious processing, the reading procedure in the present study was organized in a way to maximally direct the participants away from error detection and the conscious analysis of errors. The readers had not been informed of errors in the stimuli, and due to the

comprehension questions following each item, the participants concentrated on the contents of the sentences.

Evidence that comprehension effectively took place can be found in the results of the "yes" or "no" questions following each stimulus sentence. The whole experiment was organized as a comprehension test. The participants answered correctly in 82% of the cases. Native speakers scored only a little better than the German learners (83% and 80% correspondingly), which also indicates that the language material was clear enough for non-native speakers.

In an error evaluation study, by contrast, participants are instructed about errors and are asked to assess the grammaticality of sentences. In the following sections, I will compare the ratings and the processing cost for the same sentences in two different perception tasks.

6.2 Non-standard features

As already mentioned in section 5.5.1, native and non-native speakers differ in their perception of the L1, ESL type of errors. For both groups of participants, the distinction between the wrong and correct stimuli is significant, but the processing cost for the non-native speakers is twice as large as for the native speakers. Does it also mean that learners would have assessed non-standard features twice as severe?

To answer this question, I made a selection of 50 sentences out of the L1, ESL category and set up a web-based survey[27]. Native speakers of English L1 varieties (total 78, including 15 from the UK, 38 from the USA, 9 from Canada, 4 from Australia, and 12 from other countries) and non-native speakers (total 24; 12 of them grew up in Germany) rated the acceptability of the sentences containing non-standard features based on a 5-point scale from "This is bad English" (1 star) to "This is a perfect sentence in English" (5 stars). The ratings (5100 observations) were analyzed based on mixed-effects models with the participants' age (20-35), gender (65% female and 35% male) and language studies (language majors 57%) as fixed effects.

According to the native speakers, the following sentences are considered least acceptable (for the full table see Appendix 7).

Rank	Sentence	Rating
1	***They're not left** school yet.*	*1.25*
1	*I'll tell you what we might **should** do about it.*	*1.25*
2	*This is the man **what** painted my house.*	*1.39*

[27] https://kwiksurveys.com/s.asp?sid=s82ycgyu7dqzkcz488409#/0

2	*Ron said there last night that you **was** his ex-wife and his friend only.*	*1.39*
2	*And it is good to be back with your friends and get **catched** up with school again.*	*1.39*
3	*I **gone** on a mission trip with my church.*	*1.41*
4	*My daughter, **what** lives in London, is currently a student.*	*1.43*
5	*I probably wouldn't let him go just because I **wants** him here.*	*1.45*
6	*I am in the mortal pain, I have **the toothache**.*	*1.46*
7	*As for Tant, he **weren't** so scared as he pretended.*	*1.59*

Table 6.1. *Items with lowest ratings in native speakers*

The informants were quite consistent in their evaluations in that examples of the same error type received a similar ranking, for instance, in the use of an interrogative pronoun *what* in the function of a relative pronoun (*the man what painted my house* and *my daughter, what lives in London*), wrong subject-verb agreement in number or person (*you was, I wants* and *he weren't*), or wrong form of an irregular verb (*get catched up with school* and *I gone on a mission trip*). In most of the severest errors there are two adjacent words that do not match, and the second word should be corrected.

There were no control conditions in the survey to prove that correct sentences would receive high ratings. However, the following non-standard sentences considered by the native speakers the closest to standard English are, in fact, quite common and almost function as controls:

Rank	Sentence	Rating
31	*I mean, he's so wrong in this thing, and he's **like**, oh, the deficit is out of control.*	*3.04*
32	*Some of the top models in the world aren't **the most pretty** girls.*	*3.17*
33	*I'm wondering what **am I** going to do after Easter.*	*3.2*
34	*Patients **are wanting** a lot of personal attention.*	*3.22*
35	*Let me just ask you, **Were you ever** in a car with a pair of binoculars near their house?*	*3.36*
36	*We've **had nice** conversation with people from all over the world.*	*3.41*
37	*My father said that was the only way you could get a **real good** look at the faces, and he was right.*	*3.44*
38	*Take them **to market**, please.*	*3.45*
39	*We realized that each of us was talking about something quite different **than what** the other thought.*	*3.59*
40	*You better come **quick**, I'm not sure how long it'll last.*	*3.9*

Table 6.2. *Items with highest ratings in native speakers*

Among the most acceptable items in native speakers appear the *be like* as a quotative (considering how common it is in the spoken language, the only reason to qualify it as an 'error' may be that it looks unusual in writing), the extension of the analytic marking of comparison to a 2-syllable adjective (*the most pretty*), the inverted subject-verb order in an indirect question (*I'm*

wondering what am I going to do), the extension of progressive to a stative verb (*patients are wanting*), the levelling of the difference between the Present Perfect and the Simple Past (*were you ever…?*), the use of the zero article (*we've had nice conversation* and *take them to market*), and an adverb having the same form as an adjective (*come quick* and *a real good look*). Many of these features, especially the latter, have a high attestation rate in the eWAVE (Kortmann and Lunkenheimer 2013).

The estimated mean (intercept) of the native speakers' ratings is 2.22. Non-native speakers rated the sentences a bit lower – 2.19 on the average.

The sentences with the lowest ratings by the non-native speaking participants are presented in the following table.

Rank	Sentence	Rating
1	*Ron said there last night that you **was** his ex-wife and his friend only.*	*1.37*
2	*"Jack!" I **says** putting my elbows on that cushion. "How are you doing?"*	*1.4*
3	*My daughter, **what** lives in London, is currently a student.*	*1.44*
4	***They're not left** school yet.*	*1.55*
5	*This is the man **what** painted my house.*	*1.56*
6	*I've always fancied **meself** married to a lordship and swanning up and down the stairs as lady.*	*1.57*
6	*The man **I met's** girlfriend is a real beauty.*	*1.57*
7	*Next morning we went and shook hands at the Ford factory **for to thank** them for their support.*	*1.59*
7	*And quite frankly, Bill Clinton is a man trying to get **he** and his wife back in the White House.*	*1.59*
8	*I **gone** on a mission trip with my church.*	*1.6*

Table 6.3. *Items with lowest ratings in non-native speakers*

Out of the ten sentences that received the lowest ratings by the non-native speakers, five items overlap with the corresponding list of the native speakers (Table 6.1). Again, *you was* and *I says* – wrong subject-verb agreement (Ranks 1 and 2), the relativizer *what* instead of *who* (Ranks 3 and 5), *I gone* and *they're not left school*. Interestingly, the latter example only needs to repair one letter to be correct, and this failure is qualified as a severe grammatical error and not excused as a typographical mistake. The new error types in the non-native speaker list are the non-standard uses of pronouns, such as *meself* – an object pronoun form serving as base for a reflexive (Rank 6) and *get he* – an object pronoun used instead of a reflexive (Rank 7), the wrong attachment of possessive (Rank 6) and insertion of a preposition (Rank 7).

The following sentences were found by the non-native speakers the closest to perfect English.

Rank	Sentence	Rating
32	*Let me just ask you, **Were you ever** in a car with a pair of binoculars near their house?*	*2.71*
33	*If **I'd have made** the decision myself I'd have either got disqualified or got a two-shot penalty.*	*2.79*
34	*Patients **are wanting** a lot of personal attention.*	*2.82*
34	*We've **had nice** conversation with people from all over the world.*	*2.82*
35	*My brother **and me** were late for school.*	*2.89*
36	*I mean, he's so wrong in this thing, and he's **like**, oh, the deficit is out of control.*	*3.06*
37	*Some of the top models in the world aren't **the most pretty** girls.*	*3.15*
38	*You better come **quick**, I'm not sure how long it'll last.*	*3.56*
39	*We realized that each of us was talking about something quite different **than what** the other thought.*	*3.57*
40	*My father said that was the only way you could get a **real good** look at the faces, and he was right.*	*3.72*

Table 6.4. *Items with highest ratings in non-native speakers*

In the lists of the ten most acceptable deviations from Standard written English, native and non-native speakers share eight sentences. The German learners considered the form of conditional *If I'd have made* more tolerable than it appeared in the assessment of the native speakers (Rank 29), but the rating (2.79) is still closer to the mean (2.19) than to the maximum value (3.72). The evaluation of *my brother and me* showed a greater disagreement between the L1 and L2 participants. The use of *me* instead of *I* in a coordinate subject was assigned Rank 35 in non-native speakers, while native speakers found it less acceptable placing it at Rank 21 with the Rating 2.21. Although, according to the eWave, this feature has a high attestation rate of 89% and pervasiveness of 83%, its grammaticality was judged slightly below the mean (2.22).

By calculating the difference between the evaluation ranks for the same items (native speakers' rank "minus" non-native speakers' rank), we can analyze the sentences in the assessment of which the L1 and L2 participants mostly disagreed. In Tables 6.5 and 6.6 below is a selection of sentences from the top and the bottom of the list, with largest rank differences. A positive value indicates that the item was low in the native speakers' table (was judged as more acceptable) and higher in the non-native speakers' table (less acceptable). In other words, the learners found these deviations from the norm 'worse' than they are perceived in L1.

	Sentence	Rank difference
1	*The man I **met's** girlfriend is a real beauty.*	*22*
2	*"Jack!" I **says** putting my elbows on that cushion. "How are you doing?"*	*14*
3	*Take them **to market** please.*	*12*

4	*And quite frankly, Bill Clinton is a man trying to get **he** and his wife back in the White House.*	8
5	*I **seen** him a while ago, but I don't know where he is now.*	7
6	*I've always fancied **meself** married to a lordship and swanning up and down the stairs as lady.*	7
7	*I'm ready to tell you, child, you **done** got your wish.*	5
8	*It's hot and there's a storm **a-coming**. This is a little tender moment.*	4
9	*I'm wondering what **am I** going to do after Easter.*	4
10	*That President has two Secretary of **States**.*	3

Table 6.5. *Items rated more acceptable by native speakers*

The largest possible rank difference could be 49, and the largest rank difference elicited in this analysis is less than half as high – 22. Besides that, the values rapidly decrease, leaving a substantial gap between 22 and 14 and then between 12 and 8. Thus, it can be argued that there are only three items rated by the native speakers as much more acceptable than they were judged by the non-native speakers: *the man I met's girlfriend* (the absolute leader)*, I says* and *to market*.

The negative values in Table 6.6 mark the features which the native speakers rated stricter than the non-native speakers did. Note that the rank difference is less than one third from the largest possible value (less disagreement) and subsides more gradually.

	Sentence	Rank difference
1	*And it is good to be back with your friends and get **catched** up with school again.*	-14
2	*My brother **and me** were late for school.*	-14
3	*I am in the mortal pain, I have **the toothache**.*	-13
4	*As for Tant, he **weren't** so scared as he pretended.*	-11
5	*I'll tell you what we **might should** do about it.*	-11
6	*I **don't have no** reason to worry about it.*	-10
7	*Then Oleta's boy got **hisself** shot the day after Rondell and DeMarcus delivered their messages.*	-7
8	*He's making the battle a lot **more easier** for me.*	-7
9	*They've been promoting **theirself** as the best women's club in the world.*	-6
10	*Well, back in **them** days I was a reporter for the Daily Mirror.*	-6

Table 6.6. *Items rated less acceptable by native speakers*

It is not surprising that the regularization of the verb *catch* in *get catched up* led to disagreement between the native speakers and learners – presumably, the frequent *caught* is entrenched in L1. But the use of *me* instead of *I* in a coordinate subject surely does not involve psycholinguistic reasons but rather social ones. Interestingly, the double negation in *I don't have no reason* is perceived more negatively in L1 than learners think. It is not clear why *he*

weren't is so different from *I says* (in Table 6.5) to produce an opposite reaction being overestimated by one group and underestimated by the other. Also, for a non-native speaker of English, it is difficult to assess what is so dramatically wrong about *I have the toothache* and why it is considerably worse than *to market* which native speakers placed at Rank 6 and Rank 38 correspondingly (rank difference 32!), compared to non-native speakers' Ranks 19 and 26 (rank difference only 7). Maybe that is exactly what illustrates the point – the lack of constructional knowledge inherent in L1. Native speakers' competence includes knowledge about frequencies, transitional probabilities (e.g. in COCA, *have a* occurs 2.7 times more frequent than *have the*), subtle differences in the meaning and social attitudes that are difficult to acquire in the classroom.

On the whole, the correlation of native speakers' and learners' ratings is r = 0.85, which indicates a very strong relationship. However, processing times for the same sentences (based on the RPD) correlate only at 46%. It illustrates that native and non-native speakers share common knowledge about grammaticality and correctness in English, but when it comes to internal processing without pointing to errors, L1 and L2 readers operate differently. At the same time, both groups are similar in relation to the task. The fact that an error rated very low on the acceptability scale requires more processing time is only to some extent true (cost : evaluation, r = -0.36 in native speakers and r = -0.41 in non-native speakers). This negative correlation means that longer reading times caused by the error only moderately predict poor grammaticality judgements in both L1 and L2 participants.

Such low correlation values presuppose that there are a number of sentences for which the high processing cost did not correspond to low ratings, and, vice versa, some easily processed errors were judged as not acceptable. I have compared the ranks in the two tasks for each of the 50 sentences and calculated the difference between the rank in the reading experiment and the rank in the acceptability judgment. The ranks for both tasks are organized in such a way that the most costly and most severe errors are at the top (Rank 1, 2, 3…). In the table below, a high positive value means that the sentence was down on the processing scale (required little time), but was high on the judgment scale (received low ratings). The results refer to the native speakers.

	Sentence	Rank difference
1	*And it is good to be back with your friends and get **catched** up with school again.*	*42*
2	*And it's crazy, and he **don't** even know how to spell my name.*	*32*
3	*Ron said there last night that you **was** his ex-wife and his friend only.*	*29*
4	*My daughter, **what** lives in London, is currently a student.*	*28*

5	*I **don't have no** reason to worry about it.*	*23*
6	*I am in the mortal pain, I have **the toothache**.*	*22*
7	*I **seen** him a while ago, but I don't know where he is now.*	*22*
8	*I **gone** on a mission trip with my church.*	*21*
9	*Then Oleta's boy got **hisself** shot the day after Rondell and DeMarcus delivered their messages.*	*21*
10	*As for Tant, he **weren't** so scared as he pretended.*	*16*

Table 6.7. *Items with largest positive rank differences in native speakers*

The following table presents the results for the non-native speakers.

	Sentence	Rank difference
1	*I'll tell you what we **might should** do about it.*	*32*
2	*And it is good to be back with your friends and get **catched** up with school again.*	*29*
3	*Ron said there last night that you **was** his ex-wife and his friend only.*	*24*
4	*My daughter, **what** lives in London, is currently a student.*	*20*
5	*That President has **two Secretary of States**.*	*20*
6	*I **don't have no** reason to worry about it.*	*19*
7	*I am in the mortal pain, I have **the toothache**.*	*19*
8	*Hey, you **ain't** even real cops. What kind of cop's got three first names?*	*18*
9	*They've been promoting **theirself** as the best women's club in the world.*	*16*
10	*He **say** he'll take care of everything.*	*16*

Table 6.8. *Items with largest positive rank differences in non-native speakers*

Native and non-native speakers share a half of the Top 10 sentences that were processed relatively fast, but were judged as containing severe errors. Among them is a regularized past tense form of an irregular verb (*get catched up*), wrong subject-verb agreement (*you was*), the interrogative *what* used for the relative *who*, double negation (*I don't have no reason*) and the use of a definite article where Standard English requires an indefinite article (*I have the toothache*). Besides, both participant groups have examples of possessive pronouns as base for reflexives (*got hisself, promoting theirself*). But only the native speakers had past participles of an irregular verb used in the past tense context (*I seen, I gone*) and *he don't*, the latter being common in many varieties of English. Instead, the non-native speakers easily processed *you ain't* for which the rank difference in native speakers is only 6. In L1, it was a more costly feature (Rank 29) than in L2 (Rank 41), although the evaluation in both language groups coincided at Rank 23.

As was already mentioned, there are opposite cases, namely that were considered more acceptable examples of English but at the same time complicated processing. In the following tables, a large negative value indicates that the sentence had a high rank on the processing scale (cost much time) but

a low rank in the evaluation (was rated close to 'perfect' English). The results refer to the native speakers (Table 6.9) and to the non-native speakers (6.10).

	Sentence	Rank difference
1	*Take them **to market** please.*	*-30*
2	*We've **had nice conversation** with people from all over the world.*	*-26*
3	*We realized that each of us was talking about something quite **different than** what the other thought.*	*-14*
4	*My father said that was the only way you could get a **real good** look at the faces, and he was right.*	*-12*
5	*Tell me how that makes you feel and what **yall** think we should do at this point.*	*-12*
6	*The man I **met's** girlfriend is a real beauty.*	*-11*
7	*It's nice of all **youse** to have me here tonight.*	*-11*
8	*I've always fancied **meself** married to a lordship and swanning up and down the stairs as lady.*	*-11*
9	*Listen to me, son, I **been** thinking of what to do with you.*	*-9*
10	*Let me just ask you, **Were you ever** in a car with a pair of binoculars near their house?*	*-5*

Table 6.9. *Items with largest negative rank differences in native speakers*

In this analysis, native and non-native speakers share six sentences.

	Sentence	Rank difference
1	*Tell me how that makes you feel and what **yall** think we should do at this point.*	*-25*
2	*Let me just ask you, **Were you ever** in a car with a pair of binoculars near their house?*	*-23*
3	***If I'd have made** the decision myself I'd have either got disqualified or got a two-shot penalty.*	*-21*
4	*We realized that each of us was talking about something quite **different than** what the other thought.*	*-15*
5	*We've **had nice** conversation with people from all over the world.*	*-14*
6	*My father said that was the only way you could get a **real good** look at the faces, and he was right.*	*-11*
7	*I hope people **are liking** my voice and my music.*	*-11*
8	*Listen to me, son, I **been** thinking of what to do with you.*	*-10*
9	*Well, back in **them** days I was a reporter for the Daily Mirror.*	*-9*
10	*So be kind to him, he **be** a good man.*	*-8*

Table 6.10. *Items with largest negative rank differences in non-native speakers*

It is hard to believe that such a trivial error as an article deletion (*take them to market* and *had nice conversation*) is so difficult to process, but these particular examples, in fact, were looked at longer than most others (in L1 Rank 8 and Rank 10 correspondingly). This is not in line either with Figure 5.21 demonstrating that deletions are the easiest errors for processing, nor with Figure 5.23 showing that errors affecting articles require less time than those

involving other parts of speech. The similar holds for the adjectives used as an adverb (*a real good look*), although the cost ranks are much lower for this stimulus (25 in L1 and 29 in L2). It is not clear either why *different than* was looked at much longer than *different from* (Cost Rank 25 in L1 and 24 in L2), while both language groups had to acknowledge that it was close to a perfect English sentence (evaluation Rank 39 in both L1 and L2). In the same category of slowly processed but highly rated errors are non-standard pronouns (*yall, youse, meself, in them days*).

The reason why these results do not look consistent with the previous sections of the thesis may be due to the fact that the tables above refer only to the 50 sentences containing non-standard features, whereas all the previous findings (including Figures 5.21 and 5.23) present the analyses of the whole sample, including learner errors (more closely discussed in the following section). Obviously, when learner errors fall out of the list, non-standard uses of articles, pronouns and adjectives get promoted to higher cost ranks.

There is another implication. Participants look at such non-standard features remarkably longer not because they 'suffer' from 'processing difficulty', otherwise they would not have rated these sentences as close to 'perfect English'. Rather, readers gazed at them because it is unusual to see non-standard features in writing, in the context of a psychological experiment in a non-English speaking country. I suggest that the term "processing difficulty" is more appropriate for the items with a small rank difference and high ranks on both time-cost and evaluation scales, such as the following sentences for the native speakers.

	Sentence	Cost rank	Evaluation rank	Rank difference
1	*This is the man **what** painted my house.*	*5*	*2*	*3*
2	*They'**re not left** school yet.*	*3*	*1*	*2*
3	*I'll tell you what we might **should** do about it.*	*7*	*1*	*6*
4	*Then maybe you're thinking of that crowd that **was after hanging** about here crooning your name.*	*9*	*10*	*-1*

Table 6.11. *Items with high cost and evaluation ranks in native speakers*

These are the examples of processing difficulty for the non-native speakers.

	Sentence	Cost rank	Evaluation rank	Rank difference
1	*The man I **met's** girlfriend is a real beauty.*	*1*	*6*	*-5*
2	*Next morning we went and shook hands at the Ford factory **for to thank** them for their support.*	*2*	*7*	*-5*

3	*"Jack!" I **says** putting my elbows on that cushion. „How are you doing?"*	*7*	*2*	*5*

Table 6.12. *Items with high cost and evaluation ranks in non-native speakers*

Interestingly, the sentences and the deviation types that attracted attention but were rated as acceptable largely overlap for the native and non-native speakers (Tables 6.9 and 6.10), but the non-standard features for which "processing difficulty" may reasonably be applied are different for the L1 and L2 participants (Tables 6.11 and 6.12).

On the other extreme, there are sentences that received low ranks in both eye-tracking experiment and the evaluation study. The following items proved to be the easiest for native speakers.

	Sentence	Cost rank	Evaluation rank	Rank difference
1	*I'm wondering what **am I** going to do after Easter.*	*39*	*33*	*6*
2	*I hope people **are liking** my voice and my music.*	*34*	*30*	*4*
3	*You better come **quick**, I'm not sure how long it'll last.*	*37*	*40*	*-3*
4	*Let me just ask you, **Were you ever** in a car with a pair of binoculars near their house?*	*30*	*35*	*-5*

Table 6.13. *Items with low cost and evaluation ranks in native speakers*

The sentences below reflect the perception ease for non-native speakers:

	Sentence	Cost rank	Evaluation rank	Rank difference
1	*You better come **quick**, I'm not sure how long it'll last.*	*44*	*38*	*6*
2	*Some of the top models in the world aren't **the most pretty** girls.*	*43*	*37*	*6*
3	*Patients **are wanting** a lot of personal attention.*	*39*	*34*	*5*

Table 6.14. *Items with low cost and evaluation ranks in non-native speakers*

Thus, it can be argued that such non-standard features as using an adjective for an adverb (*come quick*) and extension of progressive to a stative verb (*people are liking*) are perceived as closest to Standard English by both native and non-native speakers.

It should be mentioned that the groups of native and non-native speakers in the evaluation study were not homogeneous as far as their varieties are concerned. The distribution of the mother tongue participants and foreign learners as well as their estimated ratings as a group (with confidence intervals)

can be found below in Figure 6.1 ("Other-NS" stands for native speakers and "Other-NNS" for non-native speakers coming from the countries not listed in the survey).

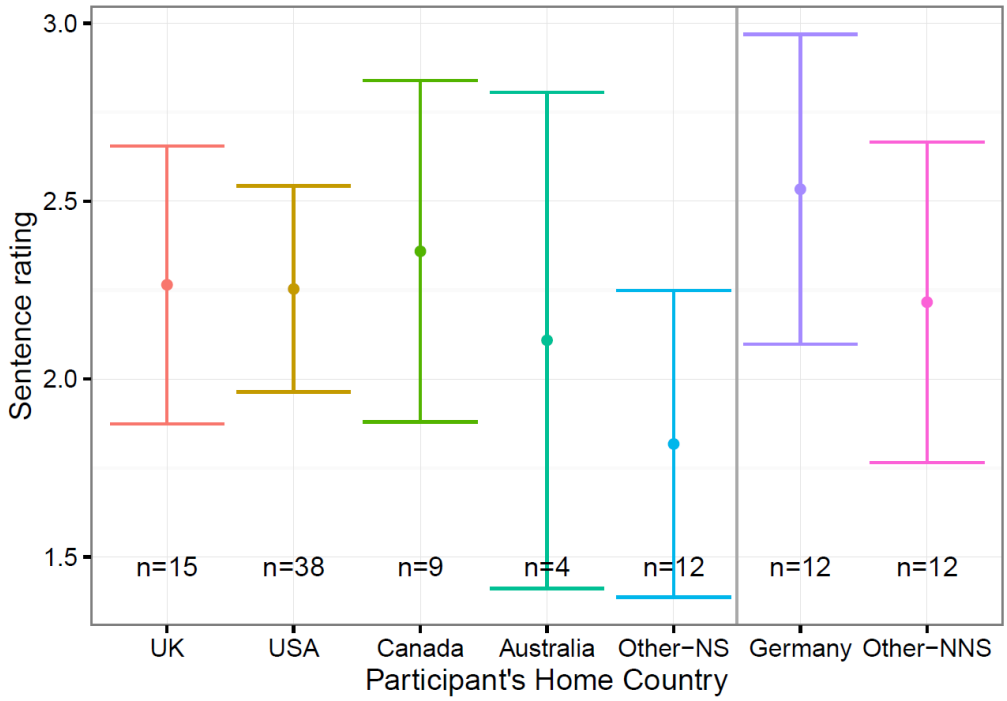


Figure 6.1. *Acceptability ratings depending on the variety*

The participants representing the standard L1 varieties (UK, USA, Canada and Australia) appear to have no significant differences in the overall rating of the 50 sentences containing non-standard features. The means are very close, and the stretch of the confidence intervals proportionally corresponds to the number of the informants. Remarkably apart stand the results of the participants who introduced themselves as native speakers and grew up in "other" countries. For technical reasons, there was no text box allowing to type in the particular place of origin, but the automatic tracking of the country where the survey was filled out shows that two participants of that category were in Singapore. Logically, if native speakers of English do not belong to the largest L1 varieties, there is a probability that they could be ESL speakers. If that is the case, it is interesting why their ratings are significantly lower than those of the participants from the UK, US and Canada. A suggestion can be made that ESL speakers are more concerned about correct language, and their attention is constantly drawn to typical errors native and non-native speakers make in English. They not only rate the sentences significantly lower but also show a slightly different pattern in the assessment of the same error types (the correlation of the Other-NS with the UK is 74% and with the US 75%). The evaluations of the Other-NS correlate more with the non-native speakers (Other-NNS) – 79%.

Compared to the L1 participants, English learners from Germany tend to give higher ratings, i.e. assessed the sentences to be closer to perfect English. Their evaluations correlate with those of the UK at only 67% and of the US at 65%. According to Figure 6.1 (that only illustrates the means), the other group of non-native speakers (coming from Italy, Greece, Russia, Poland and other countries) rated almost in the same way as the representatives of the standard L1 varieties. The ratings of the Other-NNS correlate with the UK at 75% and with the US at 83% – it is even higher than the correlation with the non-native speakers from Germany (68%). For some reason, the assessments of the German learners show the lowest correlations with all other speaker groups and tend to be too 'liberal'.

We can zoom in the largest groups of participants among the L1 varieties and focus on the differences in the evaluation of the non-standard features given by the speakers from Great Britain and the Unites States. In Figure 6.1, the UK and US participants show a remarkably uniform overall rating – the estimated means are exactly the same. The correlation of the UK and US ratings is also very high – 92%. It is the highest correlation value in this analysis. Again, by calculating the rank differences (UK rank "minus" US rank), we can find out in the evaluation of which sentences the representatives of these L1 varieties mostly disagreed.

	Sentence	UK rank	US rank	Rank difference
1	*Well, back in **them** days I was a reporter for the Daily Mirror.*	*18*	*12*	*6*
2	*They**'re** not left school yet.*	*4*	*1*	*3*
3	*And it is good to be back with your friends and get **catched** up with school again.*	*5*	*2*	*3*
4	*This is the man **what** painted my house.*	*7*	*4*	*3*
5	*Then maybe you're thinking of that crowd that was **after hanging** about here crooning your name.*	*13*	*10*	*3*

Table 6.15. *Items that UK speakers found more acceptable than US speakers*

As the rank difference of 3 is too miserable to contrast the varieties, the only item that the British speakers found slightly more acceptable than the Americans is using *them* instead of the demonstrative *those* as in *back in them days*. According to eWAVE (Kortmann and Lunkenheimer 2013), this feature is attested and pervasive in both North America and the British Isles. Anyway, the rank difference of 6 is small in comparison with the evaluation of other features which the US participants found more acceptable than their British counterparts. The latter turned out to be stricter in their assessment of many sentences relatively to the others in the sample. Thus, if the rank difference of

3 in Table 6.15 falls on the second item, the rank difference of -3 is arrived at only on the 29th position.

	Sentence	UK rank	US rank	Rank difference
1	*I'm ready to tell you child, you **done** got your wish.*	*5*	*20*	*-15*
2	*We realized that each of us was talking about something quite different **than** what the other thought.*	*25*	*38*	*-13*
3	*He **say** he'll take care of everything.*	*10*	*21*	*-11*
4	*Tell me how that makes you feel and what **yall** think we should do at this point.*	*16*	*27*	*-11*
5	*We've **had nice** conversation with people from all over the world.*	*26*	*37*	*-11*
6	*And quite frankly, Bill Clinton is a man trying to get **he** and his wife back in the White House.*	*8*	*18*	*-10*
7	*That President has two Secretary of **States**.*	*15*	*25*	*-10*
8	*Anybody knows that there**'s all** kinds of ways you get delayed as you get ready for flight.*	*14*	*23*	*-9*
9	*You better come **quick**, I'm not sure how long it'll last.*	*31*	*39*	*-8*
10	*I hope people **are liking** my voice and my music.*	*21*	*28*	*-7*

Table 6.16. *Items that US speakers found more acceptable than UK speakers*

The table is headed by the completive/perfect *done* (*you done got your wish*) – the feature that is not attested in the British Isles, but is one of the diagnostic features of AmE (Kortmann and Schröter, in press). *Different than* is much closer to 'perfect' English in the US, too. Second person plural forms (*yall, youse*) are also more pervasive in America. The zero marking of the third person singular (*he say*) as well as the use of the zero article where Standard English has an indefinite article (*we've had nice conversation*) are common neither in the UK, nor in the US, but rather in the Caribbean, Africa and Australia (see eWAVE).

Although it is curious to see in which sentences the UK and the US participants differed in their acceptability ratings, and the results are also plausible, it should be borne in mind that with such a high correlation of 92%, statistically speaking, there are no differences between these two varieties in the assessment of non-standard features. Moreover, there seems to be no difference between the participants from the UK, US, Canada and Australia, with the caveat of the relatively low number of the informants. The speakers of the largest L1 varieties appear as a solid group with very similar reactions. Who stands apart is the group of "native speakers" from "other" countries whose status as ESL speakers, unfortunately, cannot be confirmed.

Another curious detail concerning the perception of non-standard features in a foreign language is the dependence of the ratings on the level of proficiency and language/linguistics studies. The analysis refers to the 24 non-native speakers who assessed their proficiency in English on a 10-point scale. As in the eye-tracking experiment, they also checked whether they studied a language or linguistics as a major subject. The result of the mixed-effects model is presented in Figure 6.2.

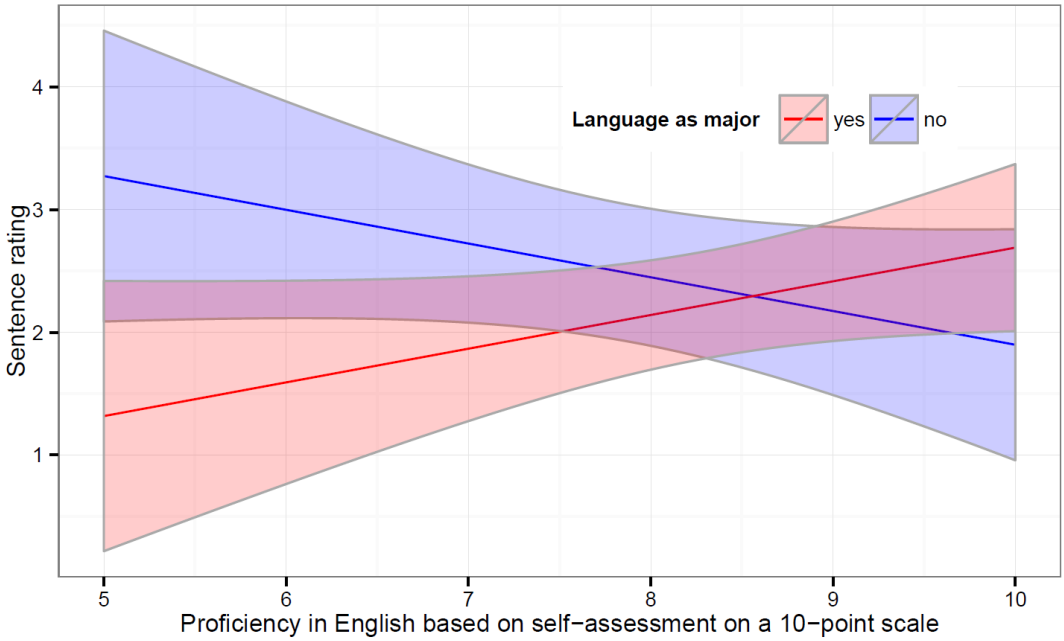


Figure 6.2. *Acceptability ratings depending on proficiency and language studies*

The interaction of the level of proficiency and studying a language as a major subject is statistically significant (t = -2.36). It indicates that as proficiency grows, linguists and non-linguists rate non-standard features differently. Those who assessed their proficiency as slightly over the middle in the range from "just a couple of phrases" to "near-native" and who are probably just starting their language studies prefer lower ratings, which signals that they perceive non-standard features as severe errors (violations of rules). However, learners with a better command of English who study languages at a professional level tend to perceive non-standard features as more acceptable and rate them higher. By contrast, participants who major in subjects other than language and linguistics react in the opposite way. In lower proficiency (school English?), the ratings tend to be higher (vague recollection of the rules learned at school, uncertainty about correctness?), but with the improvement of the language skills, non-standard features are perceived as less tolerable.

This finding is partly in line with the results of Birdsong and Kassen's (1988) study on error gravity judgments: less proficient learners assessed the errors as not serious, more proficient learners found them more severe, and

non-native speaking teachers were stricter than the two learner groups. In the present study of non-standard features, language professionals (also teachers?) at the highest level of proficiency do not have a negative concept of 'wrong' where they would view any deviation from the standard norm as an error, but they are more broad-minded and ready to accept features of other English varieties.

6.2 Learner errors

In this section, processing costs are compared to acceptability ratings for learner errors. The evaluation data were not collected by myself but borrowed from another study involving only native speakers of English. Hultfors (1986) investigated the perception of typical learner errors and produced a hierarchy of error types based on the evaluation of 70 erroneous sentences by 444 native speakers of British English. The task was to rate each sentence based on two 5-point scales. The first category was from "native-like" to "very foreign", and the results were discussed as the index of "foreignness". The second scale was from "very easy to understand" to "very difficult to understand", and the results indicated the "intelligibility difficulty" of the sentence. Based on the sum of mean evaluations in the two categories, each test sentence was assigned an error gravity index (op. cit.: 218-220).

I used 60 sentences from Hultfors (1986) in the eye-tracking experiment and compared his indices to the native speakers' processing cost (Appendix 8). First of all, I checked for the correlation of foreignness and intelligibility difficulty elicited in the judgment test. The pairwise Pearson correlation is r = 0.76, which indicates a very strong positive relationship. Below is the list of sentences with the largest "combined error gravity index" (out of the 60 items used in the eye-tracking experiment).

Rank	Sentence	Foreignness	Intelligibility difficulty
1	*He **thinks not** that they know what to do.*	*4.53*	*3.51*
2	*She was standing alone, **beside her** with rage.*	*4.19*	*3.03*
3	*She was here **for two years ago**.*	*4.09*	*2.93*
4	*I was **operated last** week.*	*4.05*	*2.88*
5	*Excuse me, what **is the clock**, please?*	*4.5*	*2.42*
6	*This room **smells food**, doesn't it?*	*4.26*	*2.61*
7	*He has **an own** company.*	*4.11*	*2.65*
8	*Many cities have **stopped to expand**.*	*3.83*	*2.86*
9	*He works very **hardly** and he likes his job.*	*4.07*	*2.47*
10	***Became he** a dentist?*	*4.33*	*2.2*

Table 6.17. *Learner errors rated as most severe in Hultfors (1986)*

The mean value of foreignness for the 60 sentences is 3.5 (standard deviation 0.6), and the mean for intelligibility difficulty is considerably lower – 1.9 (standard deviation 0.5). It is Hultfors' operationalization of error gravity that it consists of poor intelligibility and a foreignness impression. It remains unclear whether the same sentences would be assigned the highest gravity rates if the task was different, i.e. if the informants were asked to evaluate the grammaticality or well-formedness. In this list, compared to the non-standard features discussed in the previous section, the items consist of existing words combined in a way that obviously makes comprehension difficult. Rank 3 almost has a "garden path" effect where the reader first follows one interpretation ("for a period of two years") before bumping into the word *ago* that makes the initial interpretation impossible. Unlike such cases as *he don't*, *you was* or *he weren't*, in which the repair presupposes the selection of a competing form within the same paradigm, the learner errors in Table 6.17 are much more complex. Some sentences need syntactic reordering (Ranks 1 and 10) or inserting a preposition (Ranks 4 and 6), others require some kind of deletion (Ranks 3 and 9) or addition (Rank 2). It does not seem to be possible to find one grammatical criterion determining the high gravity rates for these errors.

The following sentences received the lowest gravity rates.

Rank	Sentence	Foreignness	Intelligibility difficulty
51	*There is one chair too **much**…*	*3.03*	*1.48*
52	*I look forward to **hear** from you.*	*2.89*	*1.38*
53	*He is **an useful** member of the team.*	*2.79*	*1.44*
54	*I **didn't buy nothing** at the supermarket.*	*2.56*	*1.63*
55	*He speaks French quite **good**…*	*2.53*	*1.26*
56	*He told me **to not worry**.*	*2.29*	*1.35*
57	*I am staying **at Sheraton** hotel for three days.*	*2.13*	*1.23*
58	*He has a blue car and I have a **red**. And you?*	*2.03*	*1.31*
59	*We have a great **deal** of problems.*	*2.02*	*1.28*
60	*It took me **a hour** to get there.*	*2.1*	*1.19*

Table 6.18. *Errors rated as least severe in Hultfors (1986)*

The result that the items above do not involve comprehension difficulty is sensible. Although Hultfors (1986) studied the perception of errors typical of Swedish learners of English, many of the sentences in Table 6.18 were not categorized as only EFL-errors in the analysis in section 5.5.1, where the effect of the typical error producer was discussed. Among the 'least foreign' and 'most intelligible' errors one can find non-standard uses typical of native speakers (e.g. the double negation in *I didn't buy nothing* and using an adjective for an adverb in *quite good*). Besides, the wrong form of the indefinite article (*a hour,*

an useful) may as well be taken for a typographical mistake. As previously mentioned in Chapter 2, the intended error in *I am staying at Sheraton hotel* was the omission of the definite article (correct: *at the Sheraton hotel*); however, when participants were asked to give their interpretation of the correct sentence (Hultfors 1987), many of them wrote *I was staying…*, *I have been staying…* without paying any attention to the article. A failure to produce agreement of the quantifier with a countable noun (*a great deal of problems, one chair too much*) was not rated as a severe error either.

How much do high gravity ratings for learner errors depend on the processing cost? It appears that the foreignness impression and the intelligibility difficulty only moderately correlate with the extra processing time caused by the error (cost : foreignness, r = 0.33; cost : intelligibility, r = 0.36). For example, the absolute 'winner' of both categories in Hultfors' sample is the sentence *He thinks not that they know what to do* (foreignness = 4.53, intelligibility difficulty = 3.5). However, in my reading experiment, the critical region *thinks not* emerges as low as Rank 35.

The correlation of Hultfors' error gravity index and the processing cost is also moderate, r = 0.37. For instance, *beside her with rage* (Rank 2 in Table 6.17) showed the greatest processing cost in the list of 60 sentences adopted from Hultfors' study and was rated second on the error gravity scale across the same sentences. However, the next most time-consuming error *When I five years ago visited London…* (cost Rank 2) appears only at the 33rd position in the error gravity table of the same 60 sentences. This demonstrates that error detection and conscious evaluation do not fully correspond to the processing cost in reading for comprehension.

Such a moderate correlation of 37% implies that there is a number of sentences for which the interdependence of the evaluation and processing cost does not work. Below are the sentences with the largest positive rank difference, which means that the items were processed relatively fast but were assigned high gravity rates.

	Sentence	Rank difference
1	*He **thinks not** that they know what to do.*	*34*
2	*He works very **hardly** and he likes his job.*	*30*
3	***Came he** to the party yesterday?*	*27*
4	*She had her radio beside **herself** and it played her favorite song.*	*26*
5	*Many cities have stopped **to expand**.*	*23*
6	*Yesterday's accident **depended** on the bad weather.*	*21*
7	*This room **smells food**, doesn't it?*	*19*
8	*He drives **badlier** than his brother.*	*16*
9	*The car was badly **injured** in the accident.*	*16*
10	*I always tried to **make** my best at school.*	*14*

Table 6.19. *Items with largest positive rank differences (cost* minus *evaluation)*

The following errors, by contrast, were processed very slowly, but the sentences were rated closer to "native-like" and "easy to understand".

	Sentence	Rank difference
1	*This is my **brother's-in-law** cap.*	*-40*
2	*He is **an useful** member of the team.*	*-39*
3	*I look forward to **hear** from you.*	*-35*
4	***When I five years ago visited London**, I could hardly speak English.*	*-31*
5	*He told me **to not worry**.*	*-30*
6	*Who **learnt** you Spanish?*	*-30*
7	*He has a blue car and I have a **red**. And you?*	*-29*
8	*There is one chair too **much**. Shall I put it away?*	*-28*
9	*You **do** good coffee!*	*-28*
10	*They **fighted** bravely in the war.*	*-25*

Table 6.20. *Items with largest negative rank differences (cost* minus *evaluation)*

It is surprising to find the same error types on both extremes of the rank differences, namely wrong lexical choices: *accident depended on the bad weather, the car was badly injured, to make by best* in Table 6.19 and *who learned you Spanish, you do good coffee* in Table 6.20. Why are some lexical errors better than others? Moreover, one missing *-ing* form in *Many cities have stopped to expand* has a positive rank difference of 23, while another missing *-ing* form in *I look forward to hear from you* has a negative rank difference of -35.

At the same time, there are sentences for which the evaluation very well corresponds to the processing cost. The following items can reasonably be claimed as causing processing difficulty, because they received high ranks in both time cost and severity judgment.

	Sentence	Cost rank	Evaluation rank	Rank difference
1	*She was standing alone, **beside her** with rage.*	*1*	*2*	*-1*
2	*Excuse me, what is **the clock**, please?*	*4*	*5*	*-1*
3	*He has **an own** company.*	*7*	*7*	*0*
4	*She was here **for two years ago**.*	*8*	*3*	*5*
5	*I was **operated last** week.*	*10*	*4*	*6*

Table 6.21. *Items with high cost and evaluation ranks*

It is interesting to notice that the 'worst' errors do not presuppose any massive syntactic reordering or any bizarre novel forms. The necessary repair is actually very small (adding *–self* to *beside her* or the preposition *on* to the verb *operated*; deleting an only 3-letter word *for* or *ago*; replacing the indefinite

article *an* with the possessive pronoun *his*), but these little 'defects' make the meaning unclear.

In contrast, the sentences in the table below are easy for perception in all senses, because they did not complicate processing and were assigned low error gravity rates.

	Sentence	Cost rank	Evaluation rank	Rank difference
1	*I **didn't buy nothing** at the supermarket.*	*48*	*54*	*-6*
2	*The house was full of **mouses**, they were all over the place.*	*42*	*46*	*-4*
3	*She is married **with** a German.*	*41*	*41*	*0*
4	*I **haven't heard something** from him for a long time.*	*50*	*47*	*3*
5	*My brother **have** worked for this company for three years now.*	*50*	*44*	*6*

Table 6.22. *Items with low cost and evaluation ranks*

Although the vast majority of the errors included in this sample can indeed be qualified as learner errors, the items that proved to be easiest for the conscious and unconscious perception are also typical of native speakers. They include double negation, subject verb/auxiliary agreement and extension of plural formation rule to irregular nouns, which can be found in children learning English as L1.

6.4 Summary

The evaluation data for non-standard features are consistent with WAVE in that items rated as most acceptable by native speakers have high attestation rates and pervasiveness across many varieties of English. The most negatively rated features involve violations of grammatical agreement between two adjacent words. Speakers of the standard L1 varieties are united in their assessments.

Based on the computed correlations, it can be argued that error evaluation is only partly dependent on the processing cost (0.36 in L1 and 0.41 in L2). It is likely that unacceptability and error gravity elicited in judgement tests are socially determined. This is confirmed by the fact that data for native speakers and foreign learners have a twice as higher correlation in the off-line survey (0.85) than in the on-line reading experiment (0.46). Explicit knowledge of the language is more similar for L1 and L2 than implicit knowledge.

Thus, assessments only partly reflect how much errors complicate processing. No evidence could be provided that the evaluations and the greater processing cost coincide even in cases of declared intelligibility difficulty (0.36 for learner errors). There are good reasons to observe a clear distinction between "error gravity" as a result of evaluation studies and the "processing cost" based on real-time measures.

It is suggested to use the term "processing difficulty" with caution because of its somewhat negative connotation. It is widely used in traditional psycholinguistics (with emotionally deprived materials) as a cover term for any latency in the reading times. However, "difficulty" is not unequivocally appropriate in relation to items that are looked at longer but are not rated as poorly comprehensible or, generally, as severe errors. Humans may be attracted to deviant words for different reasons, including positive ones. Just like users of the eye-tracking technology for marketing purposes interpret longer processing times for advertisements and packages as due to "attracting attention", psycholinguistics also needs to refine its approaches to what it calls "difficult". For example, if slowdowns in the eye-tracking data were synchronized with ERP signals, not only of lexical and syntactic processing but also indexing item evaluation and processing conflicting information (various kinds of 'surprisal'), it would probably account for a number of different reasons why a word in its particular form is looked at longer.

7. Conclusion

Detailed studies of learner errors started as a response to behaviorist assumptions that learners habitually transfer structures from their native language and that, for pedagogical purposes, errors can be predicted by contrasting the native and the target languages. In the 1960s and 1970s, Error Analysis tested the Contrastive Hypothesis by examining massive learner data. It turned out that the effect of transfer in learner errors was much smaller than the share of developmental errors common for students with different L1 backgrounds and even for first language acquisition. It implied that learners producing imperfect utterances are not just faulty imitators, but they actively work out the language, testing hypotheses about its structure and creating transitional systems. Consequently, the research field made a shift from the behaviorist view on language learning to reliance on the general cognitive processes. The idea that an error is a deficit of the learner's system was replaced by the acknowledgment that errors are defined as such with regard to the target variety, but they are fully justified within the interlanguage.

A new generation of error studies appeared with the advent of computer-based techniques and learner corpora. However, modern cognitive approaches to language analyze breaches of the code not in terms of learner errors but separately from their producers and the social background, i.e. as material – to lend insight into how language functions or to challenge cognition and elicit processing strategies for a particular kind of anomaly. Parallel to psycholinguists, whose major goal is to contribute to the theory of language processing, there are researchers whose recent studies analyze language errors for application in foreign language instruction and the teaching of writing skills to native speakers.

Some studies of errors have addressed error evaluation. They investigated error gravity depending on whether the perceiver is a native or a non-native speaker, a teacher or a non-teacher, the assumed characteristics of the writer or other variables. The present thesis has similar goals to those of error evaluation studies of the past, but instead of the paper-and-pencil methods of collecting the data, it implemented modern eye-tracking techniques. These allowed to investigate error perception DURING reading and to analyze the processing times for different error types by means of state-of-the-art statistical methods.

The results of my reading experiment support the conclusion drawn from error gravity studies that "an error is an error" (Van, Meyer and Lorenz 1984),

but not that "all errors are equally irritating" (Albrechtsen, Henriksen and Færch 1980). In other words, the distinction between an error and a non-error is rather dichotomous. Wrong stimuli are processed significantly longer, on aggregate, than their corrected versions (Fig.5.1, p.95). This relates primarily to grammatical surface-level errors (Fig.5.25a, p.151); e.g. tense/aspect inconsistencies do not have a significant effect (Fig.4.1, p.72; Fig. 5.25b, p.151). The processing cost across errors is gradient (Fig.5.3, p.97).

It would be infelicitous and circular to define an error simply as a deviation from the norm causing a significant delay in the reading time, just as it would be wrong to argue that breaches of the code which do not complicate processing should not be called errors. The original impetus driving my investigation was to determine the actual processing difficulty of 'officially recognized' errors. Although it has been shown that some errors may be processed as easily as their corrected versions, or even faster than the grammatical items (Tables 5.4 and 5.5, p.99), it does not affect the definition of error as a violation of the language system from a prescriptive point of view.

The presence of errors in the text affects the reading behaviour of both native and non-native speakers. In general, errors trigger regressions and slowdown the processing on the next few words (the spillover effect in Fig.5.1b, p.95). As a reaction to noisy conditions, participants made significantly more regressive movements at the beginning of the reading session (Fig.5.5, p.103), and these early regressive paths were also of longer duration (Fig.5.6, p.104). Towards the end of the experiment (that lasted approximately 45-55 minutes), participants made fewer and shorter regressions, but the error processing cost significantly increased in the first pass (Fig.5.4, p.102). This effect of adaptation to erroneous stimuli and to the task indicates that readers adjust themselves to repairing errors immediately after they have seen them and become reluctant to reinspect the previous parts of the sentence. Such an effect could not be produced by a particular sequence of error types because the order of the items was randomized into 48 different lists.

Divergent results of error gravity studies have shown that there can be no universal hierarchy of error types because the order of the categories changes from context to context. If one produces a hierarchy, it has to be specified for whether the perceiver is a careful and attentive reader or not, a native or a non-native speaker (at what proficiency level), a teacher or a non-teacher, whether or not s/he has a lot of experience with non-native speech, and how much the reader knows about the writer. Besides, there need to be details on the genre of the text, the error rate, the frequency of the items affected by the error, the probabilistic information on the competing interpretations, and whether this particular error is stigmatized by the

educational system. Multiplying these factors by each other, one can receive the number of possible hierarchies of errors. It is more efficient to find out the effect of each factor and apply it to the particular instance of error in order to predict its processing and evaluation.

The present eye-tracking experiment has demonstrated that the higher frequency of the item involved in an error increases the processing cost. The mismatch between a verb form in the Simple Past or the Present Perfect and the temporal adverbial, generally, does not require more processing time than the matching condition (Fig.4.1, p.72).

(7.1) *a. *I **found** good books **lately** and I can recommend you some.*
*b. *I'**ve found** good books **last year** and I can recommend you some.*
*c. I **found** good books **last year** and I can recommend you some.*
*d. I'**ve found** good books **lately** and I can recommend you some.*

However, the latency becomes significant only for mismatching adverbials used with frequent verb forms (Fig.4.9, p.84), such as the verbs *be, say, get, come, go, make* in the Simple Past and *be* in the Present Perfect. The explanation may sound somewhat behaviorist: this result implies that readers develop a 'habit' for a particular use of high-frequency items and are especially sensitive to the deviations regarding them. In cognitive terms, this could be explained as an effect of entrenchment. High-frequency forms, although used in different contexts, form distinct patterns, which could be predicted probabilistically. For low-frequency items, by contrast, the 'sample size' is too small to form distinct patterns. As a consequence, they remain more open to non-standard uses.

It is well-known that reading in the native language is faster than reading in the foreign language. The present study provides some details. Learners are significantly slower than native speakers in both correct and wrong sentences (p.95). Native speakers' high reading speed is supported by the fact that they do more skipping in correct sentences, but they significantly reduce skipping when there is an error (Fig.5.8, p.106). Almost no skipping occurs while reading in L2 in both correct and wrong sentences. The slower reading speed of L2 learners can be explained by the congestion of the working memory. It is caused by the learners' need to process more details, whereas native speakers' prior knowledge allows them to save resources.

Non-native speakers' processing is not only slower in speed but also shorter in span and shallower in depth. The shallow structure hypothesis (Clahsen and Felser 2006a) is based on a series of psycho- and neurolinguistic studies indicating that L2 learners show similarity to L1 speakers in the processing of morphology (e.g. incorrectly formed participles and noun plurals). However, there are significant differences between L1 and L2 processing for

large-scale items, such as relative clause attachments or syntactic (filler-gap) dependencies. Besides, according to this hypothesis, non-native speakers tend to rely more on semantic and pragmatic information (e.g. thematic roles for predicate-argument structures), while native speakers also compute detailed hierarchical syntactic representations.

The results of my eye-tracking study support the shallow structure hypothesis. Concerning the span, native speakers are shown to be sensitive to errors at all distances between the mismatching parts, up to a distance of six words (Fig. 5.17, p.121). However, non-native speakers were sensitive only to errors at a distance of no more than three words. There are also differences regarding the depth, i.e. complexity level of the construction affected by the error. In earlier stages of sentence processing, native speakers focus on complex large-scale syntactic errors more than they do on small morphological ones (Fig.5.26, p.153). Non-native speakers, by contrast, react more strongly to incorrect morphemes and, especially, to suppletive forms, paying slightly less attention to syntax. In the late time measure, it becomes clear that native speakers return to the simple (local) errors, so that, in sum, all violations of grammar are processed equally long (Fig.5.27, p.155). However, for non-native speakers, attention to these small-scale surface errors still dominates.

The L1/L2 differences can be illustrated by the perception of the sentence *The man I met's girlfriend is a real beauty*, analyzed against the control *The girlfriend of the man I met is a real beauty.* For the non-native speakers, the processing of this item cost 665 milliseconds (Rank 2 out of 282, i.e. second most costly), while the native speakers handled it in 124 ms (Rank 121). (Consider that an average difference in the reading speed between L1 and L2 for the same items was about 200 ms.) If a syntactic representation for this sentence is computed in L1, it becomes clear that the genitive *'s* is attached to an NP that is extended to include a relative clause. It sounds unusual, but the relations between the words are nonetheless clear. However, an L2 learner, guided by the relations between adjacent words, arrives at *met's* and is significantly delayed analyzing it. The native and non-native speakers also differed in their evaluation of this sentence. With the rating 2.64 (max. 5.0), it received Rank 28 (out of 50) in L1, whereas the rating in L2 was considerably lower – 1.57 (Rank 6). For the native speakers, this feature belongs to the group of items which are processed longer than other errors but are rated as more acceptable than others.

The differences between L1 and L2 in their perception can be explained by the fact that native speakers incline towards top-down processing, while non-native speakers tend to bottom-up decoding. The same tendency was revealed in error evaluation studies. Native speakers pay more attention to the

overall clarity, content and organization of the text, while non-native speakers, especially teachers, concentrate on the rule infringement. I believe, teachers are driven by the attitude that the rules for word agreement were explained to the students in the first lessons, and teachers rate such violations stricter because students should have known the rules. Another explanation is rooted in the cognitive constraints for an L2, making non-native speakers unable to look broadly. There are more resources involved in processing a foreign language, more effort directed to the control of the cues between L1 and L2, while the working memory capacity is limited. In reading and in error processing, native speakers rely on their complete knowledge of the language, while non-native speakers are still collecting it. Native speakers' probabilistic knowledge allows them to save resources by doing successful 'guesswork', whereas non-native speakers' insufficient competence makes them resort to the scrupulous decoding.

There is a debate whether or not the differences between L1 and L2 are 'fundamental' in the sense that a foreign learner will never reach a native speaker's level of competence in the language. In the present experiment, the significant effect of language proficiency on processing errors indicates that more advanced learners become faster not only in the general reading speed but also in recovery from errors (Fig.5.10, p.109). This is in line with previous observations that L2 learners definitely approach native speakers in their performance.

No significant differences have been found between the participants from the UK and the US in the processing of mismatching temporal adverbials (Fig.4.12, p.90), or the mix of learner errors and non-standard features (Fig.5.9, p.107), or only non-standard features (Fig.5.24, p.138). Moreover, in the evaluation of non-standard features, the correlation between the UK and the US ratings was 92% (p.170). Nonetheless, the American participants appeared to be more sensitive to errors, as the increase in their reading times for the mix of errors is significant, unlike that for the British participants. The reading times of the Canadian participants (only 3 subjects) often differed from those of the speakers from Britain and America, but in the evaluation, all three L1 varieties clustered together.

Another important conclusion to be drawn from this study is that error processing is slowed down by the incongruence (a kind of surprisal), when the meaning is clear and the syntactic structure remains intact. This is relevant for the processing of words with a homophone effect (Fig.5.18, p.125), e.g. 'eye dialect' with their orthographic incongruence, or errors involving transfer from a familiar language (Fig.5.23, p. 137). For example, in the case of non-standard features, it is likely that native speakers are familiar with most of them, and

their meaning is so transparent even for foreign learners that no processing difficulty would be expected. Yet, L1, ESL features trigger significantly longer reading times than their neutral counterparts for both native and non-native speakers (Fig.5.22, p.136). For the latter, the relevant reading times are significantly longer than for other errors. This finding may lend insight to the dilemma discussed by researchers on error evaluation: whether or not irritation is closely dependent on comprehensibility. In fact, this may often be the case, but non-standard features are fully comprehensible and still cause 'irritation' in the sense of "preoccupation with form", as they attract the gaze and require additional processing efforts for the item evaluation and processing the conflicting information (cf. Newman and Connolly 2004, discussed on p.126). I believe that the participants were looking longer because they were surprised to see non-standard uses in writing and in the context of the experiment.

Finally, as the major goal of the project, I examined the extent to which error processing cost based on an on-line reading experiment corresponds to the results of off-line judgment tests for the same sentences. It appeared that acceptability judgments correlate with the processing cost at less than 40%. The correlation values are almost the same for non-standard features (0.36 in L1) and learner errors (0.37 in L1), which implies that this could be a general tendency. Such a moderate correlation value explains why, for example, mismatching temporal adverbials, being judged as non-acceptable and attested as low-frequent in the corpus, do not impose any processing difficulty.

Table 7.1 summarizes the correlations between the processing cost and evaluations for native and non-native speakers with regard to the perception of non-standard features (for learner errors, there was no evaluation by non-native speakers available).

Cost: Evaluation		NS: NNS	
NS	NNS	cost	evaluation
0.36	0.41	0.46	0.85

Table 7.1. *The correlations of the perceptions depending on the task and L1*

It can be argued that the relationship between the processing cost and the evaluation is not dependent on whether reading occurs in L1 or L2. This, again, indicates the existence of a general tendency for the judgment/processing correlation of around 40%. Native and non-native speakers have a very similar understanding of which non-standard features are closer to standard English, but L1 and L2 speakers differ to a greater extent in the processing of the features. The processing cost was calculated based on the regression path

duration (RPD), and it has been shown that, in this time measure, native speakers focus more on large-scale errors in complex constructions, while non-native speakers are mostly preoccupied with small-scale errors in simpler constructions (Fig.5.26, p.153). In addition, learners look longer at non-standard features, whereas native speakers spend more time on learner errors (Fig.5.22, p.136). This divergence is likely to determine the moderate correlation of 0.46 for non-standard features (50 items) and 0.48 for the whole sample (281 items). However, when the processing cost is based on the total reading time, the relationship does not become stronger ($r = 0.44$ for the whole sample). This late time measure also reflects the processes of evaluation in which both speaker groups have much more agreement in an off-line task (0.85 for non-standard features). Yet, at the end of processing the sentence, the differences between L1 and L2 are not leveled out.

My initial hypothesis that longer processing times positively correlate with stricter evaluations was not supported by the data. The discrepancy between the processing cost and the evaluation may be due to the fact that some grammatical violations do not necessarily increase the reading times, and that participants tend to look longer at other deviations from the norm which they do not necessarily consider wrong. Thus the main reasons why errors bother us are twofold: one, because they challenge recognition and comprehensibility and, two, because we KNOW they are wrong. These factors overlap to a different degree for each error.

The advantage of the eye-tracking method is, of course, that it measures on-line (real-time, moment-to-moment) processing during uninterrupted, almost natural, reading (or other perception tasks). However, simple time measures do not answer all questions. This method is limited in the sense that we do not know what kinds of processes underlie longer fixations. We can observe fixations of different duration, but we never know WHY participants look longer at particular stimuli and what they experience at that moment. This problem is known to the users of eye-tracking for marketing purposes. To enhance the explanatory power of eye movement studies, eye trackers have been synchronized with the biometric devices for measuring the brain activity (EEG) and the activity of the heart (ECG), pupil size, sweat production of the skin, oxygen saturation and oxygen consumption of near distant tissues, respiration, body temperature, body position and movement, blood pressure and blood flow. If such synchronization could be done for psycholinguistics, it would help us interpret the latencies of reactions more accurately and provide essential information on processing language.

Zusammenfassung in deutscher Sprache

Fehlerwahrnehmung in Mutter- und Fremdsprache:
Bewertung vs. Verarbeitungsaufwand

Die Studie erforscht die Wahrnehmung von Lernerfehlern und umgangssprachlichen Merkmalen durch ein psycholinguistisches Blickbewegungsexperiment. Die Haupthypothese ist die folgende: je länger man den Fehler verarbeiten muss, desto gravierender wird er bewertet.

Seit den 1960er Jahren wird die Reaktion auf Lernersprache innerhalb der Fehleranalyse untersucht. Man hat herausgefunden, dass Englischmuttersprachler die Verständlichkeit und Textorganisation für am wichtigsten halten, während solche Nichtmuttersprachler, die auch Lehrer sind, vor allem die Einhaltung der grammatischen Regeln beachten. Einige Forscher haben auch Hierarchien von Fehlertypen vorgeschlagen, die als universal gelten sollten. Allerdings hat die Reihenfolge der Fehlerkategorien in den Hierarchien oft gewechselt, und es gibt bis heute keine Einigung darüber, ob zum Beispiel lexikalische oder grammatische Fehler am schwersten wiegen.

Das Ziel der vorliegenden Studie ist es, die Faktoren, die Fehlergravität bestimmen können, einzeln zu betrachten und ihre Wirkung herauszufinden. Die Daten wurden in einem Eyetracking-Experiment gesammelt, in dem 32 Englischmuttersprachler und 24 Nichtmuttersprachler aus Deutschland 330 Sätze mit Fehlern im Englischen und ihre korrigierten Kontrollsätze im Labor gelesen haben.

Im ersten Teil des Experiments wird untersucht, ob die Lesezeiten sich erhöhen entsprechend der größeren Distanz zwischen dem Verb und dem Zeitadverbial, das auf die Verbform nicht abgestimmt ist. Im Englischen können Handlungen in der Vergangenheit durch Verbformen im Simple Past (SP) oder im Present Perfect (PP) bezeichnet werden. In der Standardvarietät werden diese Zeitformen mit bestimmten Adverbialen benutzt, wie in folgenden Beispielen:

(1) *a.* SP, richtig: *I **found** good books **last year** and I can recommend you some.*
b. SP, falsch:**I **found** good books **lately** and I can recommend you some.*
c. PP, richtig: *I'**ve found** good books **lately** and I can recommend you some.*
d. PP, falsch: **I'**ve found** good books **last year** and I can recommend you some.*

Die eigene Korpusstudie bestätigt, dass der Anteil der „falschen" Adverbiale sehr gering ist (0,45% mit SP und 11,14% mit PP). In der Bewertungsstudie von Elsness (1997) haben Teilnehmer aus Großbritannien und aus den USA Sätze wie *I have seen John yesterday* und *I have seen Mary last night* sehr niedrig (d.h. als nicht akzeptabel) beurteilt. Allerdings wurden falsche

Adverbiale im Leseexperiment weder von den Englischmuttersprachlern noch den deutschen Lernern länger verarbeitet. Dabei hat sich auch der Effekt der Distanz nicht gezeigt. Die Lesezeiten waren aber signifikant höher, wenn die falschen Adverbiale mit häufigen Verbformen benutzt wurden (beispielweise *be, say, get, come, go, make* im SP und *be* im PP). Dieser Effekt bedeutet, dass je häufiger (i.S.v. Tokenfrequenzen) das vom Fehler betroffene Wort ist, desto länger wird der Fehler verarbeitet.

Im zweiten Teil des Blickbewegungsexperiments wurden 281 Sätze mit unterschiedlichen Fehlertypen gelesen. Die Ergebnisse weisen nach, dass Fehler einen signifikanten Zeitaufwand und Rücksprünge verursachen sowie die Verarbeitung von den weiteren Wörtern im Satz verlangsamen. Am Anfang des Experiments machten die Teilnehmer signifikant häufigere und längere rückwärtsgerichtete Sakkaden. Zum Ende des Lesevorgangs (ca. 50 Minuten) sind sie dagegen immer seltener und kürzer zurückgesprungen, wobei allerdings die Verarbeitungszeiten im ersten Lesedurchgang stets zugenommen haben. Offensichtlich passen die Teilnehmer ihr Leseverhalten so an, dass sie Fehler sofort nach der Begegnung reparieren und sich ungern die vorangegangenen Satzteile noch einmal anschauen.

Das Leseexperiment hat bestätigt, dass man in der Muttersprache (L1) signifikant schneller liest als in der Fremdsprache (L2). Das wird dadurch erklärt, dass L2-Lerner mehr Ressourcen einbeziehen müssen, mehr Prozesse kontrollieren, die in L1 automatisiert sind, und dadurch ihr Arbeitsgedächtnis anstrengen. Muttersprachler haben die Regionen viel mehr in korrekten Sätzen übersprungen und haben das Springen reduziert, wenn es einen Fehler gab. Nichtmuttersprachler haben immer sehr detailliert gelesen und kaum übersprungen. Die Lerner haben auch besonders stark auf umgangssprachliche Merkmale reagiert; für die Muttersprachler waren allerdings Lernerfehler schwerer zu verarbeiten. Insgesamt korrelieren die Verarbeitungszeiten der L1- und L2-Sprecher zu 48%, was heißt, dass Mutter- und Nichtmuttersprachler oft unterschiedlich lange Zeiten brauchen, um gleiche Fehler zu verarbeiten.

Die Ergebnisse unterstützen die „Hypothese der flachen Struktur" (Clahsen und Felser 2006a), nach der die Verarbeitung in der Fremdsprache eine kürzere Spanne und flachere Tiefe aufweist im Vergleich mit der Verarbeitung in der Muttersprache. Was die Spanne in der Fehlerwahrnehmung angeht, zeigt die statistische Analyse, dass Muttersprachler auf Items mit einer Distanz bis zu 6 Wörtern zwischen den Elementen des Fehlers reagiert haben. Die Lerner konnten aber nur die Fehler mit kurzen Distanzen (bis zu 3 Wörtern) erkennen.

Es hat sich in Bezug auf die Tiefe der Satzanalyse ergeben, dass die Muttersprachler sich in früheren Wahrnehmungsphasen eher auf komplexe Syntaxkonstruktionen konzentrieren, um die volle Satzstruktur analysieren zu

können. Im Gegensatz dazu beschäftigen sich Nichtmuttersprachler in früheren und auch späteren Zeitmaßen mehr mit einfachen morphologischen Konstruktionen. Die Unterschiede zwischen der L1- und L2-Verarbeitung können anhand der Wahrnehmung des folgenden Satzes (2a) illustriert werden.

(2) *a. *The man I met's girlfriend is a real beauty.*
b. The girlfriend of the man I met is a real beauty.

Der Zeitaufwand für die Verarbeitung des Items 2a (im Vergleich mit 2b) betrug für die Nichtmuttersprachler 665 Millisekunden (Rang 2 von 282, d.h. der zweitschwerste Fehler). Für die Muttersprachler betrug der Zeitaufwand dagegen nur 124 ms (Rang 121). (Im Durchschnitt waren die Nichtmuttersprachler bei korrekten Items nur um ca. 200 ms verzögert.) Wenn man die tiefe syntaktische Struktur in L1 kalkuliert, wird es klar, dass das Genitiv *'s* an die erweiterte NP angehängt ist, die einen Relativsatz einschließt. Allerdings wenn man in der L2 oberflächlich, d.h. durch das Verhältnis zwischen den angrenzenden Wörtern, geführt wird und in (2a) zu *met's* kommt, wird das Verständnis deutlich längere Zeit in Anspruch nehmen. Die Mutter- und Nichtmuttersprachler haben den Satz 2a auch unterschiedlich beurteilt. Mit der Bewertung 2,64 (max. 5,0) hat dieser Satz in L1 Rang 28 (von 50) bekommen, wohingegen die Bewertung in L2 viel niedriger war (1,57 – Rang 6).

Der dritte Teil des Experiments präsentiert eine fragebogenbasierte Studie, bei der die ausgewählten Sätze mit umgangssprachlichen Merkmalen nach Akzeptabilität bewertet wurden. Das hat ermöglicht, die Korrelation zwischen der fast unbewussten Fehlerwahrnehmung im Leseexperiment und der bewussten Beurteilung festzustellen. Sie liegt bei ca. 40% sowohl für Lernerfehler als auch für L1-Merkmale. Konkreter für umgangssprachliche Merkmale korreliert der Zeitaufwand und die Bewertung bei Muttersprachlern zu 0,36 und bei Nichtmuttersprachlern zu 0,41. Gleichzeitig korrelieren die Lesezeiten in L1 und L2 zu 0,46, aber die Bewertungen stimmen zu 0,85 überein. Das zeigt, dass Mutter- und Nichtmuttersprachler gemeinsames Wissen über die Grammatikalität teilen, während die Verarbeitung für das Verständnis oft unterschiedlich verläuft.

Diese mäßige Korrelation von ca. 40% zwischen der Verarbeitung und Bewertung spricht dafür, dass wahrscheinlich soziale Faktoren in der Fehlerwahrnehmung nicht weniger wichtig sind als die Prozessierungsschwierigkeiten. Folglich kann die Ausgangshypothese (also dass länger verarbeiteten Fehler auch als gravierender bewertet werden) nicht bestätigt werden. Dieses Ergebnis deutet darauf hin, dass die Übertretung der Grammatik der Standardvarietät für die Verarbeitung nicht unbedingt schwierig ist bzw. dass Leser auf solche Abweichungen von der Norm länger schauen, die sie nicht unbedingt als falsch bewerten.

Bibliography

Abramovici, Shimon. 1983. "Errors in proofreading: Evidence for syntactic control of letter processing?" *Memory and Cognition* 11(3): 258-261.

Abutalebi, Jubin. 2011. "Bilingualism, neurobiology of." Hogan, P.C. (ed), *The Cambridge encyclopedia of the language sciences*. New York: Cambridge University Press: 123-125.

Albrechtsen, Dorte, Birgit Henriksen and Claus Færch. 1980. "Native speaker reactions to learners' spoken interlanguage." *Language Learning* 30 (2): 365-396.

Baayen, Rolf Harald. 2008. *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R.H., Davidson, D.J. and Bates, D.M. 2008. "Mixed-effects modeling with crossed random effects for subjects and items." *Journal of Memory and Language* 59: 390-412.

Baddeley, Alan D. 2007. *Working memory, thought and action*. Oxford: Oxford University Press.

Baker, Beverly Anne. 2012. "Individual differences in rater decision-making style: An exploratory mixed-methods study." *Language Assessment Quarterly* 9(3): 225-248.

Bansal, R.K. 1969. *The intelligibility of Indian English*. Hyderabad: Central Institute of English.

Barber, Charles L. 1964. *Linguistic change in present-day English*. London: Oliver and Boyd.

Bates, Douglas, Martin Maechler, Ben Bolker and Steven Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7, URL: http://Cran.R-project.org/package=lme4

Bates, Elizabeth, Sandra McNew, Brian MacWhinney, Antonella Devescovi and Stan Smith. 1982. "Functional constraints on sentence processing: A cross-linguistic study." *Cognition* 11(3): 245-299.

Beason, Larry. 2001. "Ethos and error: How business people react to errors." *College Composition and Communication* 53(1): 33–64.

Beech, John R. and Kate A. Mayall. 2005. "The word shape hypothesis re-examined: Evidence for an external feature advantage in visual word recognition." *Journal of Research in Reading* 28(3): 302-319.

Bernhardt, Elizabeth B. 1991. *Reading development in a second language: Theoretical, empirical, and classroom perspectives.* Norwood, NJ: Ablex Publishing Corporation.

Biber, Douglas, Leech, Geoffrey, and Conrad, Susan. 2002. *Longman student grammar of spoken and written English*. Pearson Education Limited.

Birdsong, David and Margaret Kassen. 1988. "Teachers' and students' evaluations of foreign language errors: a meeting of minds?" *Modern Language Journal* 72: 1-12.

Blythe, Hazel I., Ascensión Pagán, and Megan Dodd. 2015. "Beyond decoding: Phonological processing during silent reading in beginning readers." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 41(4): 1244-1252.

Borsley, Robert D. 1991. *Syntactic Theory: A unified approach*. London: Edward Arnold.

Boyd, Jeremy K. and Adele E. Goldberg. 2009. "Input Effects Within a Constructionist Framework." *The Modern Language Journal* 93: 418-429.

Brandenburg, Laura C. 2015. "Testing the recognition and perception of errors in context." *Business and Professional Communication Quarterly* 78(1): 74-93.

Braun M., F. Hutzler, T.F. Münte, M. Rotte, M. Dambacher, F. Richlan and A.M. Jacobs. 2015. "The neural bases of the pseudohomophone effect: Phonological constraints on lexico-semantic access in reading." *Neuroscience* 295: 151-163.

Brooks, Teon L. and Daniela Cid de Garcia. 2015. "Evidence for morphological composition in compound words using MEG." *Frontiers in Human Neuroscience* 9: 215.

Bultena, Sybrine and Ton Dijkstra. 2013. "Lexical access in bilingual visual word recognition." In: Carol A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*. Chichester: Wiley-Blackwell (web edition).

Burt, Marina and Carol Kiparsky. 1974. "Global and local mistakes." In: John Schumann and Nancy Stenson (eds.), *New frontiers in second language learning.* Rowley, Massachusetts: Newbury House Publishing: 71-80.

Bybee, Joan. 2006. *Frequency of Use and the Organization of Language.* Oxford: Oxford University Press.

Bybee, Joan. 2010. *Language, Usage and Cognition.* Cambridge: Cambridge University Press.

Callan, V. and C. Gallois. 1983. "Ethnic stereotypes: Australian and Southern European youth." *The Journal of Social Psychology* 119: 287-298.

Carpenter, P.A. and M.A. Just. 1983. "What your eyes do while your mind is reading." In: K. Rayner (ed.), *Eye movements in reading: Perceptual and language processes.* New York: Academic Press: 275-307.

Carello, Claudia and George Lukatela. 2011. "Word recognition, visual." In: P.C. Hogan (ed.), *The Cambridge encyclopedia of the language sciences*. New York: Cambridge University Press: 923-924.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, Mass.: M.I.T. Press.

Clahsen, Harald and Claudia Felser. 2006a. "Grammatical processing in language learners." *Applied Psycholinguistics* 27(1): 3-42.

Clahsen, Harald and Claudia Felser. 2006b. "Continuity and shallow structures in language processing." *Applied Psycholinguistics* 27(1): 107-126.

Clifton, Charles Jr., Adrian Staub and Keith Rayner. 2007. "Eye movements in reading words and sentences." In R. Van Gompel, M. Fisher, W. Murray, and R. L. Hill (eds.) *Eye movement research: A window on mind and brain.* Oxford: Elsevier: 341-372.

Coltheart, Max. 2004. "Brain imaging, connectionism and cognitive neuropsychology." *Cognitive Neuropsychology* 2: 21-25.

Coltheart, Max, Kathleen Rastle, Conrad Perry, Robyn Langdon, and Johanne Ziegler. 2001. "DRC: A dual route cascaded model of visual word recognition and reading aloud." *Psychological Review* 108 (1): 204–256.

Connor-Linton, Jeff. 1995. "Crosscultural comparison of writing standards: American ESL and Japanese EFL." *World Englishes* 14(1): 99-115.

Cook, Vivian. 1993. *Linguistics and second language acquisition.* Macmillan, Basingstoke.

Cook, Vivian. Homepage. Accessed on June 5, 2015. Writing systems. Spelling mistakes, novel forms, etc. Novel spelling: eye dialect http://homepage.ntlworld.com/vivian.c/SpellingNovel/EyeDialect.htm

Corder, Stephen Pit. 1967. "The Significance of Learner's Errors." *International Review of Applied Linguistics* 5: 161-169.

Davies, Eirlys E. 1983. "Error evaluation: the importance of viewpoint." *English Language Teaching Journal* 37(4): 304-311.

Davies, Mark. 2008-. *The Corpus of Contemporary American English: 450 million words, 1990-present*. Available online at http://corpus.byu.edu/coca/

Delisle, Helga. 1982. "Native speaker judgment and the evaluation of errors in German." *Modern Language Journal* 66 (1): 39-48.

Dimroth, Christine. 2013. "Learner Varieties". In: Carol A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*. Chichester: Wiley-Blackwell (web edition).

Dulay, Heidi, Marina Burt and Stephen Krashen. 1982. *Language two*. New York: Oxford University Press.

Duyck, W., E. Van Assche, D. Drighe and R. Hartsuiker. 2007. "Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33: 663-679.

Eckes, Thomas. 2008. "Rater types in writing performance assessments: A classification approach to rater variability." *Language Testing* 25(2): 155-185.

Ehrlich, Susan F. and Keith Rayner. 1981. "Contextual effects on word perception and eye movements during reading." *Journal of Verbal Learning and Verbal Behavior* 20: 641-655.

Ellis, Nick C. 2002. "Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition." *Studies in Second Language Acquisition* 24(2): 143–188.

Ellis, Rod. 1985. *Understanding Second Language Acquisition.* Oxford: Oxford University Press.

Ellis, Rod. 1996/2008. *The study of second language acquisition.* Oxford: Oxford University Press.

Elsness, Johan. 1997. *The perfect and the preterite in contemporary and earlier English* (Topics in English Linguistics 21)*.* Berlin and New York: Mouton de Gruyter.

Elsness, Johan. 2009. "The present perfect and the preterite." In: G. Rohdenburg and J. Schlüter (eds). *One language, two grammars? Differences between British and American English* (Studies in English Language), Cambridge: Cambridge University Press. 228–245.

Field, John. 2004. *Psycholinguistics. The key concepts*. London: Routledge.

Fiorentino, Robert, Yuka Naito-Billen, Jamie Bost, and Ella Fund-Reznicek. 2014. "Electrophysiological evidence for the morpheme-based combinatoric processing of English compounds." *Cognitive Neuropsychology* 31: 123–146.

Fishman, Gideon, Arye Rattner and Gabriel Weimann. 1987. "The effect of ethnicity on crime attribution." *Criminology* 25: 507-524.

Fox, John. 2003. "Effect displays in R for generalized linear models." *Journal of Statistical Software* 8(15): 1-27.

Fraser, Catriona and Barbara F. Kelly. 2012. "Listening between the lines: Social assumptions around foreign accents." *Australian Review of Applied Linguistics* 35(1): 74-93.

Frazier, Lyn and Keith Rayner. 1982. "Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences." *Cognitive Psychology* 14(2): 178-210.

Frenck-Mestre, Cheryl and Joël Pynte. 1997. "Syntactic ambiguity resolution while reading in second and native languages." *Quarterly Journal of Experimental Psychology* 50A(1): 119-148.

Fried, Mirjam and Jan-Ola Östman (eds). 2004. *Construction grammar in a cross-language perspective*. Amsterdam: Benjamins.

Frost, Ram and Johannes C. Ziegler. 2007. "Speech and spelling interaction: the interdependence of visual and auditory word recognition". In: M. Gareth Gaskell (ed.), *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press. 107-118.

Fuertes, Jairo N., Jodi C. Potere and Karen Y. Ramirez. 2002. "Effects of speech accents on interpersonal evaluations: implications for counseling practice and research." *Cultural Diversity and Ethnic Minority Psychology* 8(4): 346-356.

Futrell, Richard and Edward Gibson. 2016. "L2 processing as noisy channel language comprehension." Bilingualism: Language and Cognition, doi: 10.1017/S1366728916001061

Garrett, Peter and Christopher Austin. 1993. "The English genitive apostrophe: judgements of errors and implications for teaching." *Language Awareness* 2: 61-75.

Gass, Susan M. and Larry Selinker. 2008. *Second language acquisition: an introductory course.* New York: Routledge.

Gass, S., R. Roots and J. Lee. 2006. "Inhibition and working memory capacity in a second language". Paper presented at the 16th European Second Language Association Conference, Antalya, Turkey, September.

Gathercole, Susan E. and Alan D. Baddeley. 1993. *Working memory and language*. Hove/Hillsdale: Erlbaum.

Gibson, Edward. 1998. "Linguistic complexity: locality of syntactic dependencies." *Cognition* 68(1): 1–76.

Gibson, Edward, Leon Bergen and Steven T. Piantadosi. 2013. "Rational integration of noisy evidence and prior semantic expectations in sentence interpretation." *Proceedings of the National Academy of Sciences U.S.A.* 110(20): 8051-8056.

Gibson, Edward, Caitlin Tan, Richard Futrell, Kyle Mahowald, Lars Konieczny, Barbara Hemforth and Evelina Fedorenko. In press. "Don't underestimate the benefits of being misunderstood." *Psychological Science.*

Goldberg, Adele E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago: Chicago University Press.

Goldberg, Adele E. 2003. "Constructions: A new theoretical approach to language." *Trends in Cognitive Sciences* 7: 219-224.

Goldberg, Adele E. 2006. *Constructions at Work: The nature of generalization at language*. Oxford: Oxford University Press.

Goldberg, Adele E. 2011. "Construction grammars." In: P.C. Hogan (ed.), *The Cambridge encyclopedia of the language sciences*. New York: Cambridge University Press. 216-219.

Gonzalez, Andrew. 1983. "When does an error become a feature of Philippine English?" In: Richard B. Noss (ed.), *Varieties of English in Southeast Asia* (RELC anthology series 11). Singapore: RELC/Singapore University Press. 150-172.

Goodman, Kenneth S. 1967. "Reading: A psycholinguistic guessing game." *Journal of the Reading Specialist* 6: 126-135.

Gries, Stefan Th., and Dagmar Divjak (eds.). 2012. *Frequency Effects in Language Learning and Processing*. Berlin/Boston: Mouton de Gruyter.

Haber, Ralph N. and Robert M. Schindler. 1981. "Errors in proofreading: Evidence of syntactic control of letter processing?" *Journal of Experimental Psychology: Human Perception and Performance* 7(3): 573-579.

Hadden, Betsy L. 1991. "Teacher and non-teacher perceptions of second language communication." *Language Learning* 41(1): 1-24.

Hammerly, Hector. 1991. *Fluency and Accuracy: toward Balance in Language Teaching and Learning.* Clevedon: Multilingual Matters.

Haspelmath, Martin. 2001. "Word Classes and Parts of Speech." In: Paul B. Baltes and Neil J. Smelser (eds.), *International Encyclopedia of the social and behavioral sciences*. Amsterdam: Pergamon. 24: 16538-16545.

Hirotani, Masako, Lyn Frazier and Keith Rayner. 2006. "Punctuation and intonation effects on sentence and clause wrap-up: Evidence from eye movements." *Journal of Memory and Language* 54(3): 425-443.

Hairston, Maxine. 1981. "Not all errors are created equal: Nonacademic readers in the professions respond to lapses in usage." *College English* 43(8): 794–806.

Hanulíková, Adriana, Van Alphen, P. M., Van Goch, M. M., and Weber, A. 2012. "When one person's mistake is another's standard usage: The effect of foreign accent on syntactic processing." *Journal of Cognitive Neuroscience* 24(4): 878-887.

Havik, Else, Leah Roberts, Roeland van Hout, Robert Schreuder and Marco Haverkort. 2009. "Processing subject-object ambiguities in the L2: A self-paced reading study with German L2 learners of Dutch." *Language Learning* 59(1): 73-112.

Hawkins, John A. 1987. "Implicational universals as predictors of language acquisition." *Linguistics* 25: 453-473.

Heller, Dieter. 1982. "Eye movements in reading". In: R. Groner and P. Fraisse (eds.), *Cognition and eye movements*. Amsterdam: North Holland. 139-154.

Hilpert, Martin. 2014. *Construction Grammar and its Application to English.* Edinburgh: Edinburgh University Press.

Huang, Laura, Marcia Frideger and Jone L. Pearce. 2014. "How non-native speakers can crack the glass ceiling." *Harvard business review* 92(6): 27-28.

Huddleston, Rodney D., Pullum, Geoffrey K. 2005. *A student's introduction to English grammar*. Cambridge: Cambridge University Press.

Hughes, Arthur and Chryssoula Lascaratou. 1982. "Competing criteria for error gravity." *ELT Journal* 36: 175-182.

Hultfors, Pär. 1986. *Reactions to non-native English, Part 1: Acceptability and Intelligibility*. Stockholm Studies in English 67. Stockholm: Almqvist & Wiksell International.

Hultfors, Pär. 1987. *Reactions to non-native English, Part 2: Foreigner Role and Interpretation*. Stockholm Studies in English 71. Stockholm: Almqvist & Wiksell International.

Hundt, Marianne and Nicholas Smith. 2009. "The present perfect in British and American English: Has there been any change, recently?" *ICAME Journal* 33: 45-63.

Hyland, Ken and Eri Anan. 2006. "Teachers' perceptions of error: The effect of first language and experience." *System* 34: 509-519.

Ilin, Natalia. 2008. *Readers' perception of errors in texts written in their L1 and L2*. Unpublished Master's thesis. University of Freiburg, Germany.

Indefrey, Peter. 2006. "A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean?" *Language Learning* 56: 279-304.

Inhoff, Albrecht Werner and Keith Rayner. 1986. "Parafoveal word processing during eye fixations in reading: Effects of word frequency." *Perception & Psychophysics* 40: 431-439.

James, Carl. 1972. "The diagnosis and analysis of errors – some insights from linguistics." *Audio-Visual Language Journal* 10: 75-79.

James, Carl. 1977. "Judgments of error gravities." *English Language Teaching Journal* 31: 116-124.

James, Carl. 1998. *Errors in Language Learning and Use: Exploring Error Analysis.* London: Longman.

Janopoulos, Michael. 1992. "University faculty tolerance of NS and NNS writing errors: A comparison." *Journal of Second Language Writing* 1(2): 109–121.

Johansson, Stig. 1978. *Studies of error gravity: native reactions to errors produced by Swedish learners of English*. Gothenburg Studies in English 44. Gothenburg: Acta Universitatis Gothoburgensis.

Johnson, Rebecca L., Manuel Perea and Keith Rayner. 2007. "Transposed-letter effects in reading: Evidence from eye movements and parafoveal preview." *Journal of Experimental Psychology: Human Perception and Performance* 33(1): 209-229.

Juffs, Alan and Michael Harrington. 1996. "Garden path sentences and error data in second language processing research." *Language Learning* 46: 286-324.

Just Marcel A. and Patricia A. Carpenter. 1980. "A theory of reading: from eye fixation to comprehension." *Psychological Review* 87: 329–354.

Just Marcel A. and Patricia A. Carpenter. 1992. "A capacity theory of comprehension: Individual differences in working memory." *Psychological Review* 99: 122-49.

Just, Marcel A., Patricia A. Carpenter and Jacqueline D. Woolley. 1982. "Paradigms and processes in reading comprehension." *Journal of Experimental Psychology* 111(2): 228-238.

Kachru, Braj B. 1985. "Standards, codification and sociolinguistic realism: The English language in the Outer Circle". In: R. Quirk and H.G. Widdowson (eds.), *English in the World: Teaching and Learning the Language and Literatures.* Cambridge: Cambridge University Press. 11-30.

Kang, Hyun-Sook and Hillary Veitch. 2016. "Mainstream teacher candidates' perspectives on ESL writing: The effects of writer identity and rater background." *TESOL Quarterly*, doi: 10.1002/tesq.289

Kaplan, Robert. 1966. "Cultural thought patterns in intercultural education". In: *Language Learning* 16: 1-20.

Katz, Leonard. 2011. "Writing and reading, neurobiology of." In: P.C. Hogan (ed.), *The Cambridge encyclopedia of the language sciences*. New York: Cambridge University Press. 932-935.

Keating, Gregory D. 2009. "Sensitivity to violations in gender agreement in native and nonnative Spanish: An eye-movement investigation." *Language Learning* 59: 503-535.

Kobayashi, Toshihiko. 1992. "Native and nonnative reactions to ESL compositions." *TESOL Quarterly* 26(1): 81-112.

Konieczny, Lars. 2000. "Locality and parsing complexity." *Journal of Psycholinguistic Research* 29(6): 627-645.

Konieczny, Lars, Barbara Hemforth, and Christoph Scheepers. 1994. "Reanalyses vs. internal repairs: nonmonotonic processes in sentence perception." In: B. Hemforth, L. Konieczny, C. Scheepers, and G. Strube (eds.), *First Analysis, Reanalysis and Repair*. IIG-Berichte 8, University of Freiburg. 1-23.

Konieczny, Lars, Barbara Hemforth, Christoph Scheepers, and Gerhard Strube. 1997. "The role of lexical heads in parsing: Evidence from German." *Language and Cognitive Processes* 12: 307–348.

Kortmann, Bernd and Kerstin Lunkenheimer (eds.) 2013. *The Electronic World Atlas of Varieties of English*. Leipzig: Max Planck Institute for Evolutionary Anthropology (Available online at http://ewave-atlas.org, Accessed on 2015-07-08).

Kortmann, Bernd and Verena Schröter. In press. "Varieties of English." In: R. Hickey (ed.), *The Cambridge Handbook of Areal Linguistics.* Cambridge: Cambridge University Press.

Krashen, Stephen D. 1982. *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon.

Kresovich, Brant M. 1988. "Error gravity: Perceptions of native-speaking and non-native speaking faculty in EFL." *Bulletin of College of Education* (University of the Ryukyus) 33(2).

Kroll, Judith F. and Cari A. Bogulski. 2013. "Organization of the second language lexicon." In: Carol A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics.* Chichester: Wiley-Blackwell (web edition).

Kroll, Judith F. and E. Stewart. 1994. "Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations." *Journal of Language and Memory* 33: 149-174.

Lado, Robert. 1957. *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press: Ann Arbor.

Larson, Kevin. 2004. *The Science of Word Recognition.* Microsoft Typography: http://www.microsoft.com/typography/ctfonts/wordrecognition.aspx

Leech, Geoffrey. 2006. *A Glossary of English Grammar*. Edinburgh: Edinburgh University Press.

Leinenger, Mallorie. 2014. "Phonological Coding During Reading." *Psychological Bulletin* 140(6): 1534-1555.

Leonard, Donald J. and Jeanette W. Gilsdorf. 1990. "Language in change: Academics' and executives perceptions of usage errors." *Journal of Business Communication* 27(1): 137–158.

Lesch, Mary F. and Alexander Pollatsek. 1993. "Automatic access of semantic information by phonological codes in visual word recognition". *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19: 285-294.

Lev-Ari, Shiri and Boaz Keysar. 2010. "Why don't we believe non-native speakers? The influence of accent on credibility." *Journal of Experimental Social Psychology* 46(6): 1093-1096.

Levy, Roger. 2008. "A noisy-channel model of rational human sentence comprehension under uncertain input." Proceedings of the 13th Conference on Empirical Methods in Natural Language Processing (Association for Computational Linguistics): 234-243.

Littlewood, William T. 1984. *Foreign and second language learning: language-acquisition research and its implications for the classroom*. Cambridge: Cambridge University Press.

Livingston, Beth A., Pauline Schilpzand and Amir Erez. 2014. "Not What You Expected to Hear. Accented Messages and Their Effect on Choice." *Journal of Management* 43(3): 804-833.

Ludwig, Jeannette. 1982. "Native-speaker judgments of second-language learners' efforts at communication: A review". *Modern Language Journal* 66: 274-283.

Lunkenheimer, Kerstin. 2012. "Typological profile: L2 varieties." In: Bernd Kortmann and Kerstin Lunkenheimer (eds.), *The Mouton World Atlas of Variation in English*. Berlin/Boston: de Gruyter. 843-872.

Luque-Agulló, Gloria and Lucas González-Fernández. 2012. "The use of newspapers for L2 reading: Practical activities." *Creative Education* 3(4): 471-478.

MacWhinney, Brian, Elizabeth Bates and Reinhold Kliegl. 1984. "Cue validity and sentence interpretation in English, German, and Italian." *Journal of Verbal Learning and Verbal Behavior* 23(2): 127-150.

Marefat, Fahimeh and Mojtaba Heydari. 2016. "Native and Iranian teachers' perceptions and evaluation of Iranian students' English essays." *Assessing Writing* 27: 24-36.

Mahoney, Sean. 2011. "Exploring gaps in teacher and student EFL error evaluation." *JALT Journal* 33(2): 107-130.

Marinis, Theodoros. 2013. "Online psycholinguistic methods in second language acquisition research". In: Carol A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*. Chichester: Wiley-Blackwell (web edition).

McClelland, James L., and James C. Johnston. 1977. "The role of familiar units in perception of words and nonwords." *Perception and Psychophysics* 22(3): 249-261.

McCretton, Elena and Nigel Rider. 1993. "Error gravity and error hierarchies." *International review of Applied Linguistics* 31: 177-88.

McNorgan, Chris, Sarah Chabal, Daniel O'Young, Sladjana Lukic, James R. Booth. 2015. "Task dependent lexicality effects support interactive models of reading: A meta-analytic neuroimaging review." *Neuropsychologia* 67: 148-158.

Mollin, Sandra. 2007. "New variety or learner English? Criteria for variety status and the case of Euro-English." *English World-Wide* 28(2): 167-185.

Monk, Andrew F. and Charles Hulme. 1983. "Errors in proofreading: Evidence for the use of word shape in word recognition." *Memory and Cognition* 11(1): 16-23.

Murphy, M. Lynne. 2011. "Words." In: P.C. Hogan (ed.), *The Cambridge encyclopedia of the language sciences*. New York: Cambridge University Press. 924-925.

Murray, Heather. 2002. "Developing language awareness and error detection: What can we expect of novice trainees?" In: H. Trappes-Lomax and G. Ferguson (eds.), *Language in language teacher education*. Amsterdam/Philadelphia: John Benjamins. 187-198.

Murray, Wayne S. and Alan Kennedy. 1988. "Spatial coding in the processing of anaphor by good and poor readers: Evidence from eye movement analyses." *Quarterly Journal of Experimental Psychology Section A* 40(4): 693-718.

Nazir, Tatjana A. 2000. "Traces of print along the visual pathway." In: A. Kennedy, R. Radach, D. Heller and J. Pynte (eds.), *Reading as a Perceptual Process*, Oxford: Elsevier. 3-22.

Newman, Randy Lynn and John F. Connolly. 2004. "Determining the role of phonology in silent reading using event-related brain potentials". *Cognitive Brain Research* 21: 94-105.

Odlin, Terence. 2003. *Language transfer: cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.

Olsson, Margareta. 1972. *Intelligibility*. The GUME Project, No.12. Gothenburg: Department of Educational Research, Gothenburg School of Education.

Olsson, Margareta. 1973. "The effects of different types of errors in the communication situation". In: J. Svartvik (ed.), *Errata*. Lund: Gleerup. 153-160.

Olsson, Margareta. 1977. *Intelligibility. An evaluation of some features of English produced by Swedish 14-Year-Olds*. Gothenburg Studies in English 40. Gothenburg: Acta Universitatis Gothoburgensis.

O'Regan, Kevin. 1979. "Eye guidance in reading: Evidence for the linguistic control hypothesis." *Perception & Psychophysics* 25: 501-509.

Paap, Kenneth R., Sandra L. Newsome, and Ronald W. Noel. 1984. "Word shape's in poor shape for the race to the lexicon." *Journal of Experimental Psychology: Human Perception and Performance* 10(3): 413-428.

Page, B. 1990. "Why do we have to get it right away?" In: B. Page (ed.), *What do you mean it's wrong?* London: Centre for Information on Language Teaching. 103-107.

Papadopoulou Despina and Harald Clahsen. 2003. "Parsing strategies in L1 and L2 sentence processing: A study of relative clause attachment in Greek." *Studies in Second Language Acquisition* 25: 501-528.

Patterson, K. E., J. C. Marshall and M. Coltheart. 1985. "General introduction". In: J.C. Marshall, M. Coltheart, and K.E. Patterson (eds.), *Surface dyslexia*. Hillsdale, NJ: Erlbaum.

Perfetti, Charles A., Laura C. Bell and Suzanne M. Delaney. 1988. "Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking." *Journal of Memory and Language* 27: 59-70.

Petti, Vincent. 1979. *Code of errors. An assessment of errors in English made by Swedish students*. Stockholm: Akademilitteratur.

Pienemann, Manfred. 1998. *Language processing and second language development. Processability Theory*. Amsterdam: John Benjamins.

Pienemann, Manfred. 2008. "A brief introduction to processability theory." In: Jörg-U. Keßler (ed.), *Processability Approaches to Second Language Development and Second Language Learning*. Newcastle: Cambridge Scholars Publishing. 9-30.

Pienemann, Manfred. 2013. "Processability Theory and Teachability." In: Carol A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*. Chichester: Wiley-Blackwell (web edition).

Plaut, David C., James L. McClelland, Mark S. Seidenberg, and Karalyn Patterson. 1996. "Understanding normal and impaired word reading: Computational principles in quasi-regular domains". *Psychological Review* 103(1): 56–115.

Politzer, Robert. 1978. "Errors of English speakers of German as perceived and evaluated by German natives." *Modern Language Journal* 62: 253-61.

Porte, Graeme. 1999. "Where to draw the red line: Error toleration of native and non-native EFL faculty." *Foreign Language Annals* 32(4): 426-434.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. New York: Longman.

Randall, Mick. 2007. *Memory, psychology and second language learning.* Amsterdam: John Benjamins Publishing Company.

Rastle, Kathleen. 2007. "Visual word recognition". In: M. Gareth Gaskell (ed.), *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press. 71-88.

Rastle Kathleen, Matthew H. Davis and Boris New. 2004. "The broth in my brother's brothel: morpho-orthographic segmentation in visual word recognition." *Psychonomic Bulletin and Review* 11(6): 1090–1098.

Rayner, Keith. 1977. "Visual attention in reading: Eye movements reflect cognitive processes." *Memory & Cognition* 4: 443-448.

Rayner, Keith. 1998. "Eye movements in reading and information processing: 20 years of research." *Psychological Bulletin* 124: 372–422.

Rayner, Keith and Susan A. Duffy. 1986. "Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity." *Memory & Cognition* 14: 191-201.

Rayner, Keith and Susan A. Duffy. 1988. "On-line comprehension processes and eye-movements during reading". In: M. Daneman, G.E. MacKinnon, T.G. Waller (eds.), *Reading research: Advances in theory and practice,* Vol. 6. London: Academic Press. 13-66.

Rayner, Keith and George W. McConkie. 1976. "What guides a reader's eye movements? *Vision Research* 16(8): 829-837.

Rayner, Keith, Sara C. Sereno, Robin K. Morris, A. Réne Schmauder, and Charles Clifton, Jr. 1989. "Eye movements and on-line language comprehension processes." *Language and Cognitive Processes* 4 (3-4): SI21-SI49.

Rayner, Keith, Sara C. Sereno and Gary E. Raney. 1996. "Eye movement control in reading: A comparison of two types of models." *Journal of Experimental Psychology: Human Perception and Performance* 22(5): 1188-1200.

Rayner, Keith and Arnold D. Well. 1996. "Effects of contextual constraint on eye movements in reading: A further examination." *Psychonomic Bulletin & Review* 3(4): 504-509.

Reicher, Gerald M. 1969. "Perceptual recognition as a function of meaningfulness of stimulus material." *Journal of Experimental Psychology* 81(2): 275-280.

Rifkin, Benjamin and Felicia Roberts. 1995. "Error Gravity: A critical review of research design". *Language Learning* 45(3): 511-537.

Roberts, Felicia and Tony Cimasco. 2008. "Evaluating ESL: Making sense of university professors' responses to second language writing." *Journal of Second Language Writing* 17(3): 125-143.

Roberts, Leah. 2013. "Sentence and discourse processing in second language comprehension." In: Carol A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics.* Chichester: Wiley-Blackwell (web edition).

Roberts, Leah, Jorge González Alonso, Christos Pliatsikas and Jason Rothman. 2016. "Evidence from neurolinguistic methodologies: Can it actually inform linguistic/language acquisition theories and translate to evidence-based applications?" *Second Language Research,* Special Issue on Neurolinguistics: 1-19.

Roberts, Leah, Marianne Gullberg and Peter Indefrey. 2008. "Online pronoun resolution in L2 discourse: L1 influence and general learner effects." *Studies in Second Language Acquisition* 30: 333-57.

Roberts, Leah and Anna Siyanova-Chanturia. 2013. "Using eye-tracking to investigate topics in L2 acquisition and L2 processing." *Studies in Second Language Acquisition* 35: 213-235.

Roberts, Leah and Sarah Ann Liszka. 2013. "Processing tense/aspect-agreement violations on-line in the second language: A self-paced reading study with French and German L2 learners of English." *Second Language Research* 29(4): 413-439.

Robinson, Peter. 1973. "Testing the second-language competence of children and adults". *English Language Teaching* 27: 190-199.

Rubin, Donald L. and Melanie Williams-James. 1997. "The impact of writer nationality on mainstream teachers' judgments of composition quality." *Journal of Second Language Writing* 6(2): 139-153.

Sagarra, Nuria. 2008. "Working memory and L2 processing of redundant grammatical forms." In: Z. Han (ed.), *Understanding second language process*. Clevedon, England: Multilingual Matters. 142-159.

Sagarra, Nuria. 2013. "Working memory in second language acquisition." In: Carol A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*. Chichester: Wiley-Blackwell (web edition).

Santos, Terry. 1987. "Markedness theory and error evaluation: An experimental study." *Applied Linguistics* 8(3): 207–218.

Santos, Terry. 1988. "Professors' reactions to the academic writing of nonnative-speaking students." *TESOL Quarterly* 22: 69-90.

Saxton, Matthew. 2011. "Children's grammatical errors." In: P.C. Hogan (ed.), *The Cambridge encyclopedia of the language sciences*. New York: Cambridge University Press. 152-154.

Schachter, Jacquelyn. 1974. "An error in error analysis". *Language Learning* 24: 205-14.

Schairer, Karen E. 1992. "Native speaker reaction to non-native speech." *The Modern Language Journal* 76(3): 309-319.

Schneider, Edgar W. 2003. "The dynamics of New Englishes: From identity construction to dialect birth." *Language* 79(2): 233-281.

Schneider, Edgar W. 2007. *Postcolonial English. Varieties around the world*. Cambridge approaches to language contact. Cambridge/New York: Cambridge University Press.

Schwenter, Scott A. 1994. "'Hot news' and the grammaticalization of perfects." *Linguistics* 32: 995–1028.

Seidenberg, Mark S. 1985. "The time course of phonological code activation in two writing systems." *Cognition* 19: 1-30.

Seidenberg, Mark S. 1995. "Visual word recognition: an overview". In: J. L. Miller and P. D. Eimas (eds). *Speech, Language and Communication*. *Handbook of Perception and Cognition*, 2nd edition. San Diego, etc.: Academic Press: 137-179.

Seidenberg, Mark S., and James L. McClelland. 1989. "A distributed, developmental model of word recognition and naming." *Psychological Review* 96: 523–568.

Selinker, Larry. 1972. "Interlanguage". *International Review of Applied Linguistics* 10: 209–241.

Service, Elisabet. 2013. "Working memory in second language acquisition." In: Carol A. Chapelle (ed.), *The Encyclopedia of Applied Linguistics*. Chichester: Wiley-Blackwell (web edition).

Shebilske, Wayne. 1975. "Reading eye movements from an information-processing point of view". In: Dominic W. Massaro (ed.), *Understanding language: an information-processing analysis of speech perception, reading, and psycholinguistics*. New York, etc.: Academic Press. 291-311.

Sheorey, Ravi. 1986. "Error perceptions of nativespeaking and non-nativespeaking teachers of ESL." *ELT Journal* 40(4): 306-312.

Shillcock, Richard. 2007. "Eye movements and visual word recognition". In: M. Gareth Gaskell (ed.), *The Oxford Handbook of Psycholinguistics*. Oxford: Oxford University Press. 89-106.

Shoebottom, Paul. Accessed on 31.07.2016. *Teachers/For mainstream teachers/Understanding mistakes in written language*. Frankfurt International School. http://esl.fis.edu

Spillner, Bernd. 1991. *Error Analysis. A comprehensive bibliography*. Amsterdam/Philadelphia: Benjamins.

Stemberger, Joseph P. 1982. "Syntactic errors in speech." *Journal of Psycholinguistic Research* 11(4): 313–345.

Stewart, Mark A., Ellen B. Ryan and Howard Giles. 1985. "Accent and social class effects on status and solidarity evaluations." *Personality and Social Psychology Bulletin* 11(1): 98-105.

Swan, Michael and Bernard Smith. 1987. *Learner English: a teacher's guide to interference and other problems*. Cambridge: Cambridge University Press.

Szmrecsanyi, Benedikt. 2012. "Typological profile: L1 varieties." In: Bernd Kortmann and Kerstin Lunkenheimer (eds.), *The Mouton World Atlas of Variation in English*. Berlin/Boston: de Gruyter. 827-842.

Tomasello, Michael. 2006. "Construction Grammar for kids." In: Doris Schoenefeld (ed.), *Constructions all over the place.* Constructions-online.de, Special Volume 1.

Tomiyana, Machiko. 1980. "Grammatical errors communication breakdown." *TESOL Quarterly* 14(1): 71-79.

Touchie, Hanna Y. 1986. "Second language learning errors: Their types, causes, and treatment." *JALT Journal* 8(1): 75-80.

Trudgill, Peter. 2009. "Vernacular universals and the sociolinguistic typology of English dialects." In: M. Filpulla, J. Klemola, and H. Paulasto (eds.), *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond.* London: Routledge. 302-329.

Trudgill, Peter. 2011. *Sociolinguistic Typology: Social Determinants of Linguistic Complexity.* Oxford/New York: Oxford University Press.

Van Hell, Janet G. and Annette M.B. De Groot. 2008. "Sentence context affects lexical decision and word translation." *Acta Psychologica* 128(3): 431-451.

Vann, Roberta, Daisy Meyer and Frederick Lorenz. 1984. "Error Gravity: A study of faculty opinion or ESL errors." *TESOL Quarterly* 18(3): 427-440.

Van Orden, Guy C. 1987. "A ROWS is a ROSE: Spelling, sound, and reading." *Memory & Cognition* 15(3): 181-198.

VanPatten, Bill and Gregory D. Keating. 2007. "Getting tense: lexical preference, L1 transfer and native and non-native processing of temporal reference." Paper presented at AAAL, Costa Mesa, CA.

VanPatten, Bill, Gregory D. Keating and Michael J. Leeser. 2012. "Missing verbal inflections as a representational problem: Evidence from self-paced reading." *Linguistic Approaches to Bilingualism* 2(2): 109-140.

Wall, Susan V. and Glynda A. Hull. 1989. "The semantics of error: What do teachers know?" In: Chris M. Anson (ed.), *Writing and Response: Theory, Practice, and Research*. Urbana: NCTE. 261–92.

Walpole, Jane Raymond. 1974. "Eye dialect in fictional dialogue." *College Composition and Communication* 25 (2): 191-196.

Wartenburger, I., H.R. Heekeren, J. Abutalebi, S.F. Cappa, A. Villringer, and D. Perani. 2003. "Early setting of grammatical processing in the bilingual brain." *Neuron* 37: 159-70.

Wickham, Hadley. 2009. *ggplot2: elegant graphics for data analysis.* New York: Springer.

Williams, Joseph M. 1981. "The Phenomenology of Error." *College Composition and Communication* 32: 152–68.

Yao, Xinyue and Peter Collins. 2012. "The present perfect in world Englishes". *World Englishes* 31(3): 386–403.

Zhang X., Y. Wang, S. Li and L. Wang. 2003. "Event-related potential N270, a negative component to identification of conflicting information following memory retrieval." *Clinical Neurophysiology* 114(12): 2461-8.

Zhang, Jie. 2016. "Same text different processing? Exploring how raters' cognitive and meta-cognitive strategies influence rating accuracy in essay scoring." *Assessing Writing* 27: 37-53.

Zola, David. 1984. "Redundancy and word perception during reading." *Perception and Psychophysics* 36(3): 277-284.

Appendices

Appendix 1. *Stimuli testing the Distance Hypothesis based on mismatching adverbials* (available electronically in PDF at: http://www.filefactory.com/file/4wyymc9zwzi9/Appendix1.pdf)

Appendix 2. *Experiment participants*

Appendix 3. *Frequencies of the verbs used in the experiment in the Simple Past and the Present Perfect*

Appendix 4. *Frequencies of the verbs used in the experiment in the Simple Past and the Present Perfect with matching and mismatching adverbials*

Appendix 5. *The error-mix sample* (available electronically as an Excel table at: http://www.filefactory.com/file/66zte6ev6l7b/Appendix5.xlsx)

Appendix 6. *Error processing cost for native vs. non-native speakers*

Appendix 7. *Acceptability judgment vs. error processing cost*

Appendix 8. *Error processing cost vs. Hultfors (1986) error gravity index*

Appendix 2. *Experiment participants*

No.	Age	Sex	Dominant hand	Lin-guist	Language as major	English teacher	Subjects studied	Origin	Native language	English assessment (max. 10)	English test (max.20)
1[28]	27	f	rechts	yes	yes	yes	Linguist, English teacher	Alabama (USA)	English		
3	27	m	rechts	yes	yes	no	Anglistik, Europäische Ethnologie	Saarland	German	8	14
4	31	m	rechts	no	no	no	Doktorand Neuste Geschichte	Schleswig-Holstein	German	8	17
5	31	m	rechts	no	no	no	nothing (IT)	Frankfurt am Main	German	8	17
6	24	f	rechts	yes	yes	yes	Geschichte, Englisch (Lehramt)	Wuppertal, NRW	German	9	17
8	19	f	rechts	yes	yes	no	Germanistik, Psychologie	Rheinland, Niederrhein	German	8	14
9	21	f	rechts	no	no	no	Molekulare Medizin	Aschaffenburg (Unterfranken)	German	8	17
11	24	f	rechts	yes	yes	yes	Lehramt: Biologie, Englisch	NRW Fröndenberg nähe Dortmund	German	8	15
12	25	f	rechts	no	no	no	Political science	Philadelphia, Pennsylvania, USA	English		
13	21	f	rechts	no	yes	no	International Studies & German	Wisconsin, USA	English		
14	22	m	links	no	no	no	Philosophie, KoWi	Norddeutschland	German	5	8
15	27	f	rechts	no	no	no	Elementary school teacher	New York, USA	English		
16	20	f	rechts	yes	yes	no	Linguistics, German, Russian	Toronto, Canada	English		
17	20	m	rechts	no	no	no	Humanmedizin	Lübeck, Schleswig-Holstein	German	6	18
18	19	f	rechts	no	no	no	Molekulare Medizin	Schwaben	German	6	14
19	20	f	rechts	yes	yes	no	German and English studies	Derbyshire, England	English		
20	20	f	rechts	yes	yes	no	German, linguistics	Iowa, USA	English		
22	21	f	rechts	no	yes	no	International Studies & German	Milwaukee, WI, USA	English		
23	20	f	rechts	yes	yes	yes	Lehramt Anglistik/Germanistik	Ba-Wü	German	6	15
24	19	f	rechts	yes	yes	yes	Lehramt Englisch/Französisch	Flensburg, Norddeutschland	German	8	14
25	21	f	links	yes	yes	yes	Linguistics, TESL, German	IA, USA	English		
26	23	f	rechts	yes	yes	yes	Lehramt Englisch, Italienisch, Biologie	Ba-Wü	German	7	16
27	21	f	rechts	no	yes	no	Deutsch, Französisch	Norfolk, England	English		
29	26	f	rechts	no	no	no	Mathematik, KW	Bayern	German	7	13
30	28	f	rechts	yes	yes	no	Geschichte, Englisch	Berlin	German	7	17

[28] Participants number 2, 7, 10, 21, 28 were removed from the analysis

No.	Age	Sex	Dominant hand	Lin-guist	Language as major	English teacher	Subjects studied	Origin	Native language	English assessment (max. 10)	English test (max.20)
31	33	f	rechts	no	no	no	Kinder-Erziehung	VS-Schweinningen (schwäbisch)	German	6	11
32	22	f	links	no	no	no	Medizin	Rheinland-Pfalz	German	7	14
33	30	f	rechts	yes	yes	no	Anglistik, Germanistik	Nordrhein-Westfalen	German	9	20
34	21	f	rechts	yes	yes	no	Deutsch, Französisch	Konstanz (Alemannisch)	German	6	16
35	20	f	rechts	yes	yes	yes	English	Canada	English		
36	21	f	rechts	yes	yes	no	Englisch, Geschichte	Schwaben	German	8	17
37	25	m	rechts	no	no	yes	English literature	Wisconsin, USA	English		
38	24	m	rechts	no	yes	no	FrancoMedia, Philosophie	Niedersachsen	German	8	17
39	22	m	rechts	no	no	no	Forestry & Environmental Studies	Canada	English		
40	21	f	rechts	no	no	no	History of Science	Montana, USA	English		
41	23	f	rechts	yes	yes	no	German & Russian, Slavistik	Buckinghamshire, England	English		
42	20	f	rechts	yes	yes	no	English & American Studies	Südbaden	German	8	17
43	20	f	rechts	no	no	no	Ethnologie, Psychologie	North Carolina, USA	English		
44	23	m	rechts	no	no	no	Physics	Colorado, USA	English		
45	24	f	rechts	yes	yes	yes	English teacher	California, USA	English		
46	27	f	rechts	no	no	no	dentistry, BuisAdm, chemistry	California, USA	English		
47	24	f	rechts	no	no	no	Economics	California, USA	English		
48	30	m	rechts	no	no	no	Civil Engineering	Alaska, USA	English		
49	20	f	rechts	no	yes	no	Psychology, German	Wisconsin, USA	English		
50	31	m	links	no	no	no	Historian (PhD)	Boston, USA	English		
51	22	f	rechts	yes	yes	yes	Lehramt Englisch, Bio	Rheinhessen	German	8	19
52	20	f	rechts	yes	yes	no	Anglistik, Amerikanistik	Heidelberg	German	8	18
53	21	f	rechts	no	yes	no	German, French	Norfolk, England	English		
54	20	f	rechts	no	yes	no	German, Law	Scotland, UK	English		
55	21	f	rechts	yes	yes	no	German, linguistics	Surrey, South England	English		
56	21	f	rechts	no	yes	no	German, French, Czech	Gloucester, England	English		
57	21	f	rechts	no	yes	no	German, French	East Sussex, England	English		
58	21	f	links	no	yes	no	German, Law	Scotland, UK	English		
59	20	f	rechts	no	yes	no	German, Russian	Glasgow, UK	English		
60	20	f	rechts	no	yes	no	German, Spanish	Edinburgh, Scotland, UK	English		
61	21	m	rechts	no	yes	no	German	Kent, UK	English		

Appendix 3. *Frequencies of the verbs used in the experiment in the Simple Past and the Present Perfect (COCA as of 2010, sorted by relative frequency)*

N	Verb	Past	Past in %	Perfect	Perfect in %	Past+Perfect	Rel. verb frequency
1	be	1440353	86%	226338	14%	1666691	32,60%
2	say	908599	98%	19073	2%	927672	18,15%
3	get	186195	66%	97170	34%	283365	5,54%
4	come	183539	88%	26007	12%	209546	4,10%
5	go	169070	88%	22870	12%	191940	3,75%
6	make	160096	84%	30472	16%	190568	3,73%
7	take	156164	89%	18620	11%	174784	3,42%
8	see	118580	73%	43799	27%	162379	3,18%
9	think	135078	98%	3249	2%	138327	2,71%
10	ask	111365	96%	4631	4%	115996	2,27%
11	look	96935	97%	2860	3%	99795	1,95%
12	feel	92135	97%	2921	3%	95056	1,86%
13	seem	79237	99%	774	1%	80011	1,57%
14	hear	50996	73%	18906	27%	69902	1,37%
15	happen	58959	85%	10504	15%	69463	1,36%
16	write	50773	89%	6489	11%	57262	1,12%
17	call	47414	90%	5010	10%	52424	1,03%
18	work	42427	81%	10014	19%	52441	1,03%
19	use	38849	85%	6789	15%	45638	0,89%
20	lose	33638	75%	11309	25%	44947	0,88%
21	win	32593	83%	6634	17%	39227	0,77%
22	fall	33409	89%	4329	11%	37738	0,74%
23	play	28514	85%	5021	15%	33535	0,66%
24	receive	26329	80%	6573	20%	32902	0,64%
25	live	26152	83%	5343	17%	31495	0,62%
26	change	16428	53%	14463	47%	30891	0,60%
27	talk	23635	81%	5436	19%	29071	0,57%
28	watch	26690	93%	1859	7%	28549	0,56%
29	buy	20602	93%	1518	7%	22120	0,43%
30	return	19151	89%	2402	11%	21553	0,42%
31	read	14232	76%	4476	24%	18708	0,37%
32	notice	14334	86%	2279	14%	16613	0,32%
33	draw	14011	87%	2050	13%	16061	0,31%
34	visit	9037	89%	1128	11%	10165	0,20%
35	sleep	6878	93%	514	7%	7392	0,14%
36	purchase	2991	84%	565	16%	3556	0,07%
37	cook	2008	90%	211	10%	2219	0,04%
38	rain	801	93%	56	7%	857	0,02%
39	communicate	577	84%	106	16%	683	0,01%
40	snow	234	94%	15	6%	249	0,01%
	Average		86%		14%		
	STDEV.S		9%		9%		
	MIN		53%		1%		
	MAX		99%		47%		

Appendix 4. *Frequencies of the verbs used in the experiment in the Simple Past and the Present Perfect with matching and mismatching adverbials*

		Simple Past				Present Perfect			
		match		mismatch		match		mismatch	
N	Verb	n	%	n	%	n	%	n	%
1	say	4733	99,96%	2	0,04%	23	47,92%	25	52,08%
2	call	256	100,00%	0	0,00%	1	50,00%	1	50,00%
3	win	572	100,00%	0	0,00%	4	57,14%	3	42,86%
4	ask	264	100,00%	0	0,00%	6	60,00%	4	40,00%
5	make	1023	99,90%	1	0,10%	37	72,55%	14	27,45%
6	get	805	99,75%	2	0,25%	11	78,57%	3	21,43%
7	receive	210	100,00%	0	0,00%	15	78,95%	4	21,05%
8	happen	1368	99,27%	10	0,73%	55	83,33%	11	16,67%
9	write	430	99,77%	1	0,23%	5	83,33%	1	16,67%
10	fall	318	100,00%	0	0,00%	22	84,62%	4	15,38%
11	talk	333	100,00%	0	0,00%	30	85,71%	5	14,29%
12	draw	54	98,18%	1	1,82%	6	85,71%	1	14,29%
13	lose	489	99,80%	1	0,20%	19	86,36%	3	13,64%
14	hear	293	98,32%	5	1,68%	85	86,73%	13	13,27%
15	take	846	99,88%	1	0,12%	29	87,88%	4	12,12%
16	come	1416	99,16%	12	0,84%	40	88,89%	5	11,11%
17	see	799	99,63%	3	0,37%	192	90,57%	20	9,43%
18	play	248	100,00%	0	0,00%	20	90,91%	2	9,09%
19	be	8874	99,74%	23	0,26%	568	91,91%	50	8,09%
20	visit	220	99,10%	2	0,90%	15	93,75%	1	6,25%
21	read	49	100,00%	0	0,00%	19	95,00%	1	5,00%
22	go	934	99,79%	2	0,21%	60	96,77%	2	3,23%
23	return	155	100,00%	0	0,00%	2	100,00%	0	0,00%
24	rain	5	100,00%	0	0,00%	1	100,00%	0	0,00%
25	work	142	100,00%	0	0,00%	12	100,00%	0	0,00%
26	watch	61	96,83%	2	3,17%	6	100,00%	0	0,00%
27	snow	0	-	0	-	3	100,00%	0	0,00%
28	sleep	13	100,00%	0	0,00%	3	100,00%	0	0,00%
29	look	120	98,36%	2	1,64%	26	100,00%	0	0,00%
30	notice	48	97,96%	1	2,04%	27	100,00%	0	0,00%
31	purchase	59	100,00%	0	0,00%	1	100,00%	0	0,00%
32	use	152	100,00%	0	0,00%	13	100,00%	0	0,00%
33	buy	121	100,00%	0	0,00%	14	100,00%	0	0,00%
34	feel	104	100,00%	0	0,00%	5	100,00%	0	0,00%
35	live	162	100,00%	0	0,00%	4	100,00%	0	0,00%
36	think	161	100,00%	0	0,00%	6	100,00%	0	0,00%
37	change	96	100,00%	0	0,00%	38	100,00%	0	0,00%
38	seem	206	97,17%	6	2,83%	4	100,00%	0	0,00%
39	cook	6	100,00%	0	0,00%	0	-	0	-
40	communicate	2	100,00%	0	0,00%	0	-	0	-
	Average		99,55%		0,45%		88,86%		11,14%
	STDEV.S		0,82%		0,82%		14,46%		14,46%
	MIN		96,83%		0,00%		47,92%		0,00%
	MAX		100,00%		3,17%		100,00%		52,08%

Appendix 6. *Error processing cost for native vs. non-native speakers*

Stimulus sentence	Native speakers		Non-native speakers	
	Rank	Cost (msec)	Rank	Cost (msec)
I heard of his appointing headmaster.	1	434	38	307
"Was that Eric Sears I hear there, or?" "This was Jim."	2	382	14	434
She was standing alone, beside her with rage.	3	380	177	103
She doesn't know what she will at all.	4	357	124	163
You know, he useta get this bad press all the time, but he was a great guy.	5	353	1	916
When I five years ago visited London, I could hardly speak English.	6	352	29	351
They fix refrigerators, washing machines, TV sets and other types of technique every day.	7	351	132	155
When you went there?	8	348	5	581
Did you tell her where your going tonight?	9	321	47	284
I am sure of that he will come.	10	320	18	410
Excuse me, what is the clock, please?	11	316	32	341
This is my brother’s-in-law cap.	12	312	7	557
The in Britain with excitement awaited festival will take place next week.	13	311	3	643
We have to live in the society after all.	14	308	277	-77
You look very much as your sister.	15	307	152	134
Who learnt you Spanish?	16	305	165	119
He has an own company.	17	295	214	51
If the boys hadn't found you, you'd be watching the radish from below, all right?	18	291	244	9
I probably wouldn't let him go just because I wants him here.	19	289	17	421
You can read today's new tomorrow morning.	20	288	34	328
She was here for two years ago.	21	287	85	218
Yesterday on the table lay my book.	22	277	44	287
Did I tell you that my mother English speaks?	23	277	15	430
The disaffected nobles haff little grasp of sophisticated conspiratorial politics.	24	266	21	402
I've always fancied meself married to a lordship and swanning up and down the stairs as lady.	25	264	60	251
My watch is broken. Can you fix her please?	26	262	58	258
No planes can fly by this weather.	27	261	69	237
They're not left school yet.	28	257	52	268
"Pass me the salt, please." "Please." "Thank you."	29	249	8	541
I came for to help you.	30	249	77	224
It's nice of all youse to have me here tonight.	31	248	27	360
But he was after taking four pitchers of draft with him, so I thought he'd not be needing more.	32	248	42	293
What do you watch? It looks interesting!	33	247	62	246
You remind me on your father.	34	247	93	196
This is the man what painted my house.	35	238	57	259
It sounds plausible, but I wouldn't lay my hand in the fire for it.	36	237	181	100
He had to pay five hundreds of pounds.	37	234	78	223
I was operated last week.	38	232	228	30

Next morning we went and shook hands at the Ford factory for to thank them for their support.	39	230	10	482
At what are you looking?	40	226	82	220
Smell on these flowers!	41	226	49	278
I'll tell you what we might should do about it.	42	222	247	2
Where do you go now?	43	221	30	347
Things have went too far.	44	220	101	188
He needs not come.	45	217	4	622
Who of you can speak English?	46	217	113	174
Instead of to fight, they started to laugh.	47	216	31	347
I still didn’t read the book.	48	214	184	96
He became finally President on third attempt.	49	212	251	-3
I've just heard one interesting news on TV.	50	211	94	195
I knew she is in town.	51	211	26	362
He is an useful member of the team.	52	210	168	115
There is hot here, or wet, or both.	53	208	91	209
Can you French?	54	207	128	160
I didn’t saw him today.	55	206	139	140
But Daddy goes to First Baptist most of the time, he be over here for Sunday dinner.	56	205	19	406
He said me that you were ill.	57	200	176	106
Herman was one of those singular individuals for who there is no adequate categorical description.	58	190	80	221
I no like the way people think that special education is an easy job.	59	188	90	210
It was little else to do.	60	188	179	101
Take them to market, please.	61	187	159	130
I look forward to hear from you.	62	187	172	109
I want to explain you this in more detail.	63	183	111	175
They fighted bravely in the war.	64	177	59	252
No people are here.	65	177	217	50
He very likes Chicago.	66	177	13	445
I took the blind by the arm and led him across the street.	67	176	74	229
Then maybe you're thinking of that crowd that was after hanging about here crooning your name.	68	176	96	193
She said their not going to the swimming pool today.	69	176	103	187
Her hairs are very beautiful.	70	175	114	172
What’s that? I can’t remember it’s name.	71	174	149	135
I was blissfully sleepy. Hearing him yak on about the amazing actress was not itching me that night.	72	173	193	85
I loved very much my father, and I still do and I always will.	73	172	86	216
Allow the salat to marinate at room temperature for 10 to 15 minutes, tossing occasionally.	74	169	98	192
We've had nice conversation with people from all over the world.	75	169	87	212
He was in a road accident killed, about two years ago.	76	169	163	120
This is the goodest cake I have ever tasted!	77	169	66	239
I promise I bring it back tomorrow.	78	169	222	45
There's a long time since I saw her.	79	168	141	139
I listen different kinds of music.	80	168	92	197
I am sure he could of done it better.	81	168	20	404

This happened yesterday on the party.	82	165	250	0
She gots eyes like a cat and sort of like Chinese people you see in the National Geographic.	83	164	81	221
I think it's wery important when we talk about these issues to be specific rather than generic.	84	163	158	130
You do good coffee!	85	162	99	192
On Tuesday have we a holiday and we'll make a small trip.	86	162	83	219
Became he a dentist?	87	162	225	41
There is one chair too much. Shall I put it away?	88	160	234	20
He a good boy, and his eagerness makes him seem younger than his 21 years.	89	157	75	225
So you be kind to him, he be a good man.	90	157	23	391
Protecting the environment is a very actual topic nowadays.	91	150	134	152
I'm ready to tell you, child, you done got your wish.	92	150	11	477
Listen to me, son, I been thinking on what to do with you and I want you to listen close.	93	149	41	297
I shouted to him to look up for the car.	94	148	64	243
You mustn’t do it, but you can if you want.	95	147	198	74
This room smells food, doesn't it?	96	147	48	282
I couldn't help but laugh when I saw this.	97	146	153	133
Well, so let's begin our lesson.	98	145	119	167
As I was pushing to make the pile I almost run him over and I saw him, so we called the police.	99	145	240	15
When she will call you, tell her I need to talk to her.	100	142	112	175
If I would have known about the party, I would have gone to it.	101	141	236	19
He told me to not worry.	102	137	63	245
Tell me how that makes you feel and what yall think we should do at this point.	103	136	24	379
Actually, I smoke not and I never tried.	104	135	110	176
I'm sorry I'm late. Have you waited long?	105	135	84	218
And quite frankly, Bill Clinton is a man trying to get he and his wife back in the White House.	106	134	56	260
They made me to do it.	107	134	36	312
The newspaper is moving away from the news business and toward backgrount stories with specific profiles.	108	133	140	140
New house is being built near the cinema.	109	132	79	222
I didn't know if she is at home.	110	131	138	140
"What time does the film start?" "To my mind at seven."	111	131	65	242
I must go to Susie to cut my hair.	112	130	67	237
I am used to get up early in the morning.	113	129	229	28
When it works, you get really high? When it doesn't work, do you get really low?	114	128	215	50
"Were you ever a smoker?" "No, I never smoked, except one misfortune when I was 8."	115	128	116	171
Here are the money I owe you.	116	128	22	402
"Jack!" I says putting my elbows on that cushion. "How are you doing?"	117	126	25	369
I don’t know what improvements that are being planned.	118	125	6	571
He has a blue car and I have a red. And you?	119	125	39	305
I really must stop to smoke, it's too bad for my lungs.	120	124	197	77
The man I met's girlfriend is a real beauty.	121	124	2	665
Oh, stop hanging noodles on my ears, I wasn't born yesterday.	122	122	203	67

He works on a sugar factory.	123	121	237	18
Anyway, me brother told me not to tell anyone about it.	124	121	209	57
If he would ask me, I wouldn’t tell him anything.	125	121	257	-17
He say he'll take care of everything. Just like always.	126	119	161	122
If you let them to do whatever they want, then I guess they are your friends.	127	117	51	271
I am born in 1945.	128	116	216	50
The most people would agree with you.	129	116	173	108
"Here's a book for you." "Thank you." "Please." "When do you need it back?"	130	116	160	124
Many cities have stopped to expand.	131	115	266	-47
He is a very alone man.	132	115	43	290
Your order is processed now, please wait.	133	115	100	190
They said they will do the work by 5 p.m.	134	114	167	116
What a dreadful weather!	135	113	104	185
He said he lives here for a long time.	136	113	213	53
It is a hard work to write a book.	137	112	278	-88
Yesterday's accident depended on the bad weather.	138	112	135	148
You should of seen Daddy and me and the way we danced at my first wedding.	139	111	16	423
What do you got in Hawaii that you don't have here?	140	111	196	78
She has very limited knowledges of German.	141	110	274	-65
Well, back in them days I was a reporter for the Daily Mirror.	142	110	33	340
I may have gave you a little too much for your first time.	143	109	68	237
He thinks not that they know what to do.	144	108	71	233
I’ve lost me bike.	145	108	109	178
He drives badlier than his brother.	146	108	188	91
My brother and me were late for school.	147	107	224	44
As for Tant, he weren't so scared as he pretended.	148	106	123	164
You must to work hard, if you want better things.	149	106	40	305
He wasn't about to buy a cat in a bag, nor was she interested in selling him one.	150	106	268	-52
I tried a few times and it brought nothing in the end.	151	106	258	-20
She wouldn't have went back in the house if she was hurt.	152	106	147	136
He said he already finished work.	153	105	50	277
It’s typical for him to leave without saying a word.	154	103	255	-16
You know nuthin about me, you don't understand me!	155	102	144	138
I realized that somebody came slowly up the stairs.	156	102	122	164
My brother have worked for this company for three years now.	157	102	73	230
The German prison system improved a lot in recent years.	158	101	130	156
I always tried to make my best at school.	159	99	259	-21
I am staying at Sheraton hotel for three days.	160	98	232	22
Arizona State is definately playing good basketball, having won four games in a row.	161	98	35	324
A lot of that was, you know, I gone on a mission trip with my church.	162	95	89	211
The river was a mile or so inland from the fishervillage called Marietta.	163	94	239	16
Happy 5th year anniversary you girl!	164	94	169	112
Came he to the party yesterday?	165	93	115	172

We have a great deal of problems.	166	93	55	260
I don’t know much people in this town.	167	93	186	93
The bus stop is right before our house.	168	92	220	46
I’m doing good in school, I'm in the highest reading group.	169	92	219	47
We realized that each of us was talking about something quite different than what the other thought.	170	91	107	180
My father said that was the only way you could get a real good look at the faces, and he was right.	171	91	137	143
He's making the battle a lot more easier for me, personally, because he's so strong.	172	91	189	91
They've been promoting theirself as the best women's club in the world.	173	91	150	134
If I'd have made the decision myself I'd have either got disqualified or got a two-shot penalty.	174	90	45	286
He works very hardly and he likes his job.	175	90	264	-39
How you like the color of these walls?	176	89	223	44
I don know whose child that is until it comes out and has a blood test.	177	88	70	235
This house is building to last at least 400 years.	178	87	245	8
This cake smells well! Did you make it?	179	85	205	64
I am convinced that he will hold his word, but he faces problems, and he just has to make a choice.	180	85	175	106
How do you call one thousand lawyers at the bottom of the ocean? A good start.	181	84	170	112
The only thing what he could do is pay the full price.	182	84	108	179
When started you to play the piano?	183	84	145	138
Where are the money I gave you yesterday?	184	83	37	310
She is married with a German.	185	83	218	49
So he's betrayed the program - the platform that he had ran against for the last 15 years.	186	83	171	111
Some of the top models in the world aren't the most pretty girls, but they have the extra something.	187	83	227	30
I am in the mortal pain, I have the toothache, you know.	188	81	202	69
I’ve known her since three years.	189	80	231	23
Hey, you ain't even real cops. What kind of cop's got three first names?	190	80	221	45
The house was full of mouses, they were all over the place.	191	80	262	-36
He speaks French quite good, he had a French nanny.	192	79	260	-32
When he did that, I would be really mad!	193	79	271	-59
I remember my first nurseryschoolteacher, she was very nice.	194	78	12	460
She is a woman I told you about.	195	77	243	9
They don't know that tommorrow may be a better day.	196	76	276	-71
The global missile is a propaganda weapon. Let the Americans break their head over what I said.	197	76	148	135
I was terrible impressed with their new song.	198	76	230	27
I have seen Mary yesterday and she looks good.	199	74	280	-115
He is meaning this book, not that one.	200	74	185	94
This time tomorrow I will lie on the beach.	201	74	238	16
I had also seen proffessional and college-level performances by several musical ensembles and drama companies.	202	73	248	2
That President has two Secretary of States.	203	73	210	57
Let me just ask you, Were you ever in a car with a pair of binoculars near their house?	204	73	28	352
The books are very expensive these days.	205	73	194	83

Stores that sell junk food, soda and bier should offer healthy options or pay a junk-food tax.	206	72	208	58
The situation turned out to be more bad than I thought.	207	72	136	145
The car was badly injured in the accident.	208	71	265	-43
You better not start all that cryin, we can go back to the house right now.	209	71	106	181
When he came I slept on the couch in the living room.	210	71	226	40
Ron said there last night that you was his ex-wife and his friend only, that he had a girlfriend.	211	69	117	167
She called during you were out on your lunch break.	212	69	156	132
If she will come, I will tell her about the accident.	213	68	154	132
I met him in the stairs.	214	68	105	183
He don't sound too good, nope.	215	68	151	134
Its going to be sunny tomorrow.	216	67	212	56
My mother is doctor, a neurologist, to be exact.	217	67	206	62
I spent the evening by John and Alice.	218	66	125	163
Ladies and gentlemen, we are shortly arriving the Freiburg Train Station.	219	66	199	73
There were less people at that concert than last time.	220	65	157	131
Borodin is one of the Russian compositors well-known around the world.	221	65	133	155
I can Spanish, because I learned it at school.	222	65	178	102
What shall we make today? Do you want to go somewhere?	223	65	211	57
It appeared that they had a pretty sound marriage, yes?	224	64	53	266
She is gone out and will come back in an hour.	225	62	279	-94
It's written with a pitchfork on the water, so don't rely much on it.	226	61	200	70
Many men emphasized that their illegal acts were comitted to allow them to engage in an honorable task.	227	61	254	-16
My daughter, what lives in London, is currently a student.	228	60	97	193
"Would you like to go there?" "No, I wouldn't." "Why?"	229	57	61	247
She must been saving these egg cartons for months.	230	57	182	97
It's just one of their many schemes to embarass me and to just drag me through mud.	231	57	162	121
Can I have a piece from that cake please?	232	56	192	88
She don’t go there now.	233	56	155	132
I brought a new book yesterday and put her on the table.	234	56	118	167
My sister is nurse and she works at a hospital.	235	55	164	119
Statistics sometimes has to compare apples with pears to make the math work.	236	54	54	261
Then Oleta's boy got hisself shot the day after Rondell and DeMarcus delivered their messages.	237	52	88	211
I didn’t buy nothing at the supermarket.	238	52	241	12
"None of it was written down." "Please?" "There's no document that says it."	239	51	187	92
It took me a hour to get there.	240	51	72	230
The girl was lost, it didn't know what was going on.	241	49	201	69
Hopefully people are liking my voice and my music more than they are thinking about my age.	242	49	76	224
He was dressed with a dark suit.	243	47	253	-14
These young leaders are decisive, self-conscious, quick-thinking, and excellent public speakers when addressing peers.	244	47	142	139
Anybody knows that there's all kinds of ways you get delayed as you get ready for flight.	245	46	183	96

When we go to the party on Saturday, let’s bring a bottle of wine.	246	46	275	-65
Hey mom, us kids want to help with Christmas too.	247	45	190	90
The most important is to tell everybody at once.	248	43	191	89
I know her since we were children.	249	43	195	79
You better come quick, I'm not sure how long it'll last.	250	42	249	2
You can’t get there without car, there's no public transaportation there.	251	42	143	138
His scenario is based on the pattern of fox remains found at sites thoroughout the island chain.	252	40	263	-39
How long are you in England?	253	39	102	187
Well, what you going to do when the one you love blows away like the clouds above?	254	39	180	100
Can I become a glass of beer, please?	255	39	246	3
I never would of thought that he’d behave like that.	256	38	9	484
Tell me please how to get to the station.	257	37	129	156
She plays the piano very good and she practices every week.	258	36	204	65
I saw by my eyes that people were throwing theirselves from the helicopter.	259	35	166	117
I'm wondering what am I going to do after Easter.	260	34	146	138
I don't have no reason to worry about it.	261	32	267	-52
The company will enjoy a wealth of free advertising from this Wendesday's press coverage.	262	30	127	160
Last month Osborne was elected to the chairman of the board of The Associated Press.	263	27	233	21
But it'll be soon dark and I won't be able to dig anything if I can't see.	264	27	120	166
Are you coming to the party? The Sarah is also going to be there.	265	24	126	161
I haven’t heard something from him for a long time.	266	22	269	-53
I mean, he's so wrong in this thing, and he's like, oh, the deficit is out of control.	267	20	131	155
I seen him a while ago, but I don't know where he got to now.	268	19	46	285
Please bring those books to the library for me.	269	14	272	-60
And it's crazy, and he don't even know how to spell my name.	270	13	121	166
She had her radio beside herself and it played her favorite song.	271	13	273	-61
My friend Betty that lives across the street has her birthday tomorrow.	272	9	174	107
Our son is 17 and he still goes to the school.	273	7	256	-17
It's hot and there's a storm a-coming. This is a little tender moment.	274	6	95	194
And it is good to be back with your friends and get catched up with school again.	275	6	252	-12
You ain't stepping foot in my house until you learn to respect me.	276	5	270	-58
Do you have brother or sister?	277	-19	235	20
Patients are wanting a lot of personal attention.	278	-20	207	59
Everybody is invited to come here at 18 o'clock today and then again tomorrow.	279	-21	261	-35
Though the book has finished, he continues to find inspiration for it in the most unlikely places.	280	-22	242	11
My eyes are running in different directions, there's so much of everything.	281	-27	282	-119
"Can you speak German?" "Yes, of course." "Great!"	282	-38	281	-118

Appendix 7. *Acceptability judgment vs. error processing cost*

Stimulus sentence (non-standard features)	Evaluation data				Eye-tracking data			
	native		non-native		native		non-native	
	rank	value	rank	value	rank	cost	rank	cost
They're not left school yet.	1	1,25	4	1,55	3	257	14	268
I'll tell you what we might should do about it.	1	1,25	12	1,74	7	222	44	2
Ron said there last night that you was his ex-wife and his friend only.	2	1,39	1	1,37	31	69	25	167
This is the man what painted my house.	2	1,39	5	1,56	5	238	16	259
And it is good to be back with your friends and get catched up with school again.	2	1,39	16	1,89	44	6	45	-12
I gone on a mission trip with my church.	3	1,41	8	1,6	24	95	21	211
My daughter, what lives in London, is currently a student.	4	1,43	3	1,44	32	60	23	193
I probably wouldn't let him go just because I wants him here.	5	1,45	10	1,7	1	289	4	421
I am in the mortal pain, I have the toothache.	6	1,46	19	2	28	81	38	69
As for Tant, he weren't so scared as he pretended.	7	1,59	18	1,96	23	106	27	164
So be kind to him, he be a good man.	8	1,63	13	1,78	11	157	5	391
Next morning we went and shook hands at the Ford factory for to thank them for their support.	9	1,68	7	1,59	6	230	2	482
They've been promoting theirself as the best women's club in the world.	9	1,68	15	1,87	25	91	31	134
Then maybe you're thinking of that crowd that was after hanging about here crooning your name.	10	1,78	10	1,7	9	176	23	193
I may have gave you a little too much for your first time.	11	1,79	10	1,7	21	109	18	237
And it's crazy, and he don't even know how to spell my name.	11	1,79	11	1,71	43	13	26	166
Then Oleta's boy got hisself shot the day after Rondell and DeMarcus delivered their messages.	12	1,81	19	2	33	52	21	211
I've always fancied meself married to a lordship and swanning up and down the stairs as lady.	13	1,82	6	1,57	20	264	10	251
Well, back in them days I was a reporter for the Daily Mirror.	13	1,82	19	2	2	110	17	340
I'm ready to tell you, child, you done got your wish.	14	1,83	9	1,63	12	150	3	477
And quite frankly, Bill Clinton is a man trying to get he and his wife back in the White House.	15	1,86	7	1,59	4	134	8	260
It's nice of all youse to have me here tonight.	15	1,86	14	1,82	15	248	15	360
"Jack!" I says putting my elbows on that cushion. "How are you doing?"	16	1,87	2	1,4	16	126	7	369
I don't have no reason to worry about it.	17	1,93	27	2,52	40	32	46	-52
He say he'll take care of everything.	18	1,99	17	1,91	19	119	33	122

Anyway, me brother told me not to tell anyone about it.	19	2,03	24	2,36	18	121	40	57
I seen him a while ago, but I don't know where he is now.	20	2,11	13	1,78	42	19	13	285
He's making the battle a lot more easier for me.	21	2,21	28	2,54	25	91	36	91
My brother and me were late for school.	21	2,21	35	2,89	22	107	42	44
Listen to me, son, I been thinking of what to do with you.	22	2,22	21	2,19	13	149	11	297
That President has two Secretary of States.	23	2,32	20	2,18	30	73	40	57
Hey, you ain't even real cops. What kind of cop's got three first names?	23	2,32	23	2,28	29	80	41	45
Well, what you going to do when the one you love blows away like the clouds above?	24	2,38	22	2,26	38	39	34	100
Anybody knows that there's all kinds of ways you get delayed as you get ready for flight.	25	2,44	29	2,56	35	46	35	96
Tell me how that makes you feel and what yall think we should do at this point.	26	2,45	31	2,64	14	136	6	379
Hey mom, us kids want to help with Christmas too.	27	2,46	30	2,57	36	45	37	90
The man I met's girlfriend is a real beauty.	28	2,64	6	1,57	17	124	1	665
It's hot and there's a storm a-coming. This is a little tender moment.	29	2,67	25	2,46	26	6	12	194
If I'd have made the decision myself I'd have either got disqualified or got a two-shot penalty.	29	2,67	33	2,79	44	90	22	286
I hope people are liking my voice and my music.	30	2,93	30	2,57	34	49	19	224
I mean, he's so wrong in this thing, and he's like, oh, the deficit is out of control.	31	3,04	36	3,06	41	20	28	155
Some of the top models in the world aren't the most pretty girls.	32	3,17	37	3,15	27	83	43	30
I'm wondering what am I going to do after Easter.	33	3,2	29	2,56	39	34	30	138
Patients are wanting a lot of personal attention.	34	3,22	34	2,82	45	-20	39	59
Let me just ask you, Were you ever in a car with a pair of binoculars near their house?	35	3,36	32	2,71	30	73	9	352
We've had nice conversation with people from all over the world.	36	3,41	34	2,82	10	169	20	212
My father said that was the only way you could get a real good look at the faces, and he was right.	37	3,44	40	3,72	25	91	29	143
Take them to market please.	38	3,45	26	2,48	8	187	32	130
We realized that each of us was talking about something quite different than what the other thought.	39	3,59	39	3,57	25	91	24	180
You better come quick, I'm not sure how long it'll last.	40	3,9	38	3,56	37	42	44	2

Appendix 8. *Error processing cost vs. Hultfors (1986) error gravity index*

Stimulus sentence (learner errors)	Eye-tracking experiment		Hultfors (1986) evaluation study			
	Cost rank	Cost (msec)	Error gravity index	Foreign-ness index	Intelligibi-lity index	Rank
She was standing alone, beside her with rage.	1	380	7,22	4,19	3,03	2
When I five years ago visited London, I could hardly speak English.	2	352	5,43	3,59	1,84	33
I am sure of that he will come.	3	320	5,97	3,83	2,14	17
Excuse me, what is the clock, please?	4	316	6,92	4,5	2,42	5
This is my brother's-in-law cap.	5	312	4,94	3,17	1,77	45
Who learnt you Spanish?	6	305	5,36	3,5	1,86	36
He has an own company.	7	295	6,76	4,11	2,65	7
She was here for two years ago.	8	287	7,02	4,09	2,93	3
He had to pay five hundreds of pounds.	9	234	5,52	3,77	1,75	28
I was operated last week.	10	232	6,93	4,05	2,88	4
Smell on these flowers!	11	226	6,1	3,89	2,21	15
Things have went too far.	12	220	6,21	4,05	2,16	13
He needs not come.	13	217	5,95	3,73	2,22	18
He is an useful member of the team.	14	210	4,23	2,79	1,44	53
I didn't saw him today.	15	206	5,8	3,9	1,9	20
It was little else to do.	16	188	6,12	3,65	2,47	14
I look forward to hear from you.	17	187	4,27	2,89	1,38	52
They fighted bravely in the war.	18	177	5,03	3,5	1,53	43
I took the blind by the arm and led him across the street.	19	176	5,36	3,45	1,91	37
This is the goodest cake I have ever tasted!	20	169	5,08	3,58	1,5	42
There's a long time since I saw her.	21	168	5,98	3,91	2,07	16
You do good coffee!	22	162	4,57	3,09	1,48	50
Became he a dentist?	23	162	6,53	4,33	2,2	10
There is one chair too much. Shall I put it away?	24	160	4,51	3,03	1,48	51
I shouted to him to look up for the car.	25	148	5,62	3,38	2,24	24
This room smells food, doesn't it?	26	147	6,87	4,26	2,61	6
He told me to not worry.	27	137	3,64	2,29	1,35	56
I am used to get up early in the morning.	28	129	5,4	3,54	1,86	35
Here are the money I owe you.	29	128	5,74	3,92	1,82	21
I don't know what improvements that are being planned.	30	125	5,19	3,18	2,01	39
He has a blue car and I have a red. And you?	31	125	3,34	2,03	1,31	58
I am born in 1945.	32	116	5,54	3,86	1,68	27
The most people would agree with you.	33	116	5,57	3,57	2	26
Many cities have stopped to expand.	34	115	6,69	3,83	2,86	8
He is a very alone man.	35	115	5,73	3,71	2,02	23
What a dreadful weather!	36	113	5,41	3,71	1,7	34
It is a hard work to write a book.	37	112	4,73	3,2	1,53	48
Yesterday's accident depended on the bad weather.	38	112	6,26	3,61	2,65	12
She has very limited knowledges of German.	39	110	4,62	3,06	1,56	49

He thinks not that they know what to do.	40	108	8,04	4,53	3,51	1
He drives badlier than his brother.	41	108	5,94	3,91	2,03	19
I always tried to make my best at school.	42	99	5,74	3,7	2,04	22
I am staying at Sheraton hotel for three days.	43	98	3,36	2,13	1,23	57
Came he to the party yesterday?	44	93	6,34	4,17	2,17	11
We have a great deal of problems.	45	93	3,3	2,02	1,28	59
I don't know much people in this town.	46	93	5,46	3,66	1,8	32
He works very hardly and he likes his job.	47	90	6,54	4,07	2,47	9
This cake smells well! Did you make it?	48	85	5,5	3,67	1,83	30
She is married with a German.	49	83	5,16	3,56	1,6	41
The house was full of mouses, they were all over the place.	50	80	4,81	3,43	1,38	46
He speaks French quite good, he had a French nanny.	51	79	3,79	2,53	1,26	55
He is meaning this book, not that one.	52	74	5,47	3,62	1,85	31
The car was badly injured in the accident.	53	71	5,52	3,64	1,88	29
I met him in the stairs.	54	68	5,34	3,53	1,81	38
My sister is nurse and she works at a hospital.	55	55	5,16	3,54	1,62	40
I didn't buy nothing at the supermarket.	56	52	4,19	2,56	1,63	54
It took me a hour to get there.	57	51	3,29	2,1	1,19	60
My brother have worked for this company for three years now.	58	22	4,96	3,38	1,58	44
I haven't heard something from him for a long time.	59	22	4,8	3,06	1,74	47
She had her radio beside herself and it played her favorite song.	60	13	5,62	3,61	2,01	25