

Saliency in Welsh English grammar: A usage-based approach

Katja Roller



Salience in
Welsh English grammar
A usage-based approach

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

vorgelegt von

Katja Martina Roller
aus Alzenau

Wintersemester 2015/16

Titel der eingereichten Dissertation: *On the relation between frequency and salience in morphosyntax: The case of Welsh English*

Erstgutachter: Prof. Dr. Dr. h.c. Bernd Kortmann
Zweitgutachter: Prof. Dr. Dr. h.c. Christian Mair
Drittgutachter: Prof. Dr. Stefan Pfänder

Vorsitzender des Promotionsausschusses
der Gemeinsamen Kommission der
Philologischen, Philosophischen und Wirtschafts-
und Verhaltenswissenschaftlichen Fakultät:
Prof. Dr. Christoph Huth

Datum der Fachprüfung im Promotionsfach: 12.05.2016

*Perception is of definite and
probable things*

WILLIAM JAMES 1890

Acknowledgements

Writing a PhD thesis reminds me of composing a piano piece. What we perceive in the end is the consonance of many notes, the melody. But in constructing and playing such a piece, a range of individual keys are involved, each of them contributing in its own way to the composition. In working on this book, there were many people functioning as “keys” for me, and I would like to give them my heartfelt thanks here. First and foremost, I would like to thank my supervisors Bernd Kortmann and Christian Mair for their continuous support, constructive feedback and valuable advice. My work benefited greatly from their comments and ideas, and I am grateful for having learnt a lot in the context of linguistic research from these inspiring teachers. My thanks also go to the other professors in our research training group GRK DFG 1624 “Frequency Effects in Language” for their very helpful critical comments from various perspectives. Many thanks to Stefan Pfänder in particular, who also was a great help in organisational GRK-related matters.

Thank you ever so much to all my colleagues in the GRK: for your support in so many ways (e.g. comments on methods and statistical matters, helping distribute questionnaires, reading through dissertation chapters – thank you so much, Stephanie, Luke and Laura!), but also for the wonderful time together in Freiburg! I’m looking forward to many more barbecues at Seepark, hikes in the Black Forest, quiz nights at O’Kellys and mini wok evenings to come!

Many thanks go to various other researchers from inside and outside of Freiburg. Thank you to Jonnie Robinson and Rob Perks of the British Library in London for granting me access to the Millennium Memory Bank, and for your support when transcribing parts of the collection for my Radio Wales Corpus. Thanks a lot to Paul Kerswill for granting me access to the corpus Linguistic Innovators: The English of Adolescents in

London. Many thanks to Peter Garrett, Peredur Davies, Jakob Leimgruber, Lowri Williams, Jonnie Robinson and all the others who helped me distribute my questionnaires in Wales and London. I would also like to thank our workshop guests in the GRK for their valuable feedback on my project at different stages: Peter Garrett, Marion Krause, Adriana Hanulíková, Chris Montgomery, Elissa Pustka, Josep Àngel Mas Castells and Paul Kerswill. Thank you to Christoph Wolk for your priceless assistance in statistical matters, and to the students of the University of Freiburg's English Department (above all Marion Kwiatkowski) who helped me transcribe my corpus. Thank you to the coordinators of our research training group, Monika Schulz and Michael Schäfer, for your help with a myriad of things organisational in the context of the GRK. Thank you, Manfred Krug, for introducing me to linguistic research. Your enthusiasm for linguistics sparked my interest in carrying out such a research project myself.

Last but not least, my heartfelt thanks go to many lovely people in my private life. To my parents, my sister Miriam, my brothers Christian and Jonas, my friends from Alzenau, Bamberg, Swansea and Freiburg: thanks for your support and for always believing in me. And thanks for bringing me back down to earth at times and showing me what really matters in life. Finally, thank you, Frank! You helped me with so many PhD-related things, such as programming, and contributed inspirational thoughts on salience from a physicist's point of view. Apart from that, thanks for being a *keystone* in my everyday life and for making me laugh so often. You gave me the balance and strength to attend to this project with passion!

Katja Roller, Freiburg, November 2016

Contents

1	Introduction	1
1.1	Introduction of the topic	1
1.2	Research aims	3
1.3	Structure of the book	5
2	Salience	8
2.1	Salience in science	8
2.2	Salience in linguistics	9
2.3	Salience in sociolinguistics and dialectology	11
2.3.1	Definitions and concepts of salience	12
2.3.2	Causes of salience	17
2.3.3	Effects of salience	23
2.4	Salience in the present study	25
2.5	Summary	25
3	Welsh English	28
3.1	General information	28
3.1.1	Historical background: The anglicisation of Wales	28
3.1.2	Varieties of Welsh English	31
3.2	Welsh English morphosyntax	33
3.2.1	The features under investigation	34
3.2.2	Research on Welsh English morphosyntax	43
3.2.3	Beyond morphosyntax: Characteristics of Welsh English phonology, intonation and lexis	50
3.3	English in London	52
3.3.1	General information	53

3.3.2	London English grammar	55
3.4	Summary	58
4	Research questions and hypotheses	60
5	Methodology	64
5.1	General research design	64
5.2	Salience survey	66
5.2.1	Conceiving the survey	66
5.2.2	Conducting the survey	76
5.3	Corpus analyses	78
5.3.1	Radio Wales Corpus	78
5.3.2	Linguistic Innovators corpus	87
5.4	Further studies	88
5.4.1	Analyses of TV clips	88
5.4.2	eWAVE study	90
5.4.3	Questionnaire: Perception of deviations from Standard English	91
5.5	Statistical methods	93
5.6	Summary	95
6	Results	97
6.1	Salience in Welsh English grammar	97
6.2	Salience and frequency	101
6.2.1	Frequency in Welsh English	102
6.2.2	Frequency in Welsh English vs. London English	108
6.2.3	Frequency in Welsh English vs. Standard English	113
6.2.4	Frequency in the media	118
6.2.5	Variation of frequency in Welsh English	120
6.2.6	Interim summary: Frequency	125
6.3	Salience and pervasiveness in the British Isles	126
6.4	Salience and perceived deviation from the standard language	138
6.5	Salience and social factors	147
6.5.1	Subjects from Wales vs. subjects from London	148
6.5.2	Subjects from Wales: Native language, region in Wales	151
6.5.3	Subjects from London: Contact with Welsh English	153

6.5.4	The role of incomers	158
6.6	Summary and discussion	161
7	Conclusion and outlook	168
7.1	Relating the results to the hypotheses	168
7.2	General gains and implications for future research	171
7.2.1	Usage-based linguistics	171
7.2.2	Perceptual dialectology	175
7.2.3	Welsh English	176
7.2.4	Outlook: Towards a usage-based dialectology	177
	Bibliography	179
	Appendix	194
	German summary	211
	Welsh summary	214

List of figures

1	Main dialect areas of Wales	31
2	Overview of methods employed in the present study	66
3	Exemplary sentence from salience test I	70
4	Map showing the places of residence of the Welsh informants in the salience survey	78
5	Numbers of speakers in the RWC according to their decades of birth . .	85
6	Map showing the places where the interviews in the RWC were con- ducted	86
7	Welsh presenters and comedians named by the London subjects in the salience questionnaire	90
8	Salience values of Welsh English features, obtained in the first test of the questionnaire-based survey	99
9	Degrees to which the invariant tag question <i>isn't it</i> was assigned to different regions in the British Isles by subjects from London	100
10	Token frequencies of features in the RWC, ordered by the features' salience	105
11	Percentages of subjects who recognised Welsh English features with mean frequencies higher than or equal to/lower than the mean fre- quency of 1.37 per 10,000 words	107
12	Token frequencies of morphosyntactic features in the LI and the RWC .	110
13	Percentages of subjects from London who recognised Welsh English features with mean frequency differences higher than or equal to/lower than the average value of 0.2 per 10,000 words	112
14	Invariant tag questions and concord tag questions involving <i>isn't it</i> or <i>innit</i> in the RWC and the LI	115

15	Normalised frequencies of Welsh English grammatical features in different Welsh regions in the RWC	123
16	The relation between interlectal salience and overall pervasiveness in eWAVE for all 24 non-standard grammatical features under investigation	135
17	Degrees to which the Welsh English features were assigned to Wales, to other areas or nowhere by the informants from Wales	141
18	Test sentences in the questionnaire dealing with deviations from Standard English	143
19	The relation between the subjects' openness towards linguistic diversity and the number of Welsh English features assigned to Wales	152
20	Salience values of Welsh English features for insiders, incomers and outsiders	160
21	Salience questionnaire	194
22	Questionnaire – perception of deviations from Standard English	208

List of tables

1	Token frequencies per 10,000 words of Welsh English features in FRED	47
2	Numbers of spoken words in recordings from different dialect areas of Wales in the RWC	84
3	Salience values in per cent for subjects from Wales in tests I and II . . .	100
4	Absolute and normalised token frequencies of Welsh English features in the RWC	103
5	Normalised token frequencies of non-standard features in the RWC and the LI	109
6	Absolute token frequencies of Welsh English features in the media and the RWC	119
7	Token frequencies of Welsh English features in different age groups in the RWC	122
8	Token frequencies of Welsh English features for female and male speakers in the RWC	124
9	Indices of pervasiveness in eWAVE	127
10	Non-standard features under investigation in the eWAVE study	129
11	eWAVE-based pervasiveness ratings of non-standard features in Wales, in the rest of the British Isles and in Wales vs. the rest of the British Isles	131
12	Grammatical features of (varieties of) Irish English, Scottish English and English in England in eWAVE	134
13	Percentages of the features' perceived non-standardness and their salience values for Welsh subjects	139
14	Percentages of the features' perceived non-standardness and their salience values for London subjects	140
15	Social factors explored in the context of both intralectal and interlectal salience	149

16	Results of the generalised linear model with intralectal salience as the dependent variable and several social factors as independent variables	150
17	Social factors that the present study found to have statistically significant effects on intra- and/or interlectal salience	157
18	The feature-dependent determinants of salience explored in the present study and their relation to the features under investigation	164
19	Determinants of salience explored in the present study and their relevance for intralectal versus interlectal salience	166
20	Markup of the RWC	200
21	List of interviews in the RWC	203
22	List of TV clips analysed	208

Chapter 1

Introduction

1.1 Introduction of the topic

Imagine searching for a friend in a crowd. As one searches through the crowd, attention may be captured by elements that are inherently salient, for instance, a person in a red coat among people in black coats or by dynamic cues such as a person running. However, where one looks in the crowd also depends on one's knowledge. So, if the friend is known to be wearing a black coat and a blue hat, attention may be less likely to be captured by an otherwise salient red coat, but perhaps be misdirected to a blue coat as one searches for anything blue in the crowd (Lupyan 2010: 71).

The way we perceive the world and steer our attention does not only depend on the inherent prominence of objects, but also, and probably quite strongly, on our experiences. This general view on perception was already supported by William James in his 1890 book *The Principles of Psychology* (Vol. II): “[W]hilst part of what we perceive comes through our senses from the object before us, another part (and it may be the larger part) *always comes [...] out of our own head*” (emphasis in original; James 1890: 103). Our mind can interact with what we perceive in different ways, according to psychological research. As illustrated in the initial quote, there are situations where our attention is drawn to objects that we expect on the basis of previous experiences and/or that are relevant to us in order to achieve a personal goal (cf. e.g. Awh, Belopolsky and Theeuwes 2012; Taylor and Fiske 1978). On the other hand, items may catch our eye because they deviate from previous experiences and cognitive presuppositions, because they are unexpected or surprising (cf. e.g. Itti and Baldi 2009; Pattabhiraman

and Cerccone 1990).

The idea of prominent objects deviating from previous experiences is addressed in Schirmunski's (1928, 1929, 1930) pioneering study of dialect perceptions. Schirmunski argues that primary dialect features, those that are perceived consciously, stand out in relation to the literary language or to one's own dialect. Thus, those features attract attention that deviate from something (for example, the standard language) in the individual cognitive perception of a listener or speaker (cf. also Lenz 2010: 90).

What makes dialect features stand out, be it by contrast to or in line with previous experiences? One group of factors described by Schirmunski (1928, 1929, 1930) and later works (cf. e.g. Trudgill 1986) is language structure. For example, when a regional linguistic feature differs radically from the standard language, it may be more readily perceived (especially by speakers from outside the region) than a feature that is structurally similar to the standard language. The conspicuousness of a linguistic variable may also be enhanced when it is positioned in a prosodically or interactionally prominent position in a sentence or clause (cf. Cheshire 1996; Kerswill and Williams 2000). Apart from such language-internal determinants, the scholarly literature also points to effects of social and attitudinal factors on individual perceptions of linguistic features. For example, Mihm (1985) records that perceptions of one's own dialect can be influenced by the degree of personal sympathy for and identification with this dialect and by the consciousness of particular norms (probably conditioned by language education at schools). A factor that has received increased attention in recent studies is usage frequency, which potentially affects the salience of dialectal features in several ways. For example, when a feature occurs frequently – in relation to others – in language use, this may increase its conscious perception (cf. Labov et al. 2011). Apart from that, one might become aware of features from another dialect because they have a low probability of occurrence in one's own local way of speaking (cf. Rácz 2013). However, up to this point, there are many open questions concerning the relation between frequency and salience in regional and social dialects. For instance, most of the sociolinguistic and dialectological studies exploring salience (quantitatively) have so far focused on phonetic and phonological variation. It is not clear to what extent approaches as the one by Rácz (2013), predicting the salience of phonetic and phonological features on the basis of usage frequencies, can be transferred to the

area of dialect grammar. (Morpho-)syntactic variation differs from variation in the areas of phonology and lexis, as argued by Kortmann (2010: 842):

In contrast to phonological and lexical variation, syntactic variation can be characterized as follows: it is much subtler and less salient; it is less categorical, a matter of statistical frequency rather than presence or absence; it has a wider areal reach and is less restricted to very small areas or individual dialects.

As Jansen (2014: 108) sees it, “we are still at the beginning of understanding the phenomenon of salience in full and much more research is needed, especially on the morphosyntactic and discourse levels and on the effect that frequency can have on the salience of a feature.” What is more, to date, not much is known about how frequency interacts with other factors (such as structural or attitudinal ones) in the context of morphosyntactic salience, and whether such combinations of factors differ between individual linguistic features. In addition to that, it remains unclear how effects of frequency on salience differ between people from *inside* and *outside* of a particular speech community.

Using the example of Welsh English, this book explores how frequency, in combination with other factors, affects perceptions of non-standard grammatical features, and how such perceptions differ between insiders (people from Wales) and outsiders (people from London). The present study aims to answer the following central research questions:

- Does frequency influence the salience of dialectal grammatical features?
- Does frequency affect intralectal (people from Wales) and interlectal (people from London) salience in the same way?
- Which other factors influence salience in dialect grammar?

1.2 Research aims

The aims of this work span several linguistic disciplines. In the context of theories of “human linguistic competence” (Tomasello 2005: 5), I would like to contribute to

usage-based (cognitive-functional) theories. A core assumption in usage-based linguistics is that language use, for example in the form of usage frequencies, shapes our mental representations of language. These mental representations in turn influence, for instance, how language is perceived. As stated by Ellis (2012: 7), “[t]he more times we experience conjunctions of features, the more they become associated in our minds and the more these subsequently affect perception and categorization”. Along these lines, the present study aims to demonstrate that salience in usage-based, i.e. that it is grounded in frequencies of occurrence in language use. By focusing on morphosyntax, I seek to fill a gap, since, as stated above, not much is as yet known about the interplay of frequency and salience in dialect grammar.

My studies into laymen’s regional mappings of and attitudes towards dialectal features also aim to provide novel insights to the fields of perceptual dialectology and language attitudes research. For the former field, the (non-)accuracy with which dialectal features are assigned to Wales can contribute to the understanding of “ordinary people’s (as opposed to linguists’) beliefs about the distribution of language varieties in their own and neighbouring speech communities” (Garrett 2010: 229). By exploring morphosyntactic features, the present study can add to the findings by Williams, Garrett and Coupland (1996), who report on folk linguistic perceptions of Welsh English in the areas of pronunciation and lexis. As far as language attitudes are concerned, one of the aims of the present work – based on Garrett (2010: 176) – is to further examine how people’s perceptions and evaluations of language are shaped by their stances towards linguistic diversity. As Garrett (2010: 176) sees it, the “criterion of whether and to what extent people value sociolinguistic diversity is clearly a particularly powerful one [...]. It has not received attention in language attitudes research previously, but seemingly deserves more attention.”

Welsh English also seemingly deserves more attention in linguistic research. As discussed in Chapter 3, a relatively small number of studies has to date focused on English spoken in Wales (as compared to, for example, Irish or Scottish English). Quantitative accounts of present-day spoken Welsh English grammar are particularly rare, studies with corpora from later than the 1980s being probably very much restricted to Heli Paulasto’s databases (cf. Meriläinen and Paulasto 2014; Paulasto 2006). The present work offers a frequency-based backup for three Welsh English features discussed by

Paulasto and colleagues: focus fronting, the non-standard habitual progressive and the inverted word order in indirect questions. Furthermore, the work provides information on the use of three other grammatical features not explored by Paulasto (2006) or Meriläinen and Paulasto (2014). By that, and by relating usage frequencies to perceptions of Welsh English, I aim to contribute to research on English in Wales and to generally offer new insights for variationist sociolinguistics and dialectology.

1.3 Structure of the book

This book provides theoretical, methodological and practical accounts of salience in dialect grammar. As for theoretical accounts, **Chapter 2** (*Salience*) delineates notions of salience, proceeding from general concepts of salience in science (Section 2.1) over linguistic approaches (Section 2.2) to salience in the context of sociolinguistics and dialectology in particular (Section 2.3). Section 2.3.1 explores how salience is conceptualised in early and recent sociolinguistic and dialectological works, from pioneering studies by Schirmunski (1928, 1929, 1930) and Labov (1972b) up to more recent approaches as the one by Kerswill and Williams (2000). Furthermore, the section discusses different forms of salience in the sociolinguistic-dialectological framework, namely physiological, cognitive and sociolinguistic salience. Section 2.3.2 turns to potential language-internal and -external causes of (sociolinguistic) salience, focusing on usage frequencies, but also taking other (for example, structural and attitudinal) factors into consideration. To complete the picture, Section 2.3.3 presents an overview of studies dealing with consequences of (sociolinguistic) salience, for instance, regarding the loss or the adoption of linguistic variables in speech communities. Based on the review about salience research in the preceding sections, Section 2.4 defines salience in the context of the present study.

Chapter 3 (*Welsh English*) provides theoretical background information on English in Wales, again proceeding from the more general to the more specific. In Section 3.1, a historical overview of the anglicisation of Wales is followed by descriptions of different regional varieties of Welsh English. Section 3.2 addresses Welsh English morpho-syntax and includes structural accounts of the non-standard features investigated in this work (focus fronting, invariant tag question *isn't it*, non-standard habitual progressive, inverted word order in indirect questions, zero past tense of regular verbs,

that-clause replacing infinitival subclause). Furthermore, previous research on Welsh English grammar dealing with speech production and speech perception is outlined. The section concludes by sketching out some Welsh English features in the areas of pronunciation and lexis, since such features are likely to also shape people's perceptions of English in Wales (apart from grammatical items). As this work contrasts usage frequencies of features in Welsh English and London English, Section 3.3 provides an overview of English in London. After discussing general characteristics of London English (such as the most common accents and dialects), some grammatical features of London English are introduced and compared with Welsh English morphosyntax.

Chapter 4 (*Research questions and hypotheses*) presents central research questions of the project, discusses them in the context of the background literature provided in chapters 2 and 3, and, on the basis of these works, formulates hypotheses.

Chapter 5 (*Methods*) describes and justifies the methodological steps taken in the context of the present research project. After presenting the general research design in Section 5.1, Section 5.2 illustrates the methodological considerations at play in conceiving and conducting the salience survey. Section 5.3 deals with the two corpora used for the frequency-based analyses. As for the (1) Radio Wales Corpus, employed to analyse Welsh English, the section outlines the work steps in compiling the database and provides key data of the final corpus (for example, with regard to the age-distribution and the regional backgrounds of the speakers). Subsequently, some basic information about the (2) Linguistic Innovators corpus, consulted to analyse London speech, is provided, and it is explained how and why random samples were drawn from the corpus. Section 5.4 discusses further methods applied in this work, i.e. the analyses of TV clips, the study with the Electronic World Atlas of Varieties of English (eWAVE) and the questionnaire exploring to what extent non-standard grammatical features are perceived as deviations from Standard English. Finally, Section 5.5 explains the choice of statistical methods used to mathematically underpin the findings of the present study.

These findings are displayed and discussed in detail in **Chapter 6** (*Results*). Section 6.1 focuses on general results from the salience survey, finding that the invariant tag question *isn't it* and focus fronting are significantly more salient – to Welsh people and Londoners – than the other features under investigation. Sections 6.2 to 6.5 aim

at explaining why this is the case. A detailed analysis of token frequency as a potential determinant of salience is presented in Section 6.2. Frequency is approached from several angles, namely regarding usage frequencies in Welsh English, frequency differences between Welsh English and London English, frequencies of non-standard features as compared to functionally equivalent Standard English constructions and frequencies in “everyday” Welsh English versus in the media. Finally, the section addresses variation in Welsh English, using the examples of age, sex and regions in Wales. Section 6.3 analyses potential relations between salience and (1) a feature’s pervasiveness in Wales versus the rest of the British Isles as well as (2) the number of regions where a feature occurs in the British Isles apart from Wales. Section 6.4 elaborates on the degree to which salient features are seen as general deviations from Standard English, and it identifies several structural language-internal causes that might promote a feature’s perceived non-standardness. Section 6.5 scrutinises personal and social factors (such as age and attitudes to linguistic diversity) that contribute to interindividual differences in salience perceptions.

The findings from Chapter 6 are summarised and critically reflected on in **Chapter 7** (*Conclusion and outlook*) by referring back to the hypotheses proposed in Chapter 4. The book concludes by discussing implications of this work for research projects in the future.

Chapter 2

Salience

2.1 Salience in science

According to the *Oxford English Dictionary Online* (OED), *salience* can be defined as “[t]he fact, quality, or condition of [...] projecting beyond the general outline or surface” (OED 2014, s.v. “salience”). The term is derived from Latin *salire* which means ‘to leap’ or ‘to jump’ (cf. Vaan 2008: 535). In English, *salient* apparently emerged in the mid-16th century, first used as a heraldic term in the sense of ‘leaping’ and from the 18th century onwards also meaning ‘jutting out’ or ‘prominent’ (cf. OED 2014, s.v. “salient”). The psychological notion of *salient* as ‘standing out or prominent in consciousness’ (OED 2014, s.v. “salient”) arose in the 1840s. Nowadays, concepts of salience are investigated in several scientific disciplines such as psychology, cognitive (neuro-)science, computer science, political science and linguistics (cf. Delort 2009: 2–3). The views on what salience means exactly differ between and within those disciplines. Thus, definitions of salience always depend on the particular research context. In psychology, salience is explored, for example, in the context of attention research. According to Delort (2009: 3), “attention represents the process that enables organisms to select, among different sources of information, those that will receive cognitive processing. Information is selected according to its salience.” Thus, salience here refers to a characteristic of an object through which it attracts or catches the observer’s attention (cf. Delort 2009: 2–3). Similarly, salience (or saliency) in visual cognition refers to the attention drawn to specific locations in a visual scene (cf. Itti and Koch 2000: 1489). Itti and Koch (2000: 1489) state that “[d]espite the widely shared belief in the general public that ‘we see everything around us’, only a small fraction of the inform-

ation registered by the visual system at any given time reaches levels of processing that directly influence behaviour.” As to what motivates the choice of a specific spatial location, the authors state that there is much evidence for a two-component model of visual attention control: on the one hand, bottom-up mechanisms direct an observer’s attention towards items based on physical, sensory-driven cues; on the other hand, in top-down processing, variable selection is directed under the observer’s cognitive control, thus objects of attention are selected, for instance, based on individual expectations. In political science, the notion of saliency is debated, for example, with regard to public issues. “Public issue saliency refers to the importance and urgency that the general public ascribes to a certain issue on the political agenda in relation to other issues” (Everts 2011: 39). According to Everts (2011: 40), topics with a high public issue saliency (for example, a debate about the use of military force in a region) are likely to induce polarisation and more extreme attitudes among the interested than low-saliency topics.

What all these definitions have in common is that salient objects – material or mental – are singled out from the rest in people’s perceptions. As Auer (2014: 7–9) states, the saliency of an item cannot be restricted to the item itself, but always depends on how it is perceived. Moreover, saliency is generally context-dependent. Salient items can only be noticed in relation to the particular backdrop against which they stand out (cf. Auer 2014: 9). This makes saliency a typical instance of figure-ground dynamics (cf. Auer 2014; Kimchi and Peterson 2008).¹

2.2 Saliency in linguistics

Similar to the notions of saliency in the scientific disciplines described above, there are different concepts and definitions of linguistic saliency. A very general definition is brought forward by Delort (2009: 4): “The linguistic saliency of a segment characterizes its attractiveness from a linguistic viewpoint”. This definition focuses on saliency as a property of an item, but does not include people’s perceptions of this property.

¹According to Kimchi and Peterson (2008: 660), figure-ground segmentation, which has its origin in Gestalt psychology, can be defined as “the process by which the visual system organizes a visual scene into figures and their backgrounds”. The authors describe figures as having definite shapes whilst grounds are “shapeless near the contours they share with figures and appear to continue behind the figures” (Kimchi and Peterson 2008: 660).

The latter aspect is included in Kerswill and Williams's (2000: 63) definition of linguistic saliency: "a property of a linguistic item or feature that makes it in some way perceptually and cognitively prominent". Similarly, Lenz (2010: 94) stresses the cognitive component of saliency. According to her, linguistic saliency can be understood as the cognitive conspicuousness of a linguistic feature in the sense that a linguistic element is singled out from its context and thus is able to enter a person's linguistic awareness more easily and faster than non-salient variants.²

Saliency is explored in several linguistic subdisciplines, such as language acquisition research, pragmatics, semantics, textlinguistics, cognitive linguistics, sociolinguistics and dialectology. In language acquisition research, saliency is investigated, for example, by Ellis (2012) as a potential determinant of learning. In Ellis's usage-based approach, saliency refers to the "general perceived strength of stimuli" (Ellis 2012: 15). Stimuli with a high saliency are more readily learnt than less salient cues. As an example from the second language acquisition of English, Ellis refers to different indications of present time: in the input, *today* is perceived more easily, thus it is more salient for language learners than the third person singular *-s*, which might be overshadowed and blocked, hence more difficult to acquire. Saliency in pragmatics is discussed, for instance, in the framework of Giora's (1998) graded saliency hypothesis. When processing utterances with several possible interpretations, "salient (i.e., coded) meanings of words or expressions (whose degree of saliency is affected by e.g., frequency, familiarity, conventionality) and salient (e.g., frequent) structures should always be accessed and always first, regardless of contextual bias or speaker's intent" (Giora 1998: 85). This implies that when the salient meaning is intended, (e.g. the figurative meaning of a metaphor) it is accessed directly without processing another interpretation, whereas when a less salient interpretation is required (e.g. the verbatim meaning of a conventional idiom), it is accessed sequentially "upon which the more salient (albeit unintended) meaning should be accessed first, before the non-salient intended meaning is derived" (Giora 1998: 90). Apart from the notion of saliency connected with familiarity and conventionality, the semantic-pragmatic discussion also refers to saliency as a

²Translation from German: "Unter Salienz wird [...] die kognitive Auffälligkeit eines sprachlichen Merkmals verstanden, in dem Sinne, dass ein sprachliches Element aus seinem Kontext hervorgehoben wird und dadurch dem Sprachbewusstsein leichter und schneller zugänglich ist als nicht-saliente Varianten".

property of *new* information in an utterance. According to Zhang, Hasegawa-Johnson and Levinson (2006: 440), the semantically salient parts of speech are (content) “words containing new information neither presupposed by the interlocutor nor contained in the preceding part of the utterance”. In Delort’s (2009) textlinguistic approach, saliency refers to the “foregrounding” of objects in a text due to visual or linguistic criteria. Foregrounding, as Delort sees it, can not only be caused by irregularities (e.g. a misspelled word, a word in bold print), but also by unexpected parallels (e.g. rhymes). In any case, “a segment of a text is (subjectively) salient for a reader if he/she reacts to its content” (Delort 2009: 7). In the cognitive linguistic discussion, two broad types of saliency can be distinguished, which are reminiscent of the two-component model of attention control discussed in Section 2.1. On the one hand, as described by Schmid (2007: 119), objects can be *cognitively salient* when they are activated in a person’s working memory and hence are in the centre of attention; using a salient concept goes hand in hand with minimal cognitive efforts and processing costs. On the other hand, *ontological saliency* refers to physical properties of features not bound to the mental activation of concepts. As a consequence, “saliency” may refer both to “a temporary activation state of mental concepts (*cognitive saliency*) and an inherent and consequently more or less permanent property of entities in the real world (*ontological saliency*)” (emphasis in original; Schmid 2007: 120). The concepts of cognitive and ontological saliency also arise in the sociolinguistic-dialectological discussion of saliency. As this work is positioned at the interface of sociolinguistics and (perceptual) dialectology, the following section discusses notions of saliency in these linguistic branches in detail.

2.3 Saliency in sociolinguistics and dialectology

According to Auer (2014: 7), saliency itself, its causes and its effects are often mixed up in the (socio-)linguistic discussion of saliency. For example, some authors see language change both as a factor leading to and as a consequence following from saliency (for a detailed discussion of saliency and language change, see Section 2.3.3). To avoid such circularity, the following sections are structured as follows: (1) determining what saliency *is*, i.e. how it is defined in the context of sociolinguistics and dialectology, including the historical development of the notion(s) and different forms of saliency, (2) considering possible determinants of saliency, with a special focus on frequency, and (3) discussing consequences that may follow from saliency.

2.3.1 Definitions and concepts of saliency

Historical development of the notion(s) of saliency

A pioneering work dealing with saliency (“*Auffälligkeit*”) in dialect change was Schirmunski’s (1928, 1929, 1930) study of German-speaking linguistic enclaves in Russia. Witnessing ongoing dialect changes in several colonies, he observed that some dialectal phenomena seemed to disappear relatively easily, whilst others were much more resistant to change. Schirmunski concluded that features which are very conspicuous (*primary dialect features*) are susceptible to loss, whereas less noticeable features (*secondary dialect features*) tend to be preserved. Primary dialect features, according to Schirmunski, particularly stand out as deviations from the literary language (or another dialect). Secondary dialect features involve minor deviations from the literary language (or another dialect). Moreover, primary features can enter a speaker’s or listener’s metalinguistic awareness in that one can consciously point to a feature and use it in linguistic ridicule. This is not possible for secondary features, as speakers or listeners are either not aware of a feature at all, or they may feel that there is some dialectal colouring to an utterance, but they cannot pinpoint the exact reason for it. It should be stressed, though, that Schirmunski’s conceptualisation of “*Auffälligkeit*” is not a strictly dichotomous but rather a gradual one, with primary and secondary features as two opposing tendencies at the endpoints of a continuum. Thus, not all dialectal features he explored could be clearly assigned to one of the categories. Schirmunski’s work has been influential in (particularly German) dialectology and traces of it can still be found in current variationist linguistics (cf. e.g. Auer, Barden and Grosskopf 1998; Lenz 2010).

In English linguistics, foundations in the saliency discussion were laid by the early works of William Labov (1966, 1972a). He did not use the term “saliency” himself then – it was introduced into linguistics by Trudgill (1986) – but his classification of variables as *indicators*, *markers* and *stereotypes* significantly influenced later saliency-related projects. The following paragraph refers to his works cited above, in which Labov explored the social stratification of linguistic variables and socially motivated linguistic changes in the United States. He classified the variables involved in these changes “according to the kind of social evaluation that they receive” (Labov 1972a: 314). Thus, indicators are features that show social differentiation by social group or

age, but they receive only little (if any) social evaluation and are not involved in style shifting. An example of an indicator would be the *hock-hawk* merger (which refers to the loss of distinction between the short /o/ and the long open /oh/ in words like *hock* and *hawk*, with the “migration of /o/ to the subset of long and ingliding vowels” (Labov 2010: 13)). Markers, on the other hand, do not only show social but also stylistic stratification. Features such as the non-prevocalic /r/ lead to regular responses in reaction tests, but people may not be consciously aware of using or avoiding them – in contrast to stereotypes. These are forms “prominently labelled by society [...] [and] part of the general knowledge of adult members of the society” (Labov 1972a: 314). The label is attached to a feature even if it does not comply with objective facts. Stereotypes are part of the metalinguistic discussion in a speech community. As an example, “Bostonian” is often indicated by “Pahk your cah in the Hahvahd Yahd” (Labov 1972a: 315) due to the *r*-less pattern with the low central fronted vowel [a:] in Boston speech.

Trudgill (1986) introduced the term of linguistic “saliency”, based, for instance, on works by Labov (1972a), Timberlake (1977) and Kerswill (1985). In his explorations of dialect contact and change, he observed that some linguistic features are perceived more consciously than others, which results in modifying the more conspicuous features in particular situations. “This is because of the *saliency* which attaches to markers and indeed turns variables into markers in the first place” (Trudgill 1986: 11). Saliency is thus defined as a *property* of a marker. According to Kerswill and Williams (2000: 65), Trudgill’s concept of saliency is particularly apt in dialect contact (as opposed to language contact), since accommodating to a feature may not so much be influenced by structural difficulties or differences and more by sociolinguistic factors.

Trudgill’s notion of saliency has influenced models of saliency in several newer sociolinguistic and dialectological publications, such as the work by Kerswill and Williams (2000). They investigated dialect levelling in three urban areas in England. Saliency came into play here as one possible factor influencing the social and geographical spread of different linguistic variables (such as T-glottaling or H-dropping). In that context, the authors stress the importance of Trudgill’s concept, but also criticise the circularity of some of his arguments regarding explanatory factors of saliency (for a

more detailed discussion of causes of saliency, see Section 2.3.2). The authors' own model of saliency includes the following three components:

1. the presence of a linguistic phenomenon whose explanation we suspect may be due to the saliency of the linguistic feature or features involved. Typically, the phenomenon will be a particular pattern observed in language change, language variation, the variable behaviour of individual speakers, or the acquisition of a linguistic feature. In cases of language change and variation, the linguistic features are items being transferred from one language variety to another through diffusion; however, diffusion-type mechanisms may hold for the other types of phenomena as well.
2. language-internal explanations [...]
3. language-external [...] factors

(Kerswill and Williams 2000: 90–91)

Regarding an extension of Trudgill's model, Kerswill and Williams (2002) also point to Preston's (1996) conception of folk awareness. The *modes of awareness* discussed there "seem [...] to have the potential significantly to extend and deepen Trudgill's notion of 'saliency'" (Kerswill and Williams 2002: 173). I concur, and assume that this conception can help to operationalise a feature's degree of saliency in people's minds. The following modes of awareness can be distinguished (cf. Preston and Niedzielski 2000: 22–23). (1) *Availability* refers to the degree to which linguistic facets are available to people. This ranges from *unavailable*, where features are not commented on in the population, via *available* and *suggestible* to *common*, where features are typical topics in popular discourse on language. (2) *Accuracy* distinguishes *inaccurate* from *accurate* descriptions of language phenomena. (3) *Detail* bears on the degree of specificity in people's accounts of features, from a *global* to a *specific* pole. (4) Speakers may be more or less in *control* of the linguistic forms; this can both refer to the ability to report on a feature and to perform it. Importantly, all those categories are seen by Preston as continua.

These works – and also the present study – generally focus on *extra-situational saliency*, i.e. the conscious perception of linguistic features beyond specific conversations, as in the case of the low central fronted [a:] being a common stereotype of Boston speech (cf. Labov 1972a: 315). However, as pointed out by authors such as Cheshire

(1996) and Schmid (2014), linguistic items can also be *situationally salient* in particular communicative contexts, for example because they are important for the interaction between the interlocutors, or simply because they stand out acoustically.

Kinds of salience in the sociolinguistic-dialectological framework

As stated above, salient features stand out as figures from a ground. Auer (2014: 9) suggests to distinguish three kinds of salience depending on the particular ground from which a feature protrudes: physiological, cognitive and sociolinguistic salience.

Physiological salience is sometimes also called *physical salience*, *perceptual salience* (cf. Delort 2009: 9), *stimulus salience* (cf. Awh, Belopolsky and Theeuwes 2012: 437) or – as mentioned above – *ontological salience* (cf. Schmid 2007: 120). It generally refers to the prominence of a figure due to its physiological properties and independent of the observer's mental state (cf. e.g. Auer 2014; Awh, Belopolsky and Theeuwes 2012). Stimuli are selected according to “bottom-up” (Delort 2009: 3) mechanisms, in a “stimuli-driven” (Delort 2009: 3) manner. In attention research, directing attention to an object on the basis of its low-level characteristics is also referred to as “exogenous attentional control” (Awh, Belopolsky and Theeuwes 2012: 437). General examples of physically salient objects are colour contrasts or sudden loud noises (cf. Awh, Belopolsky and Theeuwes 2012; Delort 2009). In linguistics, physical salience can play a role, for instance, when listening to an unknown language. In that case, according to Auer (2014: 9), attention is directed towards linguistic elements that can be isolated from their surroundings particularly well, for example, because they are articulated more slowly or more loudly. Thus, a possibility of operationalising and measuring physiological salience would be to construct tests where the subjects are not familiar with the language of the stimuli (cf. Auer 2014: 12).

Cognitive salience is also called *conceptual salience* (cf. Delort 2009: 3). When a concept is cognitively salient, as stated above, it is mentally activated, and this happens on the basis of a person's previous experiences and/or needs. Cognitive salience hence is clearly subjective, relating to an individual's interpretations, memory characteristics and emotions (cf. Delort 2009: 3). “Top-down” processes steer the observer's attention in a goal-driven manner, in other words, the observer exerts “endogenous attentional control” (cf. Awh, Belopolsky and Theeuwes 2012: 437). There are two forms

of cognitive linguistic saliency discussed in the literature. First, a familiar object can be salient when it is deeply entrenched in the mind, because it is cognitively preactivated (cf. Schmid 2007: 120). Second, however, a feature can also be salient and thus attract attention when it is unexpected or surprising, hence deviating from expectations and experiences with language (cf. Auer 2014; Rácz 2013). In the following additional illustrations of cognitive saliency, I refer to Auer (2014). As opposed to physiological saliency, cognitive (as well as sociolinguistic) saliency is always based on a contrast between two ways of speaking – features only become conspicuous when compared with a different mode of speaking, for example, in comparing a regional dialect with the Standard. However, the difference between expected, well-known and unexpected, surprising objects is not categorical. For example, people tend to be aware of features from neighbouring dialect areas – they would not use them themselves, but would not be surprised if such a feature appeared in conversation. The central Austro-Bavarian articulation of /ɔa/ in German words like *weiß* /vɔas/ ('know', typically pronounced /vars/ in Standard German) may not be in the active vocabulary of speakers from Swabia, but they might nevertheless know it passively from their Bavarian neighbours. As to how cognitive saliency can be operationalised, one possibility would be to present subjects with spoken or written texts where they have to identify all deviations from the standard language (cf. e.g. Elmentaler, Gessinger and Wirrer 2010: 115–116). The degree of saliency would depend on the number of deviations correctly identified. A more advanced test design could include measuring reaction times (cf. Auer 2014: 12).

Sociolinguistic saliency. According to Rácz (2013: 37), a dialectal marker is socially salient when it carries social indexation. Salient variables in this sense are labelled, for example, by region, social class or gender. For instance, definite article reduction is a feature commonly associated with the counties Lancashire and Yorkshire in the North of England (cf. Rácz 2013: 56). As Auer (2014: 10) sees it, sociolinguistic saliency always involves the social and/or (negative or positive) affective evaluation of the feature. Both Rácz (2013) and Auer (2014) assume that a feature must first be cognitively salient in order to become socially salient. Not all cognitively salient variables are also socially salient, however. Depending on the social labelling or evaluation, only some features are able to pass through the sociolinguistic filter (cf. also Labov et al. 2011). Socially salient features are considered to be more conspicuous than features which are salient in a cognitive or physiological sense (cf. Auer 2014: 12). Another aspect of

sociolinguistic saliency, scrutinised in this work, refers to the question *to whom* something is salient. Auer (2014: 13–14) distinguishes *intralectal saliency*, the conscious perception of features of one’s own dialect or language, and *interlectal saliency*, the conscious perception of features from other dialects or languages. Methods of measuring sociolinguistic saliency involve, for example, exploring people’s speech behaviour in situations with different degrees of formality, metalinguistic utterances, hyperforms (e.g. employed by people to follow a particular speech norm), dialect writings (under the assumption that the most salient features can be found in writing), perception experiments, analyses of code-switching and language ridicule and satire (cf. e.g. Auer, Barden and Grosskopf 1998; Lenz 2010: 95–96).

2.3.2 Causes of saliency

Frequency

This section sets out to explain why frequency could be an important factor influencing saliency in the sociolinguistic-dialectological framework. In other words, arguments are presented for why we can assume *frequency effects* on saliency. This does not mean that I try to explain saliency solely on the basis of frequency. Instead I explore how successful usage frequencies are as a measure of saliency, and where the limits to this approach lie. The discussion generally draws on findings and hypotheses from usage-based (also called cognitive-functional) linguistics and from cognitive science.

A core tenet of usage-based linguistics is that “language structure emerges from language use” (Tomasello 2005: 5). Mental representations of language are assumed to be organised according to our (individual) experiences with language throughout our lives. Experience with language, among other things, can mean being exposed to linguistic features with different usage frequencies. Ellis (2012: 7) assumes language users to be sensitive to input frequencies, and that these frequencies have an impact on perception. Along similar lines, Bybee (2006: 711) states that the frequency of use of constructions influences their mental representation, as evidenced, for instance, in people’s recognitions of conventionalised items. A theoretical account able to model frequency effects in language is Exemplar Theory (cf. e.g. Bybee 2006; Rácz 2013). The theory is mostly applied in the context of (socio-)phonetics, however, Bybee (2006: 714) also proposes its application in morphosyntax. In exemplar-based models, all ut-

terances – each one an *exemplar* of one or more linguistic constructions – are stored in the brain in rich detail. The theory thus posits concrete individual representations of utterances instead of mere abstract prototypes. The exemplars are stored in memory in categories on the basis of similarity. Bybee (2006: 727) suggests that frequency plays a role in the organisation of exemplars: those exemplars that are most frequent might function as the prototypes of categories while less frequent members are positioned at their margins. An exemplar-based organisation of the mental lexicon, then, is frequency-sensitive.

Carrying forward the idea that perceptions of language are influenced by usage frequencies, what does this mean exactly for saliency? Based on the findings from the cognitive sciences and cognitive linguistics discussed above, I hypothesise that frequency can influence saliency in two scenarios. It should be noted that these two scenarios do not conflict with each other, but that they can be integrated into a common approach, for example as regards dialectal markers (see below). First, a linguistic form can be salient, i.e. in the centre of attention, when it is highly frequent. As pointed out before, it can be assumed that some concepts are foremost in one's mind because they are familiar and thus deeply entrenched. The usage frequencies of these concepts could have an effect on their entrenchment, and thus on their degree of saliency. (This is a simplified argumentation, of course. As discussed below, we should expect additional factors apart from frequency to influence saliency in various ways.) The interaction of high frequency and saliency has been explored in several studies, such as Bardovi-Harlig's (1987) work. She investigated the acquisition of the two constructions preposition stranding (*Who did Eve give the pen to?*) and pied piping (*To whom did Eve give the pen?*) by learners of English as a foreign language. Surprisingly, the informants acquired the former construction faster, even though their native tongues did not allow a construction like preposition stranding. Bardovi-Harlig concludes that this is because of the construction's saliency, which is defined as "availability of data" (cf. Bardovi-Harlig 1987: 401). In other words, preposition stranding is more salient because of its high frequency in the English input available to the learners. Kerswill and Williams (2000: 66) however argue that high frequency should not be considered the only influential factor here – it might well be that the construction's non-existence in the subjects' mother languages also influences saliency. The contrast effect when encountering such a construction might direct the learners' attention towards it. Labov et al.

(2011) explore the sociolinguistic acceptability of the non-standard realisation of /ɪŋ/ as /ɪn/. Participants were asked to rate a newsreader reading several versions of a text with different frequencies of /ɪn/ in terms of professionalism. The authors find that (1) the more frequently the non-standard variant occurred in a text, the higher its saliency was, in the sense that a deviation from the norm was detected, and (2) that listeners are extremely sensitive to different input frequencies. A study that comes closer to the present work in that it explores saliency with respect to the conscious perception of dialectal grammatical features was carried out by Kerswill and Williams (2000, 2002). They conducted a “grammar saliency test” (Kerswill and Williams 2000: 86), a questionnaire with sentences containing non-standard features. Participants had to state whether or not the features occurred in their town. Afterwards, the results of this test were compared with language use data of the informants, gained during interviews. The authors found that for a range of features, high or medium use (by working-class speakers) went hand in hand with a high rate of recognition. However, some features which were strongly localised (*stereotypes* in the Labovian sense) could still receive high recognition rates when their usage frequencies were low. For example, the subjects from Hull, England, considered the preterite *done* instead of *did* to be a typical local feature even though it seems to be largely absent from the area in terms of usage frequencies.

My second hypothesis is that a linguistic form can be salient when it is very infrequent. If the occurrence of a structure is highly unlikely and/or surprising in a particular environment, this might direct the observer’s attention towards it (cf. Auer 2014). The unexpectedness of linguistic features may be due to their low usage frequencies. Hume and Mailhot (2013) stress – in their work on phonologisation – that attention is focused on items with high surprisal, i.e. a low probability of occurrence in language use. Rácz (2013) offers some quantitative evidence for the low frequency hypothesis. In his sociophonetic approach, he finds that the saliency of some phonetic-phonological features – to speakers from outside a particular geographical area or social class – can to a certain extent be predicted by the features’ unlikelihood of occurrence in the dialects/sociolects of the outsiders. One of the linguistic variables explored by Rácz is T-glottaling. It is generally a widespread feature in the South of England, however, when it occurs word-finally before a vowel, it is salient of working class but not of upper-middle class speech. In his corpus study, Rácz demonstrates that the feature’s

saliency is reflected in usage frequencies; the upper-middle class subjects produced glottal stops in word-final pre-vocalic position with a much lower frequency than the working class speakers.

Overall, it seems to me that low-frequency-induced saliency and high-frequency-induced saliency are not contradictory, but that they can be combined in different scenarios, for example with regard to the sociolinguistic saliency of dialectal variables. This is illustrated by the following assumptions. Saliency is defined in both cases as the degree to which non-standard features are consciously associated with a particular regional language variety.

1. When saliency refers to intralectal (insiders') perceptions of regional features, it goes hand in hand with *high usage frequencies* and high degrees of familiarity with the constructions in the target variety; likewise the constructions may be *less frequent* in other (e.g. bordering) language varieties which makes speakers realise that the features are distinctive for their own area.

2. When saliency refers to interlectal (outsiders') perceptions of regional features, this is accompanied by *low probabilities of occurrence* of the features in the variety of the outsiders, potentially coupled with *high usage frequencies* in the target variety.

Other causes

As stated above, Rácz (2013) demonstrated that operationalising saliency as low token frequencies yields some insights into speakers' perceptions of linguistic features. This concept cannot be applied, though, regarding derhoticisation in Glasgow (cf. Rácz 2013). Even though working-class speakers realise coda /r/ to a lesser extent than middle class speakers, derhoticisation is no salient marker of Glaswegian adolescent working-class speech, possibly due to phonetic causes: there are a range of different variants to realise coda /r/, thus speakers have difficulties noticing differences between these variants and attaching specific social meaning to them. Mihm (1985), investigating varieties of German, found that some dialectal features were not assigned to people's home regions irrespective of their usage frequencies because of negative attitudes towards the regional forms. These accounts suggest that (sociolinguistic) saliency can be impacted by structural-linguistic and social factors. Indeed, potential

causes of saliency are often classified in the literature on the basis of linguistic and extra-linguistic factors. For the following illustrations, I use Kerswill and Williams's (2000: 90–91) terms of “language-internal” and “language-external” explanations of saliency.

A *language-internal* factor named frequently in the literature is the phonetic and/or articulatory distance in the realisation of a feature between different (regional or social) varieties. If variants differ from each other radically, they may be perceived more consciously than less distant variants (cf. e.g. Auer, Barden and Grosskopf 1998; Schirmunski 1930; Trudgill 1986). Similarly, a feature might be salient because it maintains phonological contrasts. For example, according to Trudgill (1986: 11), the word *Hugh* can be realised in Norwich, England, both as [hju:] and [hʊ:]. Only the former – the salient – realisation would maintain the differences in meaning between *Hugh* and *who*. The findings about derhoticisation in Glasgow described above imply that variables with gradual or continuous variants are harder to perceive and less conspicuous than dichotomous or categorical variables (cf. e.g. Auer 2014; Auer, Barden and Grosskopf 1998; Rácz 2013). Also, an item's saliency may be influenced by its prosodic and syntactic environment. A feature might stand out, for instance, because it is placed in a prosodically prominent position (cf. Kerswill and Williams 2000) or generally at the beginning or end of a linguistic unit (an utterance or a syllable, cf. Auer 2014). Further plausible linguistic causes include prosodic parameters such as pitch and loudness (cf. Auer 2014), semantic transparency, naturalness³ (cf. Kerswill and Williams 2000), comprehensibility to outsiders,⁴ the possibility of expressing the feature orthographically in dialect writings and lexicalisation (lexicalised rules of a dialect are more salient, cf. Auer, Barden and Grosskopf 1998). Beyond these structural peculiarities, language-internal explanations of saliency can also refer to a feature's occurrence in language use. In addition to frequency, as explained above, the areal distribution of an item could play a role (cf. e.g. Lenz 2003, 2010; Schirmunski 1930). Regiolectal features restricted to small-scale geographical areas may be conspicuous to outsiders because

³*Naturalness* is described by Kerswill and Williams (2000) as the counterpart of *markedness*. The markedness or unnaturalness of features in this context is connected to (e.g. morphological) difficulty. For example, the singular-plural pair *girl/girl-s* may be less marked because of the regular plural form than *sheep/sheep* (cf. Haspelmath 2006). It is plausible that marked forms are more conspicuous than natural ones (cf. Kerswill and Williams 2000).

⁴Schirmunski (1930: p. 184) argues that regional features which are hard to comprehend for speakers of other dialects are salient, while less salient features do not impair communication.

they are unexpected and unlikely to occur. A problem child in the debate about causes of saliency is linguistic change. Some authors (e.g. Trudgill 1986) claim that linguistic items are salient *because* they are involved in linguistic change, which leads to a circularity of the argument when also assuming that linguistic change *follows from* saliency. As Kerswill and Williams (2000: 73) see it, “[t]he way in which this factor can be maintained is if it is assumed that saliency does not itself lead to change, and that it is the change itself that causes speakers to notice the feature involved”.

Language-external explanations of saliency encompass a variety of perceptual and attitudinal factors as well as interindividual differences between the speakers. A factor frequently commented on are people’s individual attitudes towards and evaluations of linguistic features (cf. e.g. Kerswill and Williams 2000, 2002). As depicted above, Mihm (1985) found that in his study differences in saliency could only restrictedly be explained by quantitative usage differences and more by language attitudes. In his perception test, the majority of subjects from the German Ruhr region would not assign a text with a local speaker to their own area, probably because they were ashamed of identifying with Ruhr German themselves. Similarly, the findings by Hettler (2014: 86), who studies the perception of a range of Northern German regiolectal features, suggest that in some cases attitudes to local speech in Bremen and Hamburg have larger effects on the perception of features than one’s own language use. Kerswill and Williams (2000), in their dialect recognition test, observed differences between working class and middle class informants in the degree to which they considered local features as being local. The authors conclude that “even within a single town, there can be a lack of shared knowledge of local norms [...]”. It follows that saliency, however defined and however caused, will be different for different social groups” (Kerswill and Williams 2000: 86). Since language attitudes and evaluations seem to be able to inhibit frequency effects at some points, they are also scrutinised in the present work. As for further language-external factors, Kerswill and Williams (2002: 176) point to individual characteristics of the subjects that might go hand in hand with saliency, such as the kind of their social network (open or close-knit), their degree of (social and regional) mobility or their sociolinguistic maturity (relating mainly to age; cf. also Hettler 2014: 86). Some authors also list stereotyping/stigmatisation as a language-external cause of saliency. Trudgill (1986: 11), for instance, assumes a high awareness of overtly stigmatised features. Auer (2014: 14) also considers stigmatisation as a cause of saliency

and describes stereotypes as pathways leading to the social evaluation of a feature. I argue – and Auer (p.c.) agrees – that existing stereotypes may lead to people becoming consciously aware of features, but that some factors must have been involved to stereotype a feature, to raise attention to it, in the first place (frequency, for example). Thus, I do not consider stereotyping as the first link in the chain of salience causes, but rather as a potential enhancing factor at a later juncture. One case where stereotypes *could* play a role from the beginning on, though, is depicted by Laferriere (1979). In the study, Boston Jews identified a phonological low-prestige variable as “Irish” even though it was most frequently used in Boston by Italian Americans. An explanation would be that “Boston Irish” is a stigmatised dialect associated with a variety of social stereotypes, which is not the case for “Boston Italian”.

All in all, it seems useful to work with a combination of language-internal and language-external factors to explain the emergence of salience, while the factors’ exact dimensions depend on the exact operationalisation of salience. The studies just presented suggest that language-internal and language-external factors affect salience on different levels. While language-internal causes such as frequency and structural conspicuousness influence the salience of a feature to *average listeners*, the perceived prominence to *individual listeners* is shaped by language-external factors as language attitudes or mobility. This argumentation is in line with Purschke (2014), who distinguishes two components of salience in the context of linguistic regionalisms: (1) a variable’s *salience potential*, affected by feature-specific characteristics such as frequency, areal distribution and acoustic properties, and (2) a feature’s *salience perception*, connected to individual listener-based interpretations. The present study thus aims to examine and interrelate both components: an item’s average salience to listeners (from inside and outside a dialect area), with a special focus on frequency as a crucial explanatory factor, and interindividual differences in salience perceptions, which I intend to explain on the basis of social and attitudinal variation.

2.3.3 Effects of salience

Consequences of (socio-dialectological) salience are not in the focus of this work, hence they are only briefly illustrated here; they are not completely left out, however, as they seem to be tightly intervoven with salience and its causes.

In sociolinguistic and dialectological explorations of salience, it is often assumed that salient features (1) are more readily lost in the own speech community, and (2) are more easily adopted in dialect contact by speakers of another speech community (cf. e.g. Labov 1972a; Schirmunski 1930; Trudgill 1986). As to case (1), Trudgill (1986: 11) argues the awareness of dialectal phenomena may lead to speakers avoiding them in particular speech situations, for instance, when talking to speakers of another language variety. Moreover, those features which deviate prominently from the standard language or which are strongly stigmatised can undergo a reduction of use or even an extinction in a linguistic community (cf. Labov 1972a; Schirmunski 1930). However, Labov (1972a) finds that some constructions are stigmatised but still resistant to change and enduring. A potential explanation for this could be that the features carry covert prestige (for example, because they indicate group solidarity) which overrides the force of the negative overt prestige in linguistic change. Along these lines, Lenz (2010: 107) argues that salient forms functioning as regional identity markers may be kept even in formal speech situations. She concludes that salience should not simply be equated with a construction's potential of being reduced. Regarding case (2), Trudgill (1986) generally expects salient features to be diffused and accommodated to more quickly and to a greater extent in a dialect contact scenario than less salient items – unless there are other intervening factors at play. Those factors might be phonotactic constraints (in one's own dialect, which may make it difficult to pronounce a particular feature, for example), homonymic clash (very similar variables may cause confusion when being adopted) and extra-strong salience (heavy stigmatisation of a feature leads to not adopting it, for instance, because it sounds “too American”). Auer, Barden and Grosskopf (1998: 167) suggest that “salience is only a necessary, but not a sufficient, condition for a linguistic feature to be affected in accommodation”. A feature, so they argue, needs to carry some prestige or some other kind of positive evaluation in the original or the new dialect community to be adopted. Even though Labov (2001) and Trudgill (2014) doubt that social-evaluative and attitudinal factors are very influential in dialect diffusion (assuming the process is rather mechanical in nature), many works stress the importance of sociolinguistic factors in linguistic change, and that these factors can override the influence of salience in some scenarios (cf. Auer, Barden and Grosskopf 1998: e.g.). Kerswill and Williams (2000: 72) conclude that salience is “a factor that is *additional* to the list” (emphasis in original) of variables affecting the outcome of dialect contact.

2.4 Saliency in the present study

Based on the review about saliency research in this chapter, I now give a working definition of saliency as understood in the present study. I investigate the sociolinguistic saliency of regional morphosyntactic features, from a usage-based perspective. Sociolinguistic saliency means that this work focuses on the social labelling of features, the question why some variables are perceived as more characteristic or typical of a language variety than others (cf. Deumert 2003; Kerswill and Williams 2000; Rácz 2013). The present study specifically investigates grammatical phenomena considered as typical of regional language varieties. Within the scope of this work, a linguistic feature is thus defined as salient when it carries *regional indexation*, i.e. when it is consciously associated with a particular regional language variety.

I hypothesise that items which are cognitively salient generally have the potential to become sociolinguistically salient, based on the assumption that an exemplar-based storage of features is sensitive to language-external information (such as context, age, gender, cf. Rácz 2013). Frequency seems to be a crucial factor in turning cognitively salient constructions into sociolinguistically salient ones. Ideally, features can thus be assumed to be (intralectally and interlectally) salient when they are familiar and highly frequent in the target dialect, coupled with the fact that the features have a lower probability of occurrence in other (e.g. surrounding) dialect areas. The low frequency outside the target dialect may be important especially for interlectal saliency, assuming that a feature becomes particularly conspicuous for outsiders when it is unlikely to occur in their own region. However, some regiolectal features are prevented from becoming sociolinguistically salient, for instance due to extra-linguistic factors such as the negative evaluation of one's own dialect or due to intra-linguistic factors such as the pervasiveness of a feature across several dialect regions, which could make it hard to assign the feature to one area in particular.

2.5 Summary

This chapter has provided an overview of notions of (linguistic) saliency with a special focus on saliency in sociolinguistic and dialectological approaches. Clearly, there is much variation in conceptions of saliency, but comparing approaches from different

scientific disciplines I found several key points upon which the works agreed. Salient objects are prominent. They attract attention in relation to (1) a particular context and (2) people's individual perceptions. Thus, saliency is necessarily a subjective concept. *Linguistic* saliency, based on the studies cited above, can generally mean that a person's attention is drawn towards a particular linguistic item. Linguistic approaches to saliency are carried out, for example, in the areas of language acquisition research, cognitive linguistics, sociolinguistics and dialectology. In sociolinguistics and dialectology, saliency is often explored in the context of language variation and change, a pioneering work here being Schirmunski's (1928, 1929, 1930) dialectological study. Trudgill (1986) introduced the term of linguistic *saliency*. More recent sociolinguistic studies on saliency were conducted, for example, by Kerswill and Williams (2000) and Rácz (2013). There are different kinds of linguistic saliency that can be distinguished from a sociolinguistic-dialectological perspective: (1) physical saliency, connected to physical properties of the stimulus, (2) cognitive saliency, attributable to individual expectations, experiences and/or needs, and (3) sociolinguistic saliency, related to social evaluations of linguistic features. Frequency, according to previous studies, appears to be an important factor influencing saliency. For example, when a construction is frequently used in a given language variety in comparison with other constructions, it may become conspicuous (cf. e.g. Kerswill and Williams 2000). Also, a feature may be salient when it is frequently used in contrast to one or more other language varieties (cf. e.g. Rácz 2013). A comprehensive approach to causes of saliency should, however, also test for further language-internal (e.g. position in sentence/clause) and language-external (e.g. language attitudes) explanations. As for the consequences of saliency, several sociolinguistic and dialectological works assume that salient features are more likely to be lost in the home speech community, and more quickly adopted in dialect contact by speakers from another speech community. These saliency effects may be leveraged, though, within a speech community (e.g. when a variable carries covert prestige) or in diffusing features to new dialect areas (e.g. when they are heavily stigmatised). In this work, a feature is defined as salient when it carries regional indexation, i.e. when it is seen as characteristic of a regional language variety. I assume that frequency has an effect on this form of saliency in that salient features are highly frequent in the target variety in comparison with (1) other features in the target variety and (2) the feature's use in (an)other variety/ies. The frequency effect(s) may

be blocked by other intra- and extra-linguistic factors such as language attitudes and geographical spread, thus these factors are also examined in this work.

Chapter 3

Welsh English

3.1 General information

3.1.1 Historical background: The anglicisation of Wales

[B]y far the most Englishes in Wales are ‘contact varieties’ arising from the interplay of Welsh and English structural influences, reflecting the historical fact that such Englishes exist as the result of the language shift of cambrophone speakers and co-exist to a greater or lesser extent in actual geolinguistic space with Welsh itself (James 2011: 49).

The indigenous language of Wales is Welsh, a Celtic language with Indo-European roots (cf. Aitchison and Carter 1994; Davies 2014). According to Davies (2014: 4), the Celtic culture emerged in Central Europe around the eighth century BC, and the Celts were “perhaps the first Indo-European people to spread across Europe”. In pre-Roman times, Celtic speakers could be found in most parts of Western Europe, but with the rise of the Roman Empire, they were supposedly driven into the peripheries (cf. Penhallurick 2008b: 105). This means that in most of what is known today as Great Britain, Brittonic (a branch of the Insular Celtic languages and the ancestor of Welsh, Cornish and Breton) continued to be spoken during the time of the Roman power, despite Latin being the language of administration and law (cf. Davies 2014: 7). However, with the invasion of Great Britain by Germanic tribes (Angles, Saxons and Jutes) in the fifth century AD, Celtic speakers were pushed to peripheral areas in the north and the west, including Wales (cf. Baugh and Cable 2010: 47). While dialects of the Germanic tribes were dispersing in England, Wales retained the Welsh language, evolving from Brittonic around the sixth century AD. The resulting cultural divide between England and

Wales was given physical expression through Offa's Dyke (cf. Aitchison and Carter 1994; Hannahs 2013). The earthwork constructed by King Offa of Mercia in the eighth century AD separated Wales from England, and, according to Aitchison and Carter (1994: 23), "serves as a base line from which to chart the slow and complex westward retreat of the Welsh language".

The first significant transmission of English beyond Offa's Dyke happened during the Anglo-Norman invasion, which reached Wales around 1070 AD (cf. Aitchison and Carter 1994: 23). The settlers built strongholds throughout the south and the north of Wales, and considerable numbers of English speakers arrived (cf. Penhallurick 2008b: 105). Presumably most affected were the southern border regions with England and some southern coastal lowland areas, such as South Pembrokeshire and the Gower Peninsula; there, dialects of Welsh English have existed from the twelfth century onwards (cf. Aitchison and Carter 1994; Penhallurick 2008b).

English received its official status in Wales in the 16th century. Under the rule of the Tudors, the Acts of Union from 1536 to 1543 annexed Wales to England, making English the sole language of government and law in Wales (cf. Davies 1994; James 2011). Although the statute did not directly concern the use of Welsh in private contexts, it implied that Welsh people who wanted to be part of polite society would have to speak English. Welsh was seen as "the language of the barbarous past, English [as] the language of the civilised future" (Aitchison and Carter 1994: 26). However, according to James (2011: 48), English only replaced Welsh as the majority language of Wales in the course of the Industrial Revolution in the late nineteenth and early twentieth century. Large numbers of workers arrived from England, which led to significant increases both in general population figures and in the proportion of English-speaking residents (cf. Penhallurick 2007: 152). According to Davies (1994: 437), in 1891, 54 per cent of the Welsh population were able to speak Welsh whilst 69 per cent could speak English. English became the language of business and technology, which resulted in the public opinion that insufficient knowledge of the language would leave people poor and isolated (cf. Aitchison and Carter 1994: 33).

Another area of public life in Wales that English encroached upon was education. The Education Act of 1870 made it obligatory for all Welsh children to go to school until

the age of thirteen, and the language of instruction was English (cf. Davies 1994: 435–437). As stated by Davies (1994: 436–437), it has often been argued that this Education Act was the main reason for the decrease in the number of Welsh-speaking people in Wales. This could be connected to the fact that the 1870 act, apart from teaching children in English, also gave “a new awareness of English values and culture, and gave a powerful institutional fillip to anglicisation” (Williams 1990: 35).

As a result of these developments, in the first half of the twentieth century, English had become the “language of progress, of equality, of prosperity, of mass entertainment and pleasure” (Williams 1990: 36) in Wales, whilst many parents refused to speak to their children in Welsh (cf. Williams 1990: 36). Consequently, the numbers of people able to speak Welsh further decreased in the following decades from 26 per cent in 1961 to 18.9 per cent in 1981 (cf. Williams 1990: 41). Since then, however, the figures of Welsh speakers have stabilised. In recent years, Welsh has regained status and popularity in Wales, due to revitalisation efforts in the areas of government, media and education (cf. James 2011: 48). The Welsh Language Act of 1993 ensures that Welsh and English are treated “as co-equal in public life in Wales and [...] [that] all public-sector institutions are legally obliged to deal with the public bilingually” (Bishop 2006: 507). The media have also been an important factor in boosting the status of Welsh, for example by establishing S4C, a Welsh language television channel, in 1982 (cf. Bishop 2006; James 2011). Moreover, the Welsh Education Act of 1988 made Welsh a compulsory school subject for all children in Wales aged between five and sixteen (cf. Williams 2014: 250). Thus, an increasing number of young people are able to speak Welsh, but it is plausible that in the future Welsh might undergo “a process of ‘Latinization’, in which its use becomes restricted to a decreasing number of social domains as its traditional regional dialects decline” (Penhallurick 2008b: 106).

Despite the revitalisation efforts, English has remained the dominant language of Wales, nowadays being the first language of about 80 per cent of the Welsh-born population (cf. Trudgill and Hannah 2002: 30). What is more, (nearly) all Welshmen and -women who are native speakers of Welsh are bilinguals in Welsh and English (cf. Welsh Government 2012; Williams 2014). After a thousand-year-long process of anglicisation, English is, as Penhallurick (2008b: 107) sees it, “a thoroughly established language of Wales, a language used by and belonging to the Welsh people”.

3.1.2 Varieties of Welsh English

“Welsh English” will be used throughout this work to describe varieties of English spoken in Wales. However, the term should not mask the fact that there is linguistic diversity in Wales, tracing back, for instance, to influences from the Welsh language and English dialects from different parts of England (cf. Penhallurick 2008b: 107). The following overview of regional varieties of Welsh English is based on Garrett, Coupland and Williams (1999), who distinguish six larger dialect areas in Wales, as portrayed in Figure 1. It should be noted that the areas’ boundaries are fuzzy, though, and that speakers of the respective dialects are not necessarily confined to these regions.

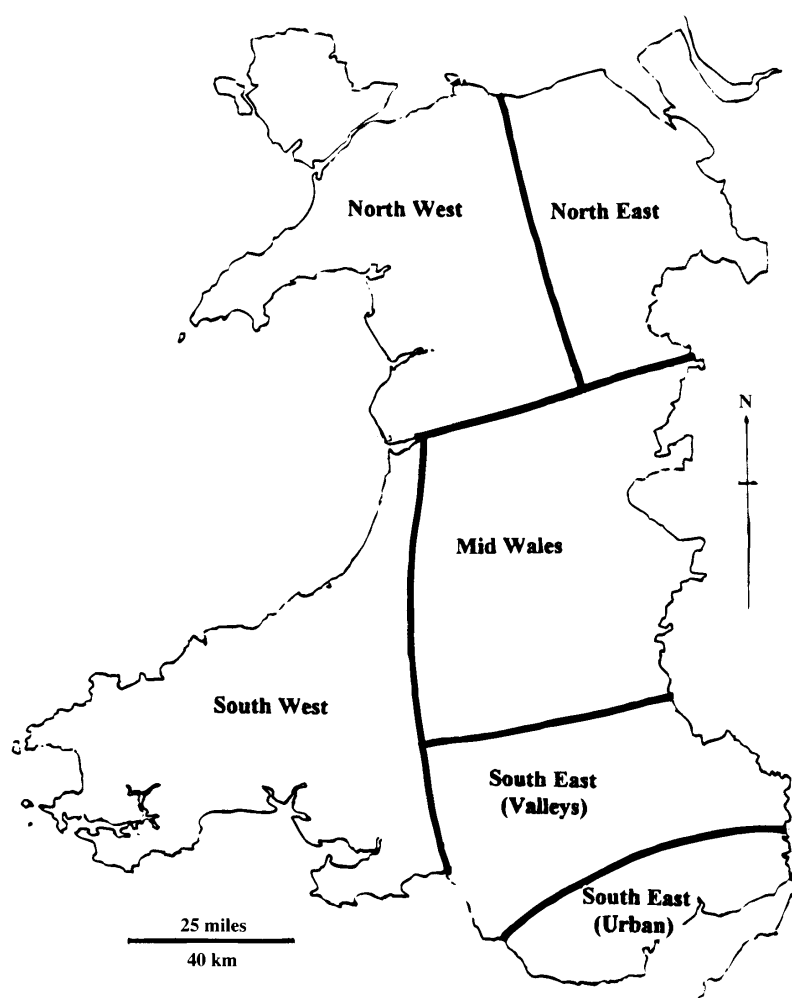


Figure 1: Main dialect areas of Wales according to Garrett, Coupland and Williams (1999: 325). Reprint of map approved by Peter Garrett (p.c.) and in line with the “Permissions Requests” policy of Cambridge University Press (2015)

The *South East (Urban)* area involves the capital city Cardiff and some nearby coastal towns such as Newport and Barry (cf. Awbery 1997: 88). The area is “historically very anglicized” (Garrett, Coupland and Williams 1999: 326), with English being spoken by the majority of the inhabitants since long before the nineteenth century (cf. Collins and Mees 1999: 186). Varieties of English here were supposedly influenced by dialects of nearby English areas like Gloucestershire, Somerset, Bristol and Avon (cf. Collins and Mees 1999; James 2011). Welsh English in the *South East (Valleys)* also exhibits structural influences from dialects of the South West of England, but traces from the Welsh language are more visible here than in the dialects of the Cardiff-Newport-Barry region.¹ Until the middle of the nineteenth century, areas such as South East Glamorgan and South West Gwent were predominantly Welsh-speaking, but with the Valleys becoming an important industrial zone for coal and steel production, a rapid linguistic transformation took place, resulting in a majority of monolingual English speakers in the area at the beginning of the twentieth century (cf. Collins and Mees 1999: 186). “The Valleys” play an important role in the political and cultural history of Wales, being “strongly associated with the (now decimated) mining industry and Welsh working-class culture” (Williams, Garrett and Coupland 1996: 187). This may be one reason why English from the South Eastern Welsh Valleys, sometimes labelled as “Valleys Voice” or “Wenglish”, is considered in other parts of the UK as “the (stereo-) typical manifestation of Welsh English *tout court*” (emphasis in original; James 2011: 47) and within Wales, it is seen as “a salient anglophone marker of Welsh identity” (James 2011: 47).

English in the Welsh *South West* appears to rely even more on Welsh-language influences. The area is considered to be a traditionally Welsh-speaking “heartland” (Garrett, Coupland and Williams 1999: 326), where, to the present day, large percentages of the population can speak Welsh (cf. Collins and Mees 1999: 185). As Coupland and Thomas (1990: 5) see it, “we must expect that the Welsh substratum will be most visible in the dialects geographically most remote from the earliest incursions of English”. West Pembrokeshire and Gower, however, are historically anglophone regions (due to early

¹This does not mean that varieties of English in the urban South East of Wales are totally free from Welsh-language influences, though. According to Thomas (1997: 67), similar to the Welsh North East, these South Eastern dialects have “substratal Welsh influences [...], but are now independent of contemporary Welsh influence, and we must expect them progressively to shed indigenous Welsh characteristics”.

arrivals of Anglo-Norman settlers, see above) with linguistic influences from South West England (cf. James 2011: 49).

In contrast to the Welsh-speaking heartland in the South West, agricultural *Mid Wales* is mainly non-Welsh speaking (cf. Garrett, Coupland and Williams 1999: 326). Varieties of English here show traces from rural dialects of adjacent English areas, such as the West Midlands, Cheshire and Shropshire (cf. Coupland and Thomas 1990; James 2011).

English dialects in the rural *North West* of Wales, similar to the South West, are heavily influenced by the Welsh substrate. However, “the” Welsh substrate can be further subdivided into different varieties of Welsh, influences of which are manifested in South-Western and North-Western Welsh English. According to Brake (2004: 8), “[d]espite its rich literary tradition, Welsh does not possess a spoken standard. The spoken language is broadly divided into North Walian and South Walian. However, these [sic] are regional differences within them.” Unlike the North West, the comparatively industrial *North East* of Wales is a relatively anglicised zone, with influences from urban English dialects of Merseyside (e.g. Liverpool) and Greater Manchester (cf. Garrett, Coupland and Williams 1999; James 2011).

In sum, it seems that what distinguishes Welsh English from other varieties of English in the British Isles is that its dialects have been shaped by an interplay of Welsh-language influences and dialects from neighbouring English areas, probably coupled with “the presence of a standard British/English English in the professions, law, government, education, media, etc.” (James 2011: 49). This unique combination of linguistic influences has brought about a range of features in the areas of phonetics and phonology, intonation, lexis and morphosyntax; the latter area is discussed in detail in the next section.

3.2 Welsh English morphosyntax

Welsh-language influences are “prominently evident in some areas of Welsh English morphology and syntax” (Penhallurick 2008a: 361). The following section provides some background information on the grammatical features explored in the present

study, with several of these showing clear parallels to the Welsh language. The study focuses on grammatical features that carry the potential to be considered typical of Welsh English, for instance, because they show influences from the Welsh substrate and/or because they are reported to be (more or less) regionally confined to Wales.

3.2.1 The features under investigation

Focus fronting

Focus fronting, or the fronting of a focused constituent in a sentence or clause (cf. Paulasto 2006; Thomas 1997), is an instance of a feature in Welsh English grammar that presumably traces back to Welsh-language influences. The construction is illustrated in examples 1 and 2 from the Radio Wales Corpus:

- (1) *And I, well, it 's because you 're speaking so much to them that uh you tend to say that but uhm 'husband' I refer to him as.*

(66-year-old informant from Pontypridd, South East Wales)

- (2) *First chance I had, yeah, yeah, I, **nineteen** I was and uh had a girlfriend who was sixteen, and I moved out of there to a flat.*

(35-year-old informant from Menai Bridge, North West Wales)

Paulasto (2006: 64) defines focus fronting in English as “a particular type of preposing, where the fronted item is a lexically governed constituent or phrase carrying the informational focus of the sentence”. Thus, as Paulasto goes on, the construction needs to be clearly distinguished from cases where the focus of the sentence or clause is on the initial element, but without affecting the canonical word order (SVO in English, cf. Paulasto 2006: 63). A sentence as in example 3 with the focus on the subject would hence not fall in the category of focus fronting.

- (3) *Our speech is different ...*

(Radio Wales Corpus, 17-year-old informant from Bethesda, North-West Wales)

It is also necessary to distinguish focus fronting from instances in Standard English where sentence elements are placed in front of the subject, but where the end-focus is maintained, as in example 4:

- (4) *Standing there was my brother.*

(Paulasto 2006: 66)

In a focus fronting construction, the fronted constituents can be (subject/object) complements, objects, adverbials and verb(phrase)s. Paulasto (2006: 65) notes that “[f]ronted verbs or verb phrases are more unusual, but not an anomaly.” According to Leech and Svartvik (2013), fronting for the purpose of prominence is not per se given with adverbials in sentence-/clause-initial positions. They state: “We do not normally consider an initial adverbial to be a ‘fronted topic’, because many adverbials can occur fairly freely in front of the subject” (Leech and Svartvik 2013: 213). Adverbials which are “not lexically governed [...] by the verb” (Paulasto 2006: 65) as in example 5 can be repositioned relatively flexibly.

- (5) *In New York, there’s always something to do.*

(Paulasto 2006: 65)

However, as Leech and Svartvik (2013: 213) see it, fronting may apply to some adverbials that are closely linked to the verb and thus are unlikely to occur in front position, as illustrated in examples 6 and 7:

- (6) ... ***in nineteen ninety-one** that was.*

(Radio Wales Corpus, 34-year-old informant from Bangor, North West Wales)

- (7) ***Due to ill health** that was.*

(Paulasto 2006: 162)

As for the origins of focus fronting in Welsh English, it seems likely to Thomas (1997: 78) that the feature “is best accounted for as an instance of interference from Welsh”. The canonical word order in Welsh is verb-subject-object (VSO), and constituents of a sentence or clause can be moved in front of the verb for emphasis (cf. Paulasto 2006; Penhallurick 2008b). As an illustration, consider examples 8 and 9 of identification sentences² in Welsh (cf. Brake 2004: 42).

²Identification sentences, according to King (2015: 176–177), involve “uses of *to be* asking or answering a question beginning *Who is/are...?* or *What is/are...?*, where a simple identification is the only information required”. Examples of identifying questions and answers in English would be: *What is this? – This is a pencil.* (cf. King 2015: 176). An important characteristic of identification sentences is that “both elements or phrases on either side of the verb *to be* refer to the same person or thing” (King 2015: 176). Sentences as the following do thus not fall in the category of identification sentences: *He is outside.*

- (8) *Mae Carl yn athro*
 be.PRS.3SG Carl PRED teacher
 ‘Carl is a teacher’

- (9) *Athro yw Carl*
 teacher be.PRS.3SG Carl
 ‘Carl is a teacher’

According to Brake (2004: 42), sentence 8 is the neutral variant, “simply stating that Carl is a teacher”, while sentence 9 stresses that Carl is a teacher instead of having another profession.

It should be noted that despite focus fronting generally being a rare feature in the British Isles, it is not confined to Wales. It also occurs in Irish English, possibly due to influences from the Irish Gaelic substrate³ (cf. Filppula, Klemola and Paulasto 2008; Mac Mathúna 1990). However, in contrast to Welsh English, the preferred way of emphasising constituents in Irish English is clefting, with focus fronting being “also used to some extent” (Filppula, Klemola and Paulasto 2008: 199).

Standard English uses of cleft and pseudo-cleft constructions are considered as functionally equivalent to (cf. Thomas 1997: 78) or at least as sharing many characteristics with focus fronting (cf. Paulasto 2006: 82). This is shown by examples 10 (focus fronting), 11 (pseudo cleft) and 12 (cleft).

- (10) *Coal they’re getting out, mostly*
 (Thomas 1997: 78)
- (11) *What they’re getting out mostly is coal*
 (Thomas 1997: 78)
- (12) *It’s coal that they’re getting out mostly*

Thus, in order to get an impression of the frequency of focus fronting in relation to (more or less) equivalent Standard English features, clefts and pseudo clefts were also analysed as part of my corpus study (see Chapter 6).

³As in Welsh, the canonical, unmarked word order in Irish Gaelic is VSO (cf. Mac Mathúna 1990: 89). To put emphasis on a constituent, “the portion of the sentence to be stressed is fronted by being placed immediately after the initial copula” (Mac Mathúna 1990: 90).

Invariant tag question *isn't it*

Another Welsh English feature which could be attributed to language contact with Welsh is the use of the tag question *isn't it* irrespective of the verb, person or tense in the preceding clause (cf. Krug 1998; Penhallurick 2008a). Such universal uses of *isn't it* are evidenced by examples 13 and 14:

- (13) *Girls wearing trousers to school, though, was not, was frowned upon for a long time, and only now recently can girls wear trousers, **isn't it**, even to secondary schools.*

(Radio Wales Corpus, 51-year-old informant from Risca, South East Wales)

- (14) *I've heard the word, **isn't it**?*

(Penhallurick 2008a: 361)

In Standard English, the tag question in example 14 would instead be *haven't I?*, with the subject and the operator of the tag question reflecting the subject and the operator of the preceding statement (cf. Quirk et al. 1985: 810). Furthermore, “[i]f the [Standard English] statement is positive, the tag is generally negative, and vice versa” (Quirk et al. 1985: 810).⁴ By contrast to that, invariant tag questions in Welsh English may ignore negation polarity, as shown in example 15 from the British National Corpus (cf. Krug 1998: 154).

- (15) *But it didn't matter as long as you could help **isn't it**.*

According to several scholars, the use of the invariant *isn't it* in Welsh English is very likely to trace back to an interference from Welsh. Penhallurick (2008a: 361) states that the construction “no doubt” arose due to Welsh-language transfer, Williams (2003: 205) considers it a “direct interference phenomenon[on][..] from Welsh”. The model for the *isn't it* tag question in the Welsh language seems to be the “generalised confirmatory interrogative tag ‘ydy fe?’” (Penhallurick 2008a: 361).

Although it seems fairly likely that the *isn't it* construction found its way into Welsh English through Welsh-language influences, the feature also occurs in other parts

⁴Exceptions to that rule include some less common tag questions, such as the one involving “constant positive polarity” (Tottie and Hoffmann 2006: 284; cf. also Quirk et al. 1985: 812). This tag question is used, for example, for the purpose of scolding or sarcasm as in *So that's your game, is it?* (Quirk et al. 1985: 812).

of the British Isles. According to the Electronic World Atlas of Varieties of English (eWAVE; Kortmann and Lunkenheimer 2013), “invariant non-concord tags” can also be found in the South East and South West of England, the Channel Islands, and, with lower degrees of pervasiveness, in Northern England and Scotland.⁵ The South East of England, especially London, is particularly well-known for the use of these tag questions. It is noticeable, though, that examples of the feature provided by different scholars for Southern English English typically include the lexicalised *innit* (/ˈɪnɪt/), while the full form *isn’t it* (/ˈɪzənt ɪt/) is common in descriptions of Welsh English (cf. e.g. Anderwald 2008; Britain 2007; Parry 1999; Penhallurick 2008a).⁶ It should be noted that in the following, unless specified differently,⁷ the notation *isn’t it* refers to the phonemic realisation /ˈɪzənt ɪt/, and *innit* to /ˈɪnɪt/. The question whether that distinction is mirrored in people’s perceptions will be discussed in Chapter 6. Generally, invariant tag questions of the types *isn’t it/innit* seem to be very interesting objects of investigation, since they are on the one hand characterised as typical of Welsh English, on the other hand they are also well-known for Southern England. If my hypothesis holds true, the features’ sociolinguistic salience in Welsh English is connected to their frequencies of use.

Non-standard habitual progressive

Non-standard habitual uses of the progressive are described by Thomas (1997: 77) as the “most characteristically ‘Welsh’ feature” in the area of Welsh English grammar. They are illustrated in the following examples from the Radio Wales Corpus, provided by a 54-year-old speaker from Holyhead in North West Wales:

- (16) *Words like that we **was using**, “you got egg on your chin”, and people would look down and then they ’d see their fly open, but “egg on your chin”...*
- (17) *Know some **are doing** “didey” up here.*

⁵Note that while invariant tags of the types *isn’t it/innit* occur in Channel Islands English, the invariant tag question *eh* is particularly characteristic of this variety of English (cf. Rosen 2012: 102). Similarly, invariant tags in Scottish English typically include *eh* (cf. Miller 2008).

⁶However, there seems to be a current trend of invariant *innit* tag questions spreading across the British Isles (cf. Krug 1998). According to Penhallurick (2008a: 361), this may have “a reinforcing effect on Welsh English”.

⁷Of course there are other ways of realising the tag question phonemically, for example as *in’t it* (/ɪnt ɪt/) or *ain’t it* (/ˈeɪnt ɪt/) (cf. Krug 1998: 145). Whenever discussions of invariant tag questions in this work concern such forms, the respective phonemic realisations are provided.

Didey in the second example refers to a Northern Welsh English term for ‘grandfather’ and *doing* to ‘using’. A speaker of Standard English might rather say: *I know that some use “didey” up here*. As can be seen in the examples, the progressive can be used in Welsh English both when referring to past and to present events that took or take place repeatedly. It should be noted that there are habitual progressives in Standard English, too, but they differ from the Welsh uses in their exact functions. Progressives expressing habituality in Standard English may be employed, for instance, when referring to events that happen(ed) within a particular limited time period, as in examples 18 and 19 (cf. Paulasto 2006; Pitkänen 2003).

- (18) *At the time she **was having** regular singing lessons*
(Quirk et al. 1985: 199)

- (19) ***We’re going** to the opera a lot these days*
(Paulasto 2006: 101)

By contrast, non-standard habitual progressives in Welsh English are used in contexts without a clear temporal limitation (cf. Pitkänen 2003: 113). In Standard English, habitual progressives also occur in constructions with an “emotive impact” (Paulasto 2006: 101), denoting somebody’s discontent with someone else’s repeated behaviour:

- (20) *She’s always **buying** far more vegetables than they can possibly eat*
(Paulasto 2006: 101)

Moreover, stance verbs such as *live*, *lie*, *sit* and *stand* can either occur in the simple or in the progressive form in Standard English (cf. Paulasto 2006: 101). Thus, sentences like *I’m living in Wales* do not fall in the category of non-standard habitual progressives. Neither do instances of a combination of habitual and progressive meaning as in example 21:

- (21) *Whenever I see her, she’s **working** in the garden*
(Paulasto 2006: 101)

Every time the observer sees “her”, her work is ongoing, hence the use of the progressive.

It is likely that the non-standard habitual progressive in Welsh English arose from language contact with Welsh (cf. Paulasto 2006; Penhallurick 2008a; Thomas 1997). In the Welsh language, the present simple and the present progressive can be expressed with the same construction, and sentences as in example 22⁸ can, but do not necessarily, indicate habituality (cf. Brake 2004; Thomas 1997):

- (22) *Mae hi 'n mynd i'r farchnad*
 be.PRS.3SG she ASP go to the market
 'She goes/is going to the market'

Non-standard uses of the progressive can also be found in other varieties of English. However, according to Paulasto (2006: 107), in most cases this means that the progressive is “used with stative verbs in a wider range of contexts than in mainstream English”.⁹ Non-standard uses of the progressive with habitual verbs are reported by Paulasto (2006: 197) to occur in Irish English, Manx English and Hebridean English. eWAVE adds Orkney and Shetland to the list, with rare occurrences of the feature (cf. Kortmann and Lunkenheimer 2013).

In Standard English, functionally equivalent constructions to past tense uses of the non-standard habitual progressive would be *used to*, *would*, simple past forms and arguably *tended to* (which “can be considered habitual in some senses” according to Paulasto 2006: 108). In the present tense, habituality can be expressed in Standard English with present simple forms of verbs.

Inverted word order in indirect questions

Indirect questions in Welsh English can maintain the word order of direct questions, with the verb coming first in the subordinate clause (cf. Penhallurick 2008a; Thomas 1997):

- (23) *Well, we had a lorry driver came down to us and asked us once **could we** tell him where Cylinderwen was.*
 (Radio Wales Corpus, 75-year-old informant from Milford Haven, South West Wales)

⁸In this example, ASP stands for aspectual marker (cf. Brake 2004: 21).

⁹An example of a stative verb in a progressive construction would be: *He's having all kinds of names* (Pitkänen 2003).

- (24) ... so we 're unloading the the sawdust on this farm and uh the chap asked me where **did I** come from and I said like, I 'm from Wales ...

(Radio Wales Corpus, 36-year-old informant from Builth Wells, Mid Wales)

As shown in example 24, the inversion of verb and subject can occur in the context of post-predicate “wh-clauses” (Biber, Conrad and Leech 2005: 324), involving relative pronouns such as *what* and *where* but also *how*. The inversion can furthermore affect indirect questions containing the conjunctions *if* and *whether* in Standard English, resulting in the loss of the conjunction, as can be seen in example 23 (cf. Thomas 1997: 79). Such indirect questions as in (23) can also be described as “Yes/No-questions” (Meriläinen and Paulasto 2014).

Similar to the constructions elaborated on above, the inverted word order in indirect questions is likely to have evolved due to Welsh structural influences. According to Penhallurick (2008a: 363), the order of the verb and the immediately following constituent in Welsh is “identical in direct questions and their equivalent indirect ones”. Thomas (1997: 79) adds that the omission of the *if/whether* conjunction may also be facilitated “by the Welsh rule of eliding the corresponding conjunction (*a/os*) in similar environments in the vernacular” (emphasis in original).

Despite the assumed influence from Welsh, the feature is not restricted to Wales. Filpula (1999) reports the existence of such indirect questions in Irish English, Scottish English, Tyneside English and Hebridean English, suggesting general Celtic influences at play (cf. also Penhallurick 2008a: 363). eWAVE additionally mentions the feature’s occurrence on the Isle of Man and in the South West of England (cf. Kortmann and Lunkenheimer 2013); in both regions, Celtic structural influences might also have played a role.

Zero past tense of regular verbs and *that*-clause replacing infinitival subclause

The last two constructions under investigation in the present study – the zero past tense of regular verbs and the *that*-clause replacing an infinitival subclause – differ from the other features discussed above in that they have not yet been investigated in the context of Welsh English. What made them interesting for my study was that they are classified as “rarissima” in eWAVE (cf. Kortmann 2012: 695). This means the

features were attested in only one variety of English in the British Isles, namely in Welsh English, thus it can be assumed that the rarissima carry strong local signs (cf. Kortmann 2012: 659).

The **zero past tense of regular verbs**, meaning that the *-ed* ending is not articulated, is depicted in examples 25 and 26 from the Radio Wales Corpus:

- (25) ... *I thought they were getting something to to pump water out of the house, and then I was amazed when you **talk** about daps.*

(54-year-old informant from Holyhead, North West Wales)

- (26) *And we **try**¹⁰ everything that was going there, piano solo, the singing solo, the uh duet solo, the the modulator, everything ...*

(66-year-old informant from North Wales)

Even though the construction seems to be very rare in the British Isles, eWAVE reports that it occurs pervasively in 25 regional varieties of English and English-based creoles and pidgins in the world, for example in the Caribbean and South East Asia (cf. Kortmann and Lunkenheimer 2013). However, it is noticeable that none of these regional varieties are traditional L1-varieties, but creoles (e.g. Jamaican Creole), pidgins (e.g. Cameroon Pidgin), indigenised L2-varieties (e.g. Hong Kong English) and high-contact L1-varieties (e.g. Colloquial Singapore English – Singlish). In eWAVE, according to Kortmann and Wolk (2012: 926), zero past tense forms of regular verbs are among “the top diagnostic features for high-contact L1-varieties” of English, since there is a big difference in how pervasive the feature is in the high-contact (such as Welsh English) and in the traditional L1-varieties (such as Scottish English or English dialects in the South East of England).

that-clauses replacing infinitival subclauses can be found in sentences as the following:

- (27) *I wanted **that** I should get leave*

(Mesthrie 2008: 629)

¹⁰According to (Biber 1999: 392), *try* (*-tried-tried*) falls in the category of regular verbs.

In such constructions, *to* plus (infinitival) verb is substituted by *that* plus substantive plus verb (cf. Mesthrie 2008: 629). Mahboob (2012: 539), who discusses the feature in the context of Pakistani English, mentions the following other verbs apart from *want* that can be used in combination with *that*: *aim*, *think*, *hesitate*, *refrain* and *fail*.

According to eWAVE, the feature seems to be rare not only in the British Isles, but worldwide (cf. Kortmann and Lunkenheimer 2013). Pervasive uses of the *that* construction are reported for none (!) of the 74 varieties of English and English-based creoles and pidgins discussed in the atlas. The feature is considered to be “neither pervasive nor extremely rare” (Kortmann and Lunkenheimer 2012: 5) in eight regions across the globe. Again, none of these regions feature traditional L1-varieties of English but creoles (e.g. Saramaccan), pidgins (e.g. Nigerian Pidgin), indigenised L2-varieties (e.g. Indian English) and high-contact L1-varieties (e.g. Rural African American vernacular English).

3.2.2 Research on Welsh English morphosyntax

The following section gives an overview of previous research on Welsh English morphosyntax with a special focus on the features discussed above, both in terms of their production or use in Welsh English and their perception. I do not claim that the list of studies presented here is exhaustive. Still, my impression is that Welsh English grammar has as yet received attention only in a small number of research projects.

Production

The Survey of Anglo-Welsh Dialects (SAWD) is described by Penhallurick (2008b: 107) as the “most comprehensive collection of Welsh English data”. The material, stored at the University of Wales, Swansea, was collected under the aegis of David Parry in 90 rural Welsh areas between 1960 and 1982 (cf. Parry 1999: 1). An urban follow-up to SAWD (SAWD 2) was conducted between 1985 and 1987 by Robert Penhallurick in four areas of Wales (cf. Penhallurick 2008b: 107). The methods for data collection were based on those applied in the Survey of English Dialects (SED), with the rural SAWD aiming to “provide material for Wales that is directly comparable with that obtained in England” (Parry 1999: 1). SAWD comprises recorded interviews with (1) informants’ responses to a questionnaire, designed to gather (mainly phonological and lexical) lin-

guistic information, and (2) “[i]ncidental material” (cf. Parry 1999: 1) from general conversation, not directly elicited from the questionnaire (cf. Parry 1999; Pitkänen 2003). Most of the informants in the rural data were non-mobile old rural males (NORMs), which means they had never moved far from the area where they grew up, they were aged older than sixty and they had received no formal education beyond the age of fifteen (cf. Parry 1999: 1). A detailed account of phonological, lexical and grammatical features found in the rural SAWD data (without frequency counts) is given by Parry (1999). He records, for instance, occurrences of the non-standard habitual progressive, the invariant tag question *isn’t it* (without phonological specifications), the inverted word order in indirect questions and focus fronting, or, as he calls it, “sentence-initial emphasis” (Parry 1999: 115, 119–120).

Despite SAWD being beyond doubt a fruitful source for many research questions, it was not used for the analyses in the present study. One reason was that the questionnaire, even though it contained some grammatical items, focused on gathering phonological and lexical information, thus “[s]yntax remains in the background” (Pitkänen 2003: 117). Moreover, the formal interview style following the questionnaire offered scant room for spontaneous conversation. Thus, frequency counts for grammatical features in these data may be strongly tied to the structure of the interviews. Furthermore, the rural SAWD informants are mainly NORMs interviewed from the 1960s to 1980s, and in the context of my study I was interested in present-day English in Wales spoken by people from different age groups.

Penhallurick (1996) uses rural SAWD recordings from the counties Clwyd and Gwynedd in North Wales, collected from 1973 to 1982, for his qualitative analyses of different progressive *be* constructions, such as the non-standard habitual progressive, but also, for example, state past forms, as in *it was holding* (Penhallurick 1996: 335). He also contrasts the use of past tense non-standard habitual progressives with habitual *do* phrases (as in *I did go to the market every week*), the latter construction possibly tracing back to influences from East Somerset English. In contrast to South Wales, where both the progressive and the *do* forms were recorded by SAWD, the *do* phrases did not occur in Penhallurick’s analysis of the North Wales data. There were many occurrences of the non-standard habitual progressive, though, mostly produced by speakers

who had Welsh as their first language.

SAWD is also used by Thomas (1984, 1997) for his descriptive overviews of Welsh English grammar. Additionally, he presents examples of features from his own knowledge, but does not name a specific dataset. Thomas (1984) mentions the non-standard habitual progressive, the invariant tag *isn't it* (without phonological specifications) and focus fronting; the inverted word order in indirect questions is described in Thomas (1997).

Williams (2000) explores focus fronting, or, as he calls it, “predicate fronting”, both regarding uses and perceptions of the feature (the latter aspect being discussed in detail below). The database consulted for uses of focus fronting consists of short stories and a fictional conversation (2,926 words in total) taken from *Talk Tidy* and *More Talk Tidy* by John Edwards (1985, 1986). Edwards’s works aim to portray “Wenglish” as spoken in the South Wales Valleys, and to restore positive attitudes towards it (cf. Williams 2000: 218). Williams’s primarily qualitative approach identifies different types of focus fronting. For example, he distinguishes frontings with relatively little new information in the fronted element from instances where the fronted element “is textually and situationally new” (Williams 2000: 227). Apart from that, he records that in his dataset, focus fronting occurs 16.74 times per 1,000 words, which “will appear out of all proportion even to those who are familiar with Hibernian English” (Williams 2000: 212). However, the validity of that number is restricted since (1) the database is very small and (2) it is unclear whether the data reflect general practices in Welsh English or just those of John Edwards.

Another study by Williams (2003) deals with invariant tag questions, such as *isn't it*, *innit* or *is it*, in Welsh English. The analysis is based on the Ceri George Corpus with interviews from the Southern Welsh Rhondda Valleys, recorded in the early 1980s. The handwritten transcripts are stored at the University of Wales, Swansea (cf. Paulasto 2006: 153). According to an estimate by Williams, the corpus comprises around 101,915 words. Williams’s study is qualitative in nature, outlining different degrees of invariability in the Welsh informants’ uses of tag questions, ranging from tags relatively close to the standard as in *Shouting now aren't I* (Williams 2003: 208) to invariant tags as in *Girl I was, isn't it* (Williams 2003: 208). The study also, again, takes a look at

focus fronting, counting 26.7 occurrences of the feature per 10,000 words. Once more, Williams's data suggest a high pervasiveness of focus fronting in Welsh English. The figure can, however, only be an estimation since no exact word count of the corpus is provided and it is not clear whether the handwritten transcripts cover everything that was said in the interviews.

Paulasto (2006) also uses the Ceri George Corpus in her study, in combination with five other corpora, which enables her to carry out a systematic quantitative analysis of grammatical features of Welsh English on a solid databasis. The features under investigation are focus fronting and non-standard uses of the progressive form (including habitual progressives). Paulasto consults (parts of) the following corpora: SAWD, SAWD 2, the Ceri George Corpus, the Llandybïe Corpus (257,500 words, compiled by herself between 1995 and 2000), the North Wales Corpus (120,000 words, compiled by herself in 2000) and, as a reference corpus representing traditional English English dialects, the Survey of English dialects (SED, interviews mainly from the 1950s and 1960s). Focus fronting occurred in the North Wales Corpus (NWC) 3.17 times and in the Llandybïe Corpus (LC) 4.66 times per 10,000 words (interestingly, the latter figure is very similar to my findings from the Radio Wales Corpus, with focus fronting occurring 4.58 times per 10,000 words). Non-standard habitual progressives appeared 1.33 times per 10,000 words in the NWC and 2.6 times per 10,000 words in the LC (again very much in line with the results from the Radio Wales Corpus, where the feature occurred 2.64 times per 10,000 words). These figures, however, may be subject to change in the future. By means of real- and apparent-time studies, Paulasto witnesses ongoing processes of dialect levelling with both features being used less frequently by the younger generations and in the more recent corpus data. Meriläinen and Paulasto (2014) use the Llandybïe Corpus to explore the inverted word order in indirect questions, or as they call it, "embedded inversion". In these data, the feature occurs 0.62 times per 10,000 words (and is only slightly more frequent than in the Radio Wales Corpus with 0.41 occurrences per 10,000 words). Both in the Llandybïe Corpus and in my data, then, the inverted word order in indirect questions is clearly less frequent than focus fronting and non-standard habitual progressives.

The Freiburg Corpus of English Dialects (FRED), compiled under the supervision of Bernd Kortmann, contains transcribed interviews from nine dialect areas in England,

Scotland and Wales, totalling approximately 2.5 million words (cf. Hernández 2006: 1). The transcriptions from Wales (Glamorgan in the South and Denbighshire in the North) encompass 88,755 words (cf. Hernández 2006: 4). FRED is a valuable source for explorations of morphological and syntactic features across Great Britain. However, it was not employed for the frequency-based analyses in the present study, the reasons being similar to those for not working with SAWD: the FRED interviews from Wales are mainly from the 1970s and most of the informants are NORMs. Nevertheless, I carried out a pilot study with FRED and determined the token frequencies of the non-standard features presented earlier in this chapter: (1) focus fronting, (2) invariant tag question *isn't it/innit*, (3) non-standard habitual progressive, (4) inverted word order in indirect questions, (5) zero past tense of regular verbs and (6) *that*-clause replacing infinitival *to*-clause. The features' relative frequencies, normalised per 10,000 words, are presented in Table 1. As can be seen, focus fronting and habitual progressives are considerably more frequent in the FRED data than the other features under investigation.

	Frequency per 10,000 words
Non-standard habitual progressive	15.55
Focus fronting	10.37
Inverted word order in indirect questions	1.13
Invariant tag question <i>isn't it/innit</i>	0.9
Zero past tense of regular verbs	0.23
<i>that</i> -clause instead of infinitival subclause	0

Table 1: Token frequencies per 10,000 words of Welsh English features in FRED

The findings for the non-standard habitual progressive have to be taken with a pinch of salt, though. (Past) habituality generally seems to be an integral part of these oral history interviews. My pilot study found the standard habitual markers *used to*, *would* and simple past/present to be very frequent in the data, too, occurring 146.92 times per 10,000 words. By contrast to that, as discussed in detail in Section 6.2, these standard habitual markers occurred 33.44 times per 10,000 words in the Radio Wales Corpus, while the non-standard habitual progressive appeared 2.64 times per 10,000 words.¹¹

¹¹These analyses show that corpus-based usage frequencies can be tied to particularities of individual corpora. In analysing the frequencies of linguistic features in a corpus, it is therefore highly advisable

Perception

Grammar has so far barely been touched upon in studies exploring perceptions of and attitudes towards Welsh English. One reason might be that different tools need to be employed when eliciting conscious perceptions and/or stereotypes of Welsh English grammar in comparison with pronunciation or lexis (for a detailed discussion of this issue, see also Chapter 5). To give an example, the folk linguistic survey by Williams, Garrett and Coupland (1996) theoretically offered the opportunity to name all kinds of linguistic features that came to mind when thinking about particular regions of Wales. Based on Preston's (e.g. 1986) draw-a-map-task, test subjects (129 teachers living in different parts of Wales) had to label dialects areas in Wales on a map and "to construct evaluative profiles for their labeled zones" (Williams, Garrett and Coupland 1996: 171). The subjects' comments thus involved specific linguistic features seen as characteristic of a particular region plus affective statements and comments dealing with status and norms, geo-social belonging or "rural versus urban" (Williams, Garrett and Coupland 1996: 186). The linguistic features named by the authors are phonetic-phonological or lexical in nature, for instance, referring to the fronted and raised long [a:] in Cardiff or to the "ai"/"aye" in Bangor. Except for one subject mentioning a "Welsh order to words", the authors do not discuss any grammatical features that arose in the study. This could mean that it is harder for informants to name, without guidance, regional grammatical features than phonological or lexical ones. The awareness of particular morphosyntactic features may need to be elicited in a more structured manner, for example, by presenting people with test sentences and asking them where, i.e. in which regional dialect or variety, they think such sentences are used (cf. Kerswill and Williams 2000). This was done in the present study (see Chapter 5). Another explanation for the scarcity of grammatical features in the study by Williams, Garrett and Coupland (1996) could be that there are no grammatical features of Welsh English that are particularly salient, or that stick out more than others. However, the present study shows that this is not the case (focus fronting and the invariant tag question *isn't it* were significantly more salient than the other features under investigation, see Chapter 6).

to compare them (1) to frequencies of other (e.g. Standard English) features and (2) to frequencies of the same features in other corpora to add more reliability to the data. Such analyses were carried out in the context of the present study (see Section 6.2).

Williams's (2003) survey on the perception of focus fronting makes use of a questionnaire with test sentences. The sentences were taken from *Talk Tidy* and *More Talk Tidy* (see above), most of them involving focus fronting constructions. The 44 subjects (mainly from Wales, but also from England) were asked if they would use the constructions themselves. Against Williams's expectations, "[i]nstead of the expected confirmation of the structures given in the examples, there was a categorical rejection" (Williams 2003: 223). Interestingly, those subjects who had Welsh as their first language had the lowest acceptability rates. Williams hypothesises that some informants might have intended to answer in line with normative criteria instead of reflecting actual personal use, possibly connected to the fact that they had to "estimate their *own* language use, not to comment on whether they expected to hear these expressions being used by other people" (emphasis in original; Williams 2003: 225). That is one reason why, in my salience survey, I asked people in which regions particular features occurred instead of letting them reflect on their own language practices (for a more detailed discussion of the methodology applied, see Chapter 5). Generally, the results of Williams's study are only restrictedly conclusive, since only one feature – focus fronting – was investigated. The validity of the acceptability ratings of focus fronting would have been increased by comparing them with ratings of other grammatical features.

All in all, it seems that, so far, not much is known about the perception of different grammatical features of Welsh English, for instance, with regard to the question whether some features are perceived more consciously than others. Moreover, the studies presented above mainly explored judgements of people from inside Wales, while it remains rather unclear how outsiders perceive Welsh English grammar.¹² In the area of speech production, only few works have as yet investigated morphosyntactic features of Welsh English from a quantitative point of view (most of them with interview data from thirty years ago or older). And those works that did, did not systematically relate the frequency results to findings from perception tests. By doing so in this work, I aim to establish a link between research on uses and on perceptions of Welsh English, contributing to both of these as yet underresearched fields of study.

¹²Except for Williams's (2003) finding that some subjects from Yorkshire indicated using focus fronting constructions in their speech, too.

3.2.3 Beyond morphosyntax: Characteristics of Welsh English phonology, intonation and lexis

My salience questionnaire for the informants from London included the question “What are the first five things about English in Wales that you can think of?” Although there were references to grammar, the largest part of the comments, apart from general descriptions (such as “friendly” or “funny”), referred to phonetic or phonological aspects of Welsh English, followed by statements on intonation. There were also a lot of references to Welsh (English) lexical items. Thus, even though this work focuses on Welsh English grammar, the following section will briefly outline some phonological, intonational and lexical features, since it can be assumed that such features also (and maybe even more so) shape folk linguistic perceptions of Welsh English (cf. also Williams, Garrett and Coupland 1996).

Phonetics and phonology

“[T]wo characteristic sounds of Welsh English” (Penhallurick 2008b: 112) are the long monophthongs [e:] and [o:], replacing the RP diphthongs [eɪ] and [əʊ] (cf. Parry 1999; Roller 2011). They can occur in words of the FACE type in Wells’s (1986) lexical set (e.g. *make, great, break*) and of the GOAT type (e.g. *spoke, road, coal*) in large parts of North and South Wales, while the diphthongal realisations are preferred in some South-Eastern border or urban areas, such as Cardiff and Newport (cf. Collins and Mees 1999; Penhallurick 2008b). In the North of Wales, the monophthongs can additionally occur in words of the STAY type (e.g. *drain, weigh, clay*) and of the SNOW type (e.g. *shoulder, cold, snow*), whereas the diphthongs (with frequent occurrences of [ɔʊ] instead of [əʊ]) are preferably used in such words in South Wales (cf. Penhallurick 2007: 157). These uses of [e:] and [o:] in Welsh English may trace back to Welsh, which possesses the monophthongs, but lacks the diphthongs [eɪ] and [əʊ]/[ɔʊ] (cf. Penhallurick 2008b: 113).

Furthermore, Welsh English features the alveolar pronunciation of /r/ as tap [ɾ] or trill [r] (cf. Penhallurick 2008b: 118). These realisations are common throughout Wales, particularly in the traditionally Welsh-speaking heartlands, but also, for example, in Cardiff (cf. Collins and Mees 1999; Wells 1986). However, in some border areas plus Southern Pembrokeshire and the Gower, the approximant [ɹ] is dominant (cf. Pen-

hallurick 2008b: 118). Alveolar [ɾ] and [r] are the two phonemes of /r/ in the Welsh language; additionally, in Northern Welsh and Welsh English, “uvular rolled [ʀ] or uvular fricative [ʁ] are sometimes to be heard” (Penhallurick 2008b: 118).

Other phonetic-phonological characteristics of Welsh English include the disyllabic realisation of words such as *fire* /'faɪjə/ or *shower* /'ʃaʊwə/ through the insertion of /j/ or /w/ in the triphthong. In RP, these words would be realised monosyllabically as /faɪə/ and /ʃaʊə/ (cf. Penhallurick 2008b; Roller 2011; Thomas 1984; Trudgill and Hannah 2002). Moreover, in Southern Welsh English, there is a tendency to use the clear [l] in all positions,¹³ leading to pronunciations such as *bell* [bel], as opposed to the dark [ɫ] in all positions in Northern Welsh English accents, resulting in pronunciations as *belly* ['beɫi] (cf. Penhallurick 2008b; Roller 2011).

Intonation

In my salience questionnaire, Welsh English was described by many subjects from London as “sing-song”, “lilting” or “musical”. Indeed, this characteristic, mainly associated with the Southern Welsh Valleys, seems to belong to “[p]opular English views about Welsh accents” (Wells 1986: 392). The “sing-song” impression of Welsh English accents is connected to a (presumably Welsh-influenced) high degree of pitch movement involving a lot of rise-falls (cf. Connolly 1990; Podhovnik 2008). What is more, while the tone movement typically takes place in the tonic syllable in RP, it also frequently occurs in pre- and post-tonic syllables in Welsh English (cf. Tench 1990: 140). A further point that was commented on by some Londoners in my questionnaire is that in Welsh English “everything sounds like a question”. Podhovnik’s (2008) study offers empirical evidence for that impression, finding that in her data from Neath, South Wales, sentence-final rises are not only used in questions, but also in statements. The Welsh language once more offers an explanation, generally using rises for adding emphasis to prominent syllables (cf. Podhovnik 2008: 387).

¹³In RP, clear and dark /l/ occur in complementary distribution, “with [l] before a vowel, and [ɫ] before a consonant or pause” (Penhallurick 2008b: 118).

Lexis

Welsh English lexis features some loanwords from Welsh and some English words used with a different meaning than in Standard English (cf. Trudgill and Hannah 2002: 33–34). The loanwords, for instance, include the terms of endearment *del* and *bach* (also commented on by Londoners in my salience survey) and *eisteddfod*, referring to a popular arts festival in Wales (cf. Thomas 1984: 193). Among the English words with different meanings in Welsh English is *tidy*, meaning ‘good, nice’, as in *A tidy car he’s got* (cf. Trudgill and Hannah 2002: 34), and employed as a vague quantifier, as in *A tidy bit of money* (cf. Thomas 1984: 194). The use of *tidy* in the titles of the above-discussed works *Talk Tidy* and *More Talk Tidy* by John Edwards could be an indicator of the word’s stereotypical status in Welsh English. *Tidy* was also mentioned by my London subjects in the salience questionnaire and discussed by some informants in the Radio Wales Corpus. Further lexical items in Welsh English, for example, include *delight* in the sense of ‘interest’ (*A delight in movies*) and *rise*, meaning ‘get, buy’ (*I will rise the food*) (cf. Edwards 1985; Trudgill and Hannah 2002).

3.3 English in London

English in London functions as the reference variety for Welsh English in the present study. The salience survey included subjects from London to investigate how outsiders perceive Welsh English grammar, and how these perceptions relate to the linguistic practices of the outsiders. One reason for choosing London English as the reference variety thus was that it differs structurally from Welsh English. It can be assumed that London English has not hosted direct transfer from the Welsh language (whereas Irish and Scottish English, for example, share some Celtic-induced features with Welsh English). Moreover, London does not border Wales, thus Welsh English might have fewer grammatical overlaps with London English than, for instance, with English dialects in the South West of England, the Western Midlands or Merseyside. Still, London is geographically not too far away from Wales, hence I hypothesise that (average) Londoners are to some extent familiar with (salient) grammatical features of Welsh English. The following sections outline English dialects in London, both in general terms and regarding morphosyntax, here referring back to Welsh English.

3.3.1 General information

At a first glance, descriptions of English in London seem contradictory. On the one hand, London is closely linked to Standard English (at least in many people's heads, cf. Anderwald 2008), while, on the other hand, the city is described as a multicultural melting pot, drawing on non-standard linguistic influences from multiple resources. Taking a closer look, it becomes clear that both observations apply to London. "Although Standard English (StE) linguistically had its source in the dialect of the East Midlands, London (the seat of the court, of Chancery, of the printing presses) is the place where the standard evolved" (Anderwald 2008: 440; cf. also Wright 1981: 13). London's status as a politically and economically powerful metropolis has, in turn, for centuries attracted large numbers of immigrants, both from the British Isles and beyond, contributing to London's multidialectal and multilingual texture (cf. Cheshire et al. 2011). The following main accents and dialects can be found in present-day London English, although many others may be used on a smaller scale, too: Received Pronunciation (RP), Standard Southern British (SSB), Estuary English, Cockney and Multicultural London English (MLE).

RP is defined by Mair (2008: 254) as "the national pronunciation standard of British English which historically developed in the Southeast of England". It is a supraregional accent indicating the affiliation to an educated higher social class (cf. Herbst, Stoll and Westermayr 1991: 207). Although spoken nowadays only by about three to five per cent of "educated native speakers" (Kortmann 2005: 22), it still serves as a reference accent both in UK education and in foreign language teaching worldwide (cf. Kortmann 2005; Mair 2008).¹⁴ According to Harrington (2006: 441), there has been a change in recent decades "from the more aristocratic form of RP, sometimes known as U- or upper-crust RP (Wells, 1982), towards mainstream RP or Standard Southern British (SSB) that is characteristic of the professional middle classes". *SSB*, among other things influenced by Cockney, is in many respects similar to RP, and speakers of both accents typically employ Standard English grammar. However, *SSB* is "associated with speakers who are younger and lower in the social hierarchy" (Harrington, Palethorpe

¹⁴RP is not the only reference accent used in education and teaching English around the globe. Another example is General American (GA) (cf. Kortmann 2005; Mair 2008).

and Watson 2000: 927).

As little as SSB can be clearly distinguished from RP, it is not possible to sharply separate these two concepts from *Estuary English*. Introduced by Rosewarne (1984), the term “Estuary English” refers to “a mixture of non-regional and local Southeastern English pronunciation and intonation. If one imagines a continuum with RP and London speech at either end, ‘Estuary English’ speakers are to be found grouped in the middle ground” (Rosewarne 1984: 29). Estuary English reflects some current trends, for instance, the spread of London working-class features into RP and the retention of regional South Eastern dialect variants by speakers otherwise expected to turn into RP speakers (cf. Altendorf and Watt 2008: 202). While sharing some non-standard phonetic features with Cockney, such as T-glottaling, Estuary English is “associated with standard grammar and usage” (Wells 1997: 2). The term “Estuary English” has given rise to controversial discussions, for instance, because it suggests that the speakers are “confined to the banks of the Thames Estuary” (Trudgill 1999: 80), while they are in fact found across the South East of England. What is more, the label may imply a new variety of English, but, as Wells (1997) sees it, it rather is the resumption of a development going on for at least 500 years, namely “the tendency for features of popular London speech to spread out geographically (to other parts of the country) and socially (to higher social classes)” (cf. also Altendorf and Watt 2008: 203).

The term *Cockney* dates back to Middle English *cokeney*, which referred to a misshaped egg and probably was “a synonym for anything odd” (Wright 1981: 11). “Cockney” was assumedly used first by villagers from around London in order to describe and ridicule people who had been born and had lived in London all their lives, “traditionally within the sound of Bow Bells – in other words, within about a quarter of a mile of the church of St Mary-le-Bow in Cheapside in east central London” (Wright 1981: 11). Today, Cockney, well-known for its rhyming slang¹⁵, more generally refers to “traditional working-class London English” (Hughes, Trudgill and Watt 2012: 75). The attribute “traditional” is worth mentioning here since recent publications see Cockney moving from London to Hertfordshire and Essex. According to Paul Kerswill in an article in the Daily Mail (2010), Cockney will have vanished within the next generation in

¹⁵In Cockney rhyming slang, words are substituted by rhymes. For example, *apples and pairs* means ‘stairs’ and *mince pie* denotes ‘eye’ (cf. Samuel 2012; Wright 1981).

much of London's East End. "People in their 40s will be the last generation to speak it and it will be gone within 30 years", a reason being that Londoners from the East End have been moving to Essex for decades, bringing their traditional dialect with them (cf. Daily Mail 2010). In London, Cockney is gradually turning itself into *Multicultural London English* (MLE) (cf. Kerswill 2014). This multiethnolect, arising in multicultural East End neighbourhoods in the 1980s, is spoken mainly by young working-class multicultural Londoners (cf. Cheshire et al. 2011; Hughes, Trudgill and Watt 2012; Kerswill 2014). Labelled by the media as "Jafaikan" (related to its perception as fake Jamaican), MLE is a "melting-pot mixture of all those people living here who learnt English as a second language" (Kerswill in Daily Mail 2010). Due to its high degree of variability, Cheshire et al. (2011) and Kerswill (2014) avoid labelling MLE as a "variety" but refer to it as a pool of features, involving influences from second-language English, Asian, Caribbean and African English, London Jamaican Creole, Cockney, Standard English and languages other than English. These diverse influences are manifested not only in pronunciation¹⁶, but also, for example, in lexical and grammatical patterns (cf. Kerswill 2014).

3.3.2 London English grammar

Characteristic features of English accents in the South East of England, and in London specifically, have been described extensively in the literature. By contrast to that, so far there is little evidence for a distinctive South Eastern dialect grammar (cf. Anderwald 2008; Edwards 1993). Surely there are some typical features, as the invariant tag question *innit*, which emerged in London. But most morphosyntactic constructions in the South East described in previous works constitute general features of non-standard English and are likely to occur in a variety of other dialects as well (cf. McArthur 1998). Note however that this lack of distinctiveness may also be connected to the small number of studies that have dealt with South Eastern grammar in detail so far (cf. Anderwald 2008: 440). The following paragraphs outline some exemplary morphosyntactic features used in the South East (among other regions).

¹⁶ An example of a phonological characteristic of MLE is the reintroduction of /h/ in stressed pronouns and lexical words, while London is a traditionally H-dropping area (cf. Kerswill 2014; Mair 2008).

In expressing the past tense, speakers of South-Eastern dialects, such as Cockney, tend to simplify irregular verb paradigms (cf. Anderwald 2008; Edwards 1993; Wright 1981). Three-way paradigms of strong verbs may become two-verb paradigms (for example, *do – done – done*), and two-verb paradigms may be reduced to just one verb (e.g. *run – run – run*), illustrated in examples (28) and (29) (cf. Anderwald 2008: 445). What (29) also shows is that irregular verbs in Standard English may be treated as regular verbs in South Eastern dialects (*runned*) by using the *-ed* ending for the past tense and the past participle. Another instance of such a non-standard *-ed* past tense form is given in (30).

- (28) ... *he went “oh whatever” . he and then he **done** it again I got up and I did smack him one ...*

(Linguistic Innovators corpus, 16-year-old informant from Havering)

- (29) *I **runned** back and my boys locked me outside because . they were scared as much as I was so I **run** I **run** I **run** .*

(Linguistic Innovators corpus, 18-year-old informant from Hackney)

- (30) *and then I interrupted them . I **comed** this year*

(Linguistic Innovators corpus, 17-year-old informant from Hackney)

Double negation (also called negative concord) generally occurs across the British Isles (cf. Kortmann and Lunkenheimer 2013). Anderwald (2008: 453), however, suggests “a robust quantitative difference between the North and the South”, with data from the FRED corpus pointing to more frequent uses in the South. The feature is common in South Eastern dialects, for example in Cockney grammar (cf. McArthur 1998). Speakers of London dialects may thus employ sentences as the following, where the negator *not* is paired with other negative forms (cf. Anderwald 2008: 454):

- (31) *. like they started on them and like .. they just **didn’t** do **nothing** my mates they just like just were trying to walk away .*

(Linguistic Innovators corpus, informant from Havering)

Variable uses of past tense BE have been recorded for Multicultural London English and South Eastern English in general (cf. Cheshire et al. 2011: 181). Plural (pro)nouns are combined with the singular verb form *was* and vice versa. In FRED (South East), according to Anderwald (2008: 448), the pairing of plural pronouns (*we*, *you*, *they*) and

the singular BE form is “almost categorical”, occurring in 80 per cent of cases. Combinations of singular subjects and plural *were* occur in FRED as well, but less frequently.

(32) ... *er they got applicants but they **wasn't** suitable ...*

(Linguistic Innovators corpus, 67-year-old informant from Hackney)

(33) *yeah at your cousin I **weren't** chucking it at you though*

(Linguistic Innovators corpus, 16-year-old informant from Havering)

A recent feature associated with young MLE speakers in London is the quotative “*this is + speaker*” (emphasis in original; Cheshire et al. 2011: 172) which is used with a similar function as *I'm like*, illustrated in example (34).

(34) *this is me “I'm from Hackney”*

(emphasis added; Cheshire et al. 2011: 172)

Cheshire et al. (2011) make use of two datasets with interviews from London, one of them being the Linguistic Innovators corpus. The new quotative occurs only in interviews with young speakers from Hackney, but not in the Havering data. It exhibits a relatively low frequency of occurrence as compared to other quotatives (e.g. *say, go*), but its use by a range of adolescents and even UK comedians and actors suggests that it is more than a transitory phenomenon (cf. Cheshire et al. 2011: 173).

The invariant tag question *innit* is another feature associated with the speech of London adolescents (cf. Anderwald 2008: 457) and Multicultural London English (cf. Kerswill 2014: 436). This lexicalised version of the invariant tag *isn't it* developed in London, but unlike the *this is + speaker* quotative, it has already spread to other parts of the British Isles and into older age groups (cf. Britain 2007; Krug 1998). As stated before, the increased presence of *innit* in Britain might strengthen the use of invariant tag questions in Welsh English (cf. Penhallurick 2008a: 361). An interesting question here, which will be dealt with in this work, is whether the invariant tag questions *isn't it* and *innit* are perceived as one single or as two separate features. Moreover, where would they preferably be localised in the British Isles by Welsh people and Londoners – in Wales, in London, or somewhere else?

(35) *cos I know a few people up there **innit***

(Linguistic Innovators corpus, 16-year-old informant from Hackney)

The other morphosyntactic features of Welsh English discussed above are not described as characteristic of South-Eastern English grammar in the literature (cf. e.g. Anderwald 2008; Britain 2007). According to eWAVE, the inverted word order in indirect questions and, in rare cases, non-standard fronting can be found in the South West of England, but not in the South East (cf. Kortmann and Lunkenheimer 2013). Based on these accounts, I hypothesise that the Welsh English features under investigation are not very salient in London English, an exception probably being *innit*.

3.4 Summary

This chapter has provided an overview of English in Wales with a focus on morpho-syntax and including comparisons with English in London. In Wales, the indigenous Welsh language has been replaced bit by bit by English throughout the last centuries. However, Welsh (1) has recently been undergoing a process of revitalisation, and (2) has left traces on present-day Welsh English that are visible, for example, in the area of grammar. These traces are more or less distinct throughout the country, with English in the traditionally Welsh-speaking heartlands in North West and South West Wales probably showing Welsh-language influences most prominently, while dialects close to the Welsh-English border also share features with dialects from neighbouring English areas. For the present study, some grammatical features of Welsh English were selected that have the potential to be seen as characteristic of Welsh English, for example, because they exhibit influences from the Celtic substrate or because they are regionally confined to Wales. The following six features were explored: focus fronting (*A man he is*), the invariant tag question *isn't it* (*She likes him, isn't it?*), the non-standard habitual progressive (*I'm going to his house every week*), the inverted word order in indirect questions (*She asked me had I called the doctor*), the zero past tense of regular verbs (*Last night I walk home*) and *that*-clauses replacing infinitival subclauses (*I wanted that I should get a higher salary*). While the first four features are discussed in other works on Welsh English, and are assumed to originate from language contact with Welsh, the last two constructions have not yet been explored in the light of Welsh English. They are interesting for this research because eWAVE reports that Wales is the only region where they occur in the British Isles (cf. Kortmann and Lunkenheimer 2013).

As for previous research on Welsh English grammar, the Survey of Anglo-Welsh Dialects (SAWD) with interviews from the 1960s to the 1980s presents a comprehensive overview of uses of non-standard features throughout Wales (cf. Parry 1999). Paulasto (2006) consults SAWD but also newer corpora for her real- and apparent-time analyses of focus fronting and non-standard uses of the progressive. Interestingly, she obtains frequency values for the selected features which are very similar to those in the present study. The perception of Welsh English is, for example, dealt with in the work of Williams, Garrett and Coupland (1996). However, grammatical features are not in the focus of this work. The study of Williams (2003) investigates the acceptability of focus fronting, but does not compare the results with perceptions of other features. Overall, it seems that uses and especially perceptions of non-standard grammatical features in Welsh English are underresearched areas of study, both of which the present study aims to contribute to.

London English features a variety of different accents and dialects, ranging from close to Standard English and associated with educated upper-class speakers (like RP) to further away from the standard and associated with the working classes (as Cockney and Multicultural London English). In terms of grammar, only one of the six Welsh English features discussed above is also mentioned in the literature on London English, namely the invariant tag question, however, realised in London typically as *innit* instead of *isn't it*. The question whether *innit* and *isn't it* constitute one or two features in people's minds will be scrutinised in my salience study (see Chapter 6).

Chapter 4

Research questions and hypotheses

This chapter presents central research questions of the present study, links them to the previously discussed theoretical background literature and past studies, and, on the basis of these accounts, formulates hypotheses.

Research question 1: Does frequency influence the sociolinguistic salience of dialectal grammatical features?

A fundamental assumption in usage-based linguistics is that language use, for example in the form of usage frequencies, shapes our mental representations of language (cf. e.g. Bybee 2006; Ellis 2012; Tomasello 2005). According to Exemplar Theory, representations of language in the mind are usage-based since individual exemplars of constructions, rather than abstract prototypes, are stored in the brain in rich detail (cf. Bybee 2006). On a practical note, it has generally been shown in perception studies that listeners are extremely sensitive to differences in input frequencies (cf. Labov et al. 2011). Rác (2013) more specifically demonstrates that the sociolinguistic salience of several phonetic-phonological features can, to a certain extent, be predicted by their probability of occurrence in language use. In the study by Kerswill and Williams (2000), the high salience ratings of several morphosyntactic features go hand in hand with high usage frequencies (for working class speakers). Consequently, I assume that the present study will evidence frequency effects on sociolinguistic salience in dialect grammar.

Hypothesis 1: The sociolinguistic salience of a dialectal grammatical feature is affected by its frequency in language use.

Research question 2: Does frequency affect intra- and interlectal salience in the same way?

I hypothesise that linguistic items are salient when they are frequent in comparison with other items, since the more frequent items are – on average – more deeply mentally entrenched and are more likely to be cognitively preactivated (cf. Schmid 2007: 120). I thus hypothesise that this effect of token frequency applies to both intralectal and interlectal perceptions. However, for interlectal perceptions, a feature's frequency in one's native as compared to the target variety may also play a crucial role. To consider a feature as typical of another variety entails its deviation from one's own way of speaking, for example in that the feature has a very low probability of occurrence in one's own variety (cf. Rącz 2013).

A further frequency-related factor that I assume applies to both intra- and interlectal perceptions is a feature's general frequency of occurrence in other, close-by language varieties. For example, the degree to which a feature is characterised as typical of Welsh English might be affected by the feature's (non-)pervasiveness in varieties of English in England, Ireland and Scotland. In order to assign a feature to Wales, instead of, say, Ireland, it needs to be more frequent in Welsh English than in Irish English. This applies to interlectal perceptions, since if somebody realises that a feature is not common in one's own variety, high(er) usage frequencies in Welsh English may direct the person to assign the feature to Wales instead of another area. Regarding intralectal salience, lower frequencies of occurrence outside one's own dialect area may lead to realising that a feature is indeed typical of one's native variety instead of being used, for example, all over the British Isles.

Hypothesis 2a: Both intra- and interlectal salience are influenced by token frequencies of dialectal grammatical features in the target variety and by the frequencies of those features across different (close-by) varieties.

Hypothesis 2b: Interlectal perceptions are affected by token frequency differences between one's own and the target variety.

Research question 3: Which other factors influence sociolinguistic salience in dialect grammar?

As early as in 1930, Schirmunski argued that primary (i.e. consciously perceived) dia-

lect features occur in small-scale geographical areas (cf. also Lenz 2003, 2010). Hypothesising that interlectally salient features deviate from what outsiders are used to in their own dialect (cf. Rácz 2013), this may not only be tied to usage frequencies but also to the fact that the features are prevalent only in restricted geographical areas (cf. Auer 2014: 13). In addition to this cognitive prominence, features which are restricted in use to specific small-scale areas may carry strong local signs (cf. Kortmann 2012: 695). I thus assume a correlation between a feature's areal distribution and its degree of regional indexation (simply put, the more restricted, the more characteristic of a specific region).¹

I furthermore assume that the prominence of non-standard items is influenced by structural language-internal factors. There is ample evidence for structural factors in the context of phonetics and phonology, such as phonetic distance or the dichotomous versus categorical character of a variable (cf. e.g. Auer 2014; Auer, Barden and Grosskopf 1998; Rácz 2013; Schirmunski 1930; Trudgill 1986). Some scholars also suggest structural factors influencing morphosyntactic salience. For example, features may be salient when they occur in an interactionally prominent position (cf. Cheshire 1996: 5–6) or generally at the beginning or end of a linguistic unit (cf. Auer 2014: 13).

In line with Purschke (2014), I argue that the factors just described – frequency, areal distribution and structure – influence perceptions of language *on the average*, but these may differ distinctly from individual and subjective perceptions of actual listeners (cf. also Blumenthal-Dramé 2012: 30). I hypothesise that individual salience perceptions are affected by a range of personal, social and attitudinal factors. Such factors prominently include, for example, age (connected to the “sociolinguistic maturity of the judges”, Kerswill & Williams 2002: 176) and people's social networks (along the gradient from close-knit or more open, cf. Kerswill & Williams 2002: 176). Also, I assume that attitudes towards linguistic diversity affect perceptions of dialect grammar (cf. Garrett 2010: 172).

¹It should be noted that the areal distribution of a feature is a relative size (as are usage frequencies), since a layman's perception of a feature's geographical spread may differ significantly from that of a dialect geographer (cf. Lenz 2003: 193). The question of areal distribution thus needs to take the perspectives of individual speakers into consideration (cf. Lenz 2003: 193).

Hypothesis 3: The sociolinguistic salience of a dialectal grammatical feature is influenced by its geographical spread, its linguistic structure and individual personal, social and attitudinal characteristics of the listeners.

I would like to point out that this work does not claim to present an exhaustive list of factors. There may well be further factors not discussed here that contribute to the conspicuousness of a dialectal feature. I rather aim to generally demonstrate that salience is a complex construct that cannot simply be ascribed to frequency, but that is shaped by the interplay of several determinants.

Chapter 5

Methodology

5.1 General research design

A range of theoretical and practical considerations directed me towards the methodological steps undertaken in this work. The overall research design rests upon the following rationales:

1. The explorations of frequency and salience should take those of Rácz (2013) as one point of departure. Rácz's quantitative study investigated salience in the context of phonetic-phonological variation (see Section 2.3). By adopting a similar approach – predicting salience by means of corpus-based usage frequencies – I aimed to test to what extent Rácz's findings can be transferred to the area of dialect grammar.
2. In the context of the corpus-based approach, it seemed necessary to include methods checking whether the token frequencies in the data were tied to particularities of the corpora used.
3. Apart from token frequencies, the project should investigate other potential causes of salience. Although I take frequency to be an important explanatory factor of salience, proceeding on the assumption that it is the *only* factor would be too simplistic in view of the range of determinants discussed in the scholarly literature (see Section 2.3).

The rationales resulted in incorporating the following methods, detailed in Figure 2. I explored perceptions of Welsh English grammar and causes of these perceptions both

for people from Wales (intralectal salience) and for people from outside of Wales (interlectal salience); here, from London. The salience of several non-standard morpho-syntactic features was determined by means of a questionnaire-based survey with 150 informants from Wales and 150 informants from London. The questionnaire also tested for a range of personal, social and attitudinal factors that may influence individual dialect perceptions.

Corpus analyses were carried out to determine the token frequencies of the grammatical features under investigation. The features were analysed in terms of their (1) frequencies in Welsh English (Radio Wales Corpus, self-compiled) and their (2) frequency ratio in Welsh English versus London English (Linguistic Innovators corpus). Corpus analyses were also employed to find out whether these token frequencies can be generalised beyond individual corpora. Furthermore, an analysis of the language of Welsh presenters, comedians and actors tested for differences in the use of grammatical features in the media versus in everyday life (as approximated by the corpora).

Two further potential determinants of salience in Welsh English, a feature's (1) pervasiveness in other British Isles areas apart from Wales and London, and (2) geographical spread were investigated using the Electronic World Atlas of Varieties of English (eWAVE; cf. Kortmann and Lunkenheimer 2013). The relation between salience and the degree to which a feature is considered a deviation from the standard language was explored in the context of a further questionnaire involving 25 subjects from Wales and 25 subjects from London.

Several statistical methods were applied in evaluating the findings from these studies. For example, correlation calculations were carried out to determine the strength of relation between salience and frequency, as well as other factors, for insiders (Welsh people) versus outsiders (Londoners). In the following sections, the individual methods are discussed in more detail and embedded in the context of the relevant literature.

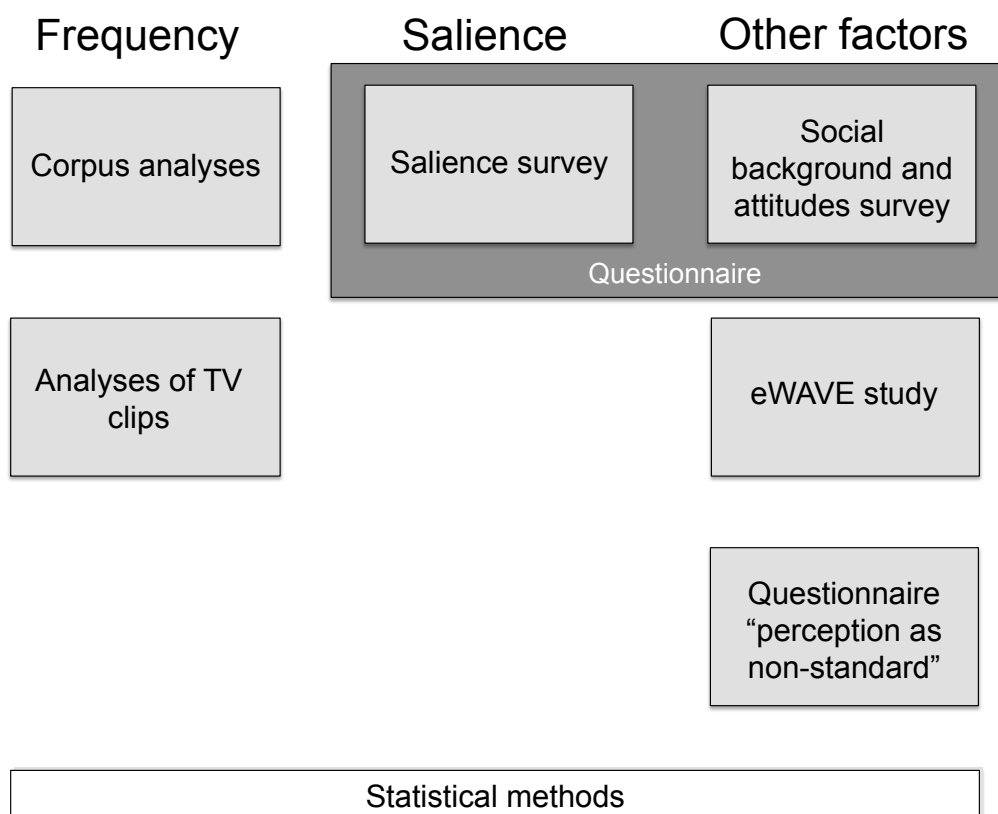


Figure 2: Overview of the methods employed in the present study

5.2 Salience survey

5.2.1 Conceiving the survey

So far, studies systematically investigating differences in the salience perceptions of regional linguistic variables between speakers from inside and outside a dialect area are rare (cf. Hettler 2014: 72). Thus, the salience survey was designed in an attempt to explore grammatical features of Welsh English both in the context of intralectal salience (speakers from Wales) and of interlectal salience (speakers from London). A reason for choosing Welsh English as the regional variety under investigation was that it features some morphosyntactic structures, assumedly tracing back to substratal influences from the Welsh language, that differ from other British Isles dialects in their form or in their frequency of use (cf. Paulasto 2006: 214; see also Chapter 3). These structures seemed to be an interesting starting point for studying the regional labelling

of linguistic variables, i.e. the degree to which features are seen as characteristic or typical of a particular geographical area. A further reason for focusing on Welsh English was that this variety of English has so far received scant attention in dialectological research, quantitative accounts of Welsh English grammar being particularly rare (see Chapter 3).

Choosing English in London as the reference variety and Londoners as the judges from outside of Wales was based on the following considerations. The dialect background of my reference group should *differ* to some extent from Welsh English, to enable a view on how outsiders perceive the target dialect. As outlined in Section 3.3, English in London differs from Welsh English in that it lacks direct transfer from the Welsh language (while Irish and Scottish English, for example, probably have more in common with Welsh English, sharing some Celtic-induced features). Moreover, London does not border Wales, and so Welsh English grammar might have fewer grammatical overlaps with London speech than, for example, dialects in the neighbouring South West of England or the Western Midlands. Nevertheless, the geographical distance between London and Wales is not too large. Thus, I assumed that my London judges would be, to a certain degree, familiar with English in Wales, which is an important precondition for studying the regional labelling of variables.¹ Choosing London speech because of its assumed proximity to Standard English would only be a pseudo argument. Although the standard presumably evolved in London, the city can nowadays be characterised as a superdiverse linguistic melting pot (see Section 3.3). What is more, in recent years, a range of working-class linguistic variables have been spreading to higher RP-speaking social classes (cf. Altendorf and Watt 2008; Anderwald 2008). In Szmrecsanyi's (2013) dialectometric study of grammatical variation in the British Isles, based on the FRED corpus, London English and Welsh English do not differ significantly with regard to morphosyntactic distance from Standard British English. The distance values range from 4.11 (closer to standard) to 8.90 (more distant from standard). London receives a value of 6.00, while the figures for the Welsh regions Glamorganshire and Denbigh-

¹Someone from the US might also not use the Welsh English features investigated here, and might perceive them consciously when they occur in speech because they differ from what he or she is used to. However, given the geographical distance between Wales and the US, I assume that this person is (on average) less likely to know that the features occur in Wales.

shire amount to 5.88 and 6.45.

How can sociolinguistic salience in dialect grammar be operationalised and tested? I decided to work with dialect recognition tests, presenting the subjects with non-standard features which they had to assign to regions, instead of free elicitations, asking the subjects to name all grammatical features of Welsh English that came to mind. The literature indicates that it is difficult to name grammatical regionalisms without any guidance. In the context of their draw-a-map survey in Wales, Williams, Garrett and Coupland (1996) report on phonological and lexical but barely any morphosyntactic features as named by subjects to describe regional dialects of Welsh English. Similarly, Garrett, Williams and Evans (2005) state that the subjects in their folk linguistic study passed linguistic comments on English in England, the US, New Zealand and Australia in the areas of pronunciation and lexis, while morphosyntactic features went unmentioned.

As for the format of the study, a questionnaire was used, presenting grammatical features in written form. The method of testing the awareness of grammatical structures by means of a questionnaire with written stimuli has been applied by other scholars, such as Cheshire, Edwards and Whittle (1995), Kerswill and Williams (2000) and Rosen (2014). A decisive reason for presenting linguistic features on paper and not audiotively was that I wanted to isolate judgements of grammatical features from their pronunciation. It can be assumed that effects of the accent used might have overshadowed effects of grammatical structures. According to Rydén (1991: 351), syntactic structures underlie social evaluations to a lesser extent than, for example, phonological and lexical features, since the former are repeated less frequently (cf. also Cheshire 1996: 2). The reference semantic neutrality of some morphosyntactic features may also make them being less prominent than, for instance, items representing lexical contrasts (cf. Auer 2014: 14). Thus, I feared to encounter one of the two following scenarios when presenting grammatical features as audio stimuli in a dialect recognition task: (1) In case the features were read with a Welsh accent, the accent might have been the prevailing reason for assigning the features to Wales, irrespective of their grammatical structure. (2) In case the features were read with a Standard English accent, they might not have been assigned to Wales at all, regardless of non-standard items in the grammatical structure. On a more practical note, a questionnaire seemed to be a convenient means

to test a large number of informants in a relatively short period of time and not dependent on a particular place (like a quiet room for audio stimuli). My informants filled in the questionnaires in parks, trains, cafes, bus stations and many other places, and it was possible to test larger groups of people at once without any further preparations.

The saliency questionnaire, which can be found in Appendix A, consists of three parts:

- Saliency test I
- Saliency test II
- Personal information and language attitudes section

Saliency test I – the main saliency test in the present study – shares some characteristics with the “grammar saliency test” by Kerswill and Williams (2000: 86). In their study, the participants were presented with a list of sentences containing non-standard features, and they had to mark the constructions that they considered to occur locally. Saliency test I in the present study also included a list of sentences containing non-standard features, however, the subjects were asked to indicate where they would locate the sentences in the British Isles. I wanted to avoid the informants knowing from the beginning that the survey focused on Welsh English. Instead, the questionnaire should appear to test general knowledge of non-standard features across the British Isles. I thus hoped that only those Welsh English features were assigned to Wales that the informants actually considered to be Welsh English, without any general bias towards Wales.

Saliency test I included 26 sentences involving one non-standard grammatical feature each (cf. Elmentaler, Gessinger and Wirrer (2010) for a similar number of test sentences in their saliency study, also including one non-standard feature each). Eight of these were Welsh English features:

- Focus fronting
- Invariant tag question *isn't it* (in the orthographic notation <isn't it>)
- Non-standard habitual progressive
- Inverted word order in indirect questions

- Zero past tense of regular verbs
- *that*-clause replacing an infinitival *to*-clause
- Non-standard habitual *do*
- *there's* replacing *how*²

The remaining sentences involved non-standard features from other British Isles varieties of English, namely Irish English, Scottish English and English English dialects. These sentences primarily served as distractors, masking the fact that the questionnaire actually explored Welsh English. The sentences, however, also provided the opportunity to be analysed in the context of salience, for example, in an attempt to back the findings obtained for Welsh English.

For each of the 26 sentences, the participants had to indicate where they thought the speaker came from. As can be seen in Figure 3, they could tick boxes for the following regions: England (North), England (South), Ireland, Scotland and Wales.

1. <i>The cat wants petted.</i>							
England	<input type="checkbox"/>	N	<input type="checkbox"/>	S	<input type="checkbox"/>	Ireland	<input type="checkbox"/>
Scotland	<input type="checkbox"/>	Wales	<input type="checkbox"/>	Other	<input type="checkbox"/>	_____	
						<input type="checkbox"/>	Don't know

Figure 3: Exemplary sentence from salience test I in the salience questionnaire

The regions were presented in an alphabetical order. Apart from these options, it was also possible to tick “Other” and write down the area one had in mind (this could also be a more specific British Isles region, such as “West Country”). Additionally, “Don’t know” could be selected. Since most of the non-standard features in the questionnaire are to be found in more than one variety of English in the British Isles (cf. Kortmann

²In the course of my research, I decided not to include the habitual *do* construction (*I do go to the theatre every week*) and *there's* replacing *how* (*There's nice he is*) in the analyses presented in this work due to several practical reasons. As for the *do* construction, its frequencies could not be determined in the Linguistic Innovators corpus since I did not have access to the audio files, and the feature could be confused with the emphatic *do* construction in written texts (*I do like him a lot*). The *there's* construction is not covered by eWAVE, hence it was not possible to test this feature regarding regional dispersion. I thus decided to focus on the remaining six features, since they fulfilled all the criteria I was interested in and could be included in the same way in all analyses. Nevertheless, the data for those two additional features could offer interesting material to scrutinise in future projects.

and Lunkenheimer 2013), the participants had the opportunity to select several areas for each grammatical construction.

Salience test II aimed at backing the tendencies of test I, using a different methodology and different test sentences. By this, I wanted to make sure, at least to some extent, that the salience ratings in test I were not tied to the specific sentences used. In test II, the subjects were presented with four pairs of sentences. Each sentence involved one non-standard grammatical feature. The subjects were told that in each pair, one of the two sentences was produced by a speaker from Wales, and that they had to determine which sentence it was. What the participants did not know was that in fact *all* sentences involved grammatical features of Welsh English.³ By opposing two features each, the test explored which of the two was seen as more characteristic of Welsh English. Such a test necessitated the creation of several versions of the questionnaire, since, for example, if every informant had to judge the pair of focus fronting and the *that* construction, it would only have been possible to gain information about how salient focus fronting is as opposed to the *that* construction, but not as opposed to the rest of the features. With eight Welsh English features in this test, there were seven combinatorial possibilities for each feature (for example, focus fronting could be in a pair with *isn't it*, the habitual progressive, the non-standard indirect question, the zero past, the *that* construction, the habitual *do* and *there's* replaces *how*). Thus, seven versions of the questionnaire were created (each of them containing four pairs of sentences) that were to be distributed to the same numbers of subjects.

The third part of the questionnaire investigated personal, social and attitudinal factors which could have influenced the salience values from tests I and II. The first questions here were identical for the subjects from London and those from Wales and partly based on the "Informant Information Sheet" provided by Krug and Sell (2013: 23). The questions concerned the following variables to be analysed in the context of (intra- and interlectal) salience: (1) sex, (2) age, (3) educational background, (4) having friends speaking another dialect than one's own and (5) attitudes to dialectal diversity.

³The features, again, were focus fronting, the invariant tag question *isn't it* (<isn't it>), the non-standard habitual progressive, the inverted word order in indirect questions, the zero past tense of regular verbs, the *that* replaces *to* construction, the non-standard habitual *do* and *there's* replaces *how*.

As for (1) sex, I was first of all interested in general differences between male and female participants in the accuracy of allocating features to the respective regions. Second, I wanted to find out if there were sex-related differences among the Welsh informants in assigning features to their *own* area. In a survey about language attitudes conducted in the context of the BBC Voices project with over 5,000 participants, women generally awarded higher social attractiveness ratings to a range of different accents than men, but men provided more favourable ratings of their own regional accent (cf. Garrett 2010: 175). This could be connected to women tending (or at least claiming) to use more standard language than men, which has been shown by a range of studies (cf. e.g. Edwards 2009; Elmentaler, Gessinger and Wirrer 2010). A reason for that may be a higher status-consciousness of women (cf. e.g. Garrett 2010). If women are “less socially secure than men [...], they may wish to gain status through the use of more standard forms” (Edwards 2009: 134).⁴ The use of more standard forms might lead to women identifying less with their local dialect than do men.

Group-internal differences may also be related to the factor (2) age. According to Kerswill and Williams (2002: 176), age is tied to the “sociolinguistic maturity of the judges”. Older speakers might simply be more experienced regarding local and non-local dialect forms than younger speakers. Moreover, apparent-time differences in dialect perception might mirror apparent-time differences in dialect production. Some older and probably nowadays infrequently used features may no longer sound familiar to the younger generations, whilst still being in the active use of older speakers.

Linguistic in-group variation can furthermore be connected to the (3) educational backgrounds of the group members. Educational background is considered to be a sub-parameter of social class by Krug and Sell (2013: 4). In the context of their dialect levelling study in urban English areas, Kerswill and Williams (2000) record differences between working-class and middle-class subjects in their perception of non-standard features. “Middle-class non-users of the features often do not recognise local features [...] [and] resort to stereotyping, resulting in the reporting of features which are not, in fact, present” (Kerswill and Williams 2000: 89). In the present questionnaire, edu-

⁴This is, of course, not the only explanation for women using more standard language than men. Another reason, for example, could be that working-class non-standard speech carries “connotations of masculinity” (Wells 1982: 20) and a covert prestige for male speakers (cf. Edwards 2009: 135).

cational background was explored by asking the subjects to name their highest educational qualification (cf. Krug and Sell 2013: 23).

The questionnaire, moreover, investigated the participants' familiarity with different dialects of English by asking them (4) whether they had friends speaking (an)other dialect(s). The theoretical assumption behind the question was that dialect forms are perceived differently by speakers who are actively in contact with other dialects than by people whose social reference group is restricted to speakers of the same dialect (cf. Kerswill and Williams 2002: 176). The latter might not become aware of some features constituting deviations from the standard language because they are not familiar with alternatives. As Johnstone, Andrus and Danielson (2006: 82) describe it in the context of their study of English in Pittsburgh,

[t]he frequency of regional variants can be correlated with being from southwestern Pennsylvania, especially from Pittsburgh, and with being working-class and male. But for socially nonmobile speakers in dense, multiplex social networks, these correlations are not noticeable, because 'everybody speaks that way.'

In addition to the familiarity with different dialects, the questionnaire investigated (5) the subjects' attitudes towards linguistic diversity. The informants had to indicate their opinion on the statement "I like hearing a range of dialects". The inspiration for this question came from the BBC Voices attitudes survey described above (cf. Garrett 2010: 172). The subjects had to express their opinion on "I like hearing a range of accents" (Garrett 2010: 172), ticking "agree", "disagree" or "undecided" (cf. Crystal 2013: 23). In my questionnaire, a seven-point Likert Scale with the endpoints "strongly agree" and "strongly disagree" was used to enable more fine-grained stances on the question. As described by Garrett (2010: 176), the BBC survey showed that participants who were more open towards linguistic diversity awarded higher social attractiveness ratings to different accents. He concludes that

[t]his general criterion of whether and to what extent people value socio-linguistic diversity is clearly a particularly powerful one and it cuts across sociodemographic variation. It has not received attention in language attitudes research previously, but seemingly deserves more attention.

The present study explores whether people's attitudes towards linguistic diversity have an effect on their performance in a dialect recognition task. A usage-based hypothesis

here would be that subjects with positive attitudes towards dialectal diversity are better at identifying linguistic regionalisms, since they deal with different dialects more frequently and, thus, are more familiar with them.

In addition to these questions, the questionnaire recorded the informants' native languages. Regarding the London informants, this information was solely used to eliminate all questionnaires from the study that came from participants with a native language other than English. As for the Welsh informants, the aim was to explore to what extent their native language(s) (English, Welsh or both) had an influence on the recognition of Welsh English features. As illustrated in Chapter 3, several of the grammatical features investigated in this work show substratal influences from the Welsh language. It is thinkable that Welsh native speakers are better at assigning these features to Wales since they also know the structures from Welsh.

Furthermore, the questionnaire asked the Welsh and London informants to indicate their place(s) of living at different ages, based on Krug and Sell (2013: 23). For the Welsh subjects, this information was used later on to determine to what extent perceptions of non-standard grammatical features differ between people from different corners of Wales. It could be that some features are more prevalent in particular Welsh regions than in others, and that this is mirrored in people's perceptions. With regard to the London participants, information about places of living could, for example, reveal whether they had lived in Wales for a particular period of time.

Apart from these questions that were identical for the Welsh and the London participants, the questionnaire for the Londoners inquired as to their familiarity with Wales and Welsh English. Two keyword tasks were applied, simply asking the Londoners to write down the first five things that came to their minds when thinking about (1) Wales and (2) English in Wales. The technique is based on Garrett, Williams and Evans (2005: 216), who assume the items named most frequently to be the most meaningful or salient to the respondents.

The London questionnaire also examined different modes of contact with Welsh English. The modes involved knowing Welsh people, having been to Wales for a particular period of time and being familiar with Welsh English in the media (in the form of

TV and radio shows, presenters and comedians). As for knowing Welsh people, the respondents had to indicate whether they had friends, family members or work colleagues from Wales. In a next step, they had to state whether they had friends, family members or work colleagues speaking Welsh English. This was based on the assumption that being from Wales does not necessitate speaking a Welsh dialect of English. Londoners having Welsh friends who speak, say, Standard English are not necessarily familiar with dialectal features of Welsh English.

If the respondents indicated that they knew people speaking Welsh English, they had to elaborate on how frequently they spoke with these people. The data here were elicited using a qualitative ordinal scale based on Krug and Sell (2013: 6–7), with the options “Every day”, “Regularly”, “Sporadically” and “Never”. For the evaluation of the data, it needs to be kept in mind that the options “cannot be translated straightforwardly into absolute numbers” (Krug and Sell 2013: 6) since no information is provided about “the quantitative distance between those options” (Krug and Sell 2013: 6, cf. also Franceschini, Galetto and Varetto 2004). For example, “Every day” and “Regularly” might be further apart from each other than “Sporadically” and “Never”. However, such scales indicate a general decline from an “imagined maximum pole” (Krug and Sell 2013: 6) to an “imagined zero” (Krug and Sell 2013: 6), thus they are often considered to be “quasi-interval-scaled” (Krug and Sell 2013: 6).

In the next question, the subjects had to indicate whether they had already been to Wales (“Yes”/“No”). Those who selected “Yes” were asked to further specify how much time in total they had spent in Wales. The scale here ranged from “Up to a week” over “Up to a month”, “Up to 3 months” and “Up to 6 months” to “Longer than 6 months”. More information on longer periods of living in Wales could be obtained from the above-mentioned question about places of residence at different ages. Here the informants could specify for how many years or decades they had lived in a particular place, for example, in Wales.

In addition to exposure to “everyday” Welsh English, it was tested to what extent the London respondents were in contact with Welsh English as spoken in the media. According to Kerswill and Williams (2002), a decisive factor for the recognition of dialects from outside one’s own speech community, apart from personal contact with speakers,

is “the influence of the broadcast media. In the contemporary world, the broadcast media are a crucial means by which familiarity with varieties is spread”. For example, in the context of the dialect recognition tasks discussed by Kerswill and Williams (2002), a range of judges stated that they recognised a Newcastle accent because they were familiar with the TV series *Byker Grove*, set in Newcastle. Apart from programmes, regional dialect forms may be promoted through the media language of individual comedians or presenters, as in the case of Welsh radio presenter Roy Noble (cf. Coupland 2001). In the present study, the Londoners had to indicate whether they knew TV or radio programmes from Wales, and if so, which programmes these were. Furthermore, in case the respondents knew any TV or radio presenters or comedians from Wales, they had to state their names.

5.2.2 Conducting the survey

The questionnaire-based salience survey was conducted between March 2013 and July 2014 in Wales and London. The study comprised 150 subjects from Wales and 150 subjects from London. The numbers of participants were based on the perception studies of Elmentaler, Gessinger and Wirrer (2010: 144 participants) and Garrett, Williams and Evans (2005: 130 participants on average in the cohorts from England, Australia, New Zealand and the US). The participants were randomly selected people in the streets as well as acquaintances of mine.⁵ I wanted to include people with different personal and social characteristics (such as age, sex and educational background) to test for effects of such characteristics on salience perceptions. The corpus of Welsh English used for the frequency-related analyses in this work (Radio Wales Corpus, see the following section) also involves speakers who differ, for example, regarding age, sex and educational background, and I aimed for some kind of comparability between the groups of *speech producers* and *speech perceivers* from Wales.

⁵For reasons of practicality, some of these acquaintances filled in the questionnaire online, while the overall majority of informants filled in the printed version of the questionnaire. This should not be a major problem, however. For some questions in the online version it was obligatory to give an answer in order to proceed with the next question. In the paper questionnaire, me and the people who helped me distribute the questionnaires ensured that all of these questions were answered, too. Likewise, in the online version, it only becomes apparent in a later part of the questionnaire that the survey focuses on Welsh English. In the paper version, this information can be found on the back side of the questionnaire. For each informant, we ensured personally that he or she did not turn the page before all questions on page one had been answered.

Moreover, the Radio Wales Corpus comprises data from different regions of Wales, so I likewise distributed my questionnaires in different areas in North, Mid and South Wales. Apart from a majority of subjects born and raised in Wales, the questionnaire-based survey also included people who moved to Wales (or to London, respectively) at a later point of time. This offered the opportunity to trace (experience-induced) perceptual differences and similarities between insiders and incomers to a region (cf. Lorenz 2014). The precondition here was, however, that all participants had grown up in the British Isles and were native speakers of English and/or Welsh.

Of the 150 subjects tested in Wales, 62.67 per cent were female and 37.33 per cent male. In the London cohort, 48.67 per cent of the subjects were female and 51.33 per cent male. The Welsh subjects had been born between 1922 and 1998. Note that although all decades of birth between the 1920s and the 1990s are covered by the subjects, the decades are not equally represented in the sample. Young people born in the 1990s constitute the largest group of informants (36 per cent), one reason being that they were easy to find in and around universities. The informants in the cohort from London were born between 1932 and 1996. All decades of birth between the 1930s and the 1990s were covered by the informants, however, again with a bias towards the younger age groups.

Most of the questionnaires in Wales were collected in Cardiff, Swansea and Bangor, but there was also a range of informants from other locations in North, Mid and South Wales. Figure 4⁶ shows the different places of residence of the Welsh participants (excluding those who had simply stated, for example, that they were from “South Wales” or from “Wales”).

Of the informants from Wales who specified their native language(s) in the questionnaire, 11 per cent indicated that Welsh was their native language, 15 per cent considered both Welsh and English as their native tongues and 74 per cent stated that English was their native language. These tendencies fit recent census data finding that Welsh is spoken by about 20 per cent of the Welsh population (cf. Williams 2014: 242).

⁶The map was created by myself using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. Copyright © Esri. All rights reserved. For more information about Esri® software, please visit www.esri.com.

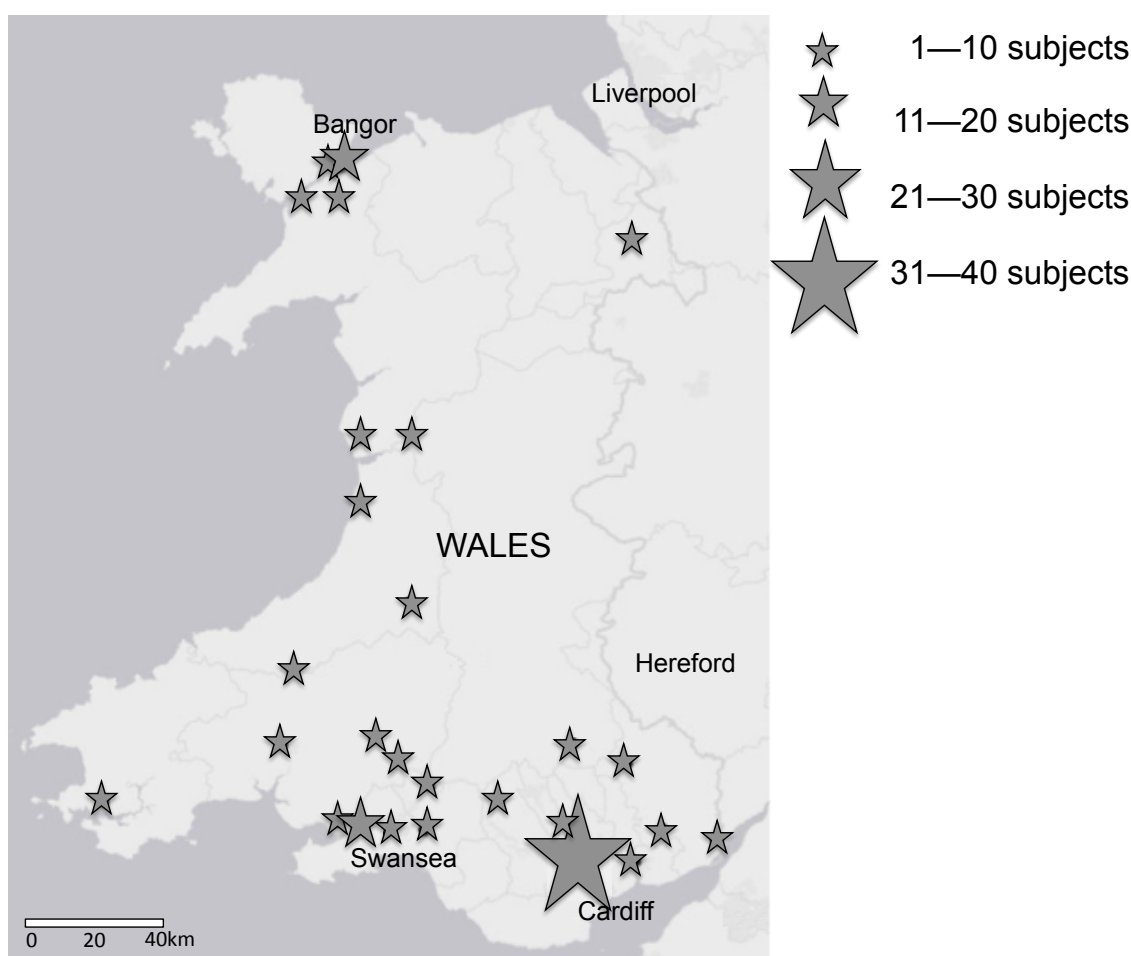


Figure 4: Map showing the places of residence of the Welsh informants in the salience survey

5.3 Corpus analyses

5.3.1 Radio Wales Corpus

For the frequency-based analyses in the present study, a corpus of spoken present-day Welsh English was needed. *Spoken* since the grammatical features investigated are reported to predominantly occur in spoken language – and there probably most in informal conversations in localised dialects (cf. Trudgill and Hannah 2002: 31). *Present-day* because I felt it desirable to have some temporal comparability between the speakers in the corpus and the informants in the salience survey. I decided to compile such

a corpus myself, namely the Radio Wales Corpus (RWC), since no suitable already existing database could be found. As hinted in Chapter 3, the only corpora of Welsh English with interviews from later than the 1980s are Heli Paulasto's corpora (Llandybïe Corpus and North Wales Corpus, cf. Paulasto 2006), but the databases were not accessible to me. Apart from being older, the Survey of Anglo-Welsh Dialects (SAWD) did not seem to be an appropriate source for testing my frequency-based hypotheses, since the interviews mainly consist of structured answers to a questionnaire primarily testing the use of phonological and lexical features, while "[s]yntax remains in the background" (Pitkänen 2003: 117). Frequency counts here would only restrictedly be able to represent everyday language use, since occurrences of particular features seem to be closely tied to the structure of the interviews. Along these lines, Penhallurick (2008a: 367) remarks that examples of focus fronting "are rare in SAWD data, because they are restricted to incidental material." What is more, the informants in SAWD are mainly non-mobile rural old males (NORMs), while my study should include different age groups as well as male and female speakers. NORMs and interviews mainly from the 1970s were also the reasons for not working with the Freiburg Corpus of English Dialects (FRED, cf. Hernández 2006).⁷ The Ceri George Corpus furthermore did not seem to be an ideal database for the present study, since, apart from being from the 1980s, it consists of handwritten transcripts and it is unclear to what extent these written accounts represent everything that was said in the interviews (cf. Williams 2003). The following sections outline the creation of and key information about the Radio Wales Corpus (RWC).

Conceiving the corpus

There were several reasons for re-using existing recordings, namely of the BBC Voices study and the Millennium Memory Bank, instead of conducting interviews myself. First of all, not having to conduct interviews saved time. Additionally, both interview collections encompass recordings from a broad range of places all over Wales. It would have been a fool's errand to attempt to engage all the interviewees for the recordings myself. Also, the interviewers in the two collections were, in contrast to myself, native speakers of English, most of whom grew up in Wales themselves. Being interviewed

⁷At least not in my main study. I carried out some preliminary corpus analyses with FRED, which provided me with a helpful general overview of occurrences of non-standard grammatical features in Wales (see Section 3.2).

by such insiders “coming from the area, often still speaking the dialect themselves [...] tends to relax the interview situation considerably” (Anderwald and Wagner 2006: 37–38). As a result, the interviewees might use more of their own local dialect than when being interviewed, for example, by someone speaking RP.

The corpus I compiled is called Radio Wales Corpus (RWC), since it consists of interviews from two BBC radio collections: BBC Voices and the Millennium Memory Bank (MMB). The BBC Voices study, conducted between 2004 and 2005, aimed to provide a “a snapshot of the linguistic landscape of the UK at the beginning of the 21st century” (British Library 2015). The project encompasses group interviews recorded in over 300 locations across the UK with more than 1,200 interviewees (cf. Elmes 2013: 8). The project being conceived by the BBC, one central objective was to generate “programmes for national and local radio and television, and material for the BBC website and publications” (Upton and Davies 2013: xii). However, the collection was also considered to have a lot of potential as a database for linguistic analyses (cf. Elmes 2013: 5), and researchers such as Clive Upton and David Crystal served as linguistic advisors in conceiving the interviews. It should be noted that the BBC Voices recordings presumably cannot fully account for the Observer’s Paradox (observing “how people speak when they are not being observed”, Labov 1972a: 113), since the interviewees knew that parts of the recordings would be used for radio broadcasts. Also, the topics of the conversations, among other things, affected local lexemes and attitudes to language. However, I assume that the formality of the situation, and the (potentially) resulting attention on one’s own speech, was reduced due to (1) having group interviews with people who knew each other well, (2) the recordings being carried out in the interviewee’s homes, and (3) the interviewers being typically from the area themselves, speaking the local dialect. Moreover, Elmes (2013: 10) remarks that even the discussions of local words led into various stories and anecdotes of the subjects’ lives, being told “naturally and unforcedly”. The BBC Voices recordings, according to Elmes (2013: 10), are “myriad conversations about ordinary life – living, loving, dying, shopping and working, eating and sleeping”. Still, before compiling the corpus, in order to make sure that the BBC Voices material was worth consulting for analyses of dialect grammar, I examined the grammatical commentaries of all BBC Voices Wales recordings provided by Jonnie Robinson from the British Library (cf. Robinson, Herring and Gilbert 2009–2012). These commentaries contain “an inventory of [grammatical] forms

that contrast with Standard British English” (British Library 2015). In these data from Wales, a range of non-standard grammatical features were mentioned, which made the Voices interviews an interesting source for the study of Welsh English dialects.⁸

The English-language BBC Voices recordings from Wales – all of which I transcribed for the RWC – comprise 22 conversations totalling about 20 hours of speech.⁹ These 22 interviews involve 110 interviewees in total. The group sizes of the interviews range from three to eight speakers (excluding the interviewer), with an average of five speakers per conversation. 62.73 per cent of the interviewees are female and 37.27 per cent are male. The informants were born between 1923 and 1989. They are relatively equally distributed according to decades of birth; however, the youngest speakers born in the 1980s constitute the largest group, while the smallest group includes the speakers born in the 1920s. The interviewees represent different strata of Welsh society (cf. Elmes 2013: 6) and exerted a wide range of professions at the time of recording. The interviews were conducted in various locations in Northern, Southern, Eastern and Western Wales (see Figure 6 for an overview of all interview locations in the RWC).

The Millennium Memory Bank (MMB) is Europe’s largest oral history archive and encompasses 6,069 interviews from across the UK, conducted between 1998 and 1999 (cf. Perks 2001; Roller 2015). The project arose from a partnership between BBC Radio and the British Library (cf. Gallwey 2013: 40). The overall goal was “to create an archival ‘snapshot’ of ‘ordinary’ Britons’ opinion and experience at the turn of the century” (Perks and Robinson 2005: 81). As was the case for BBC Voices, the BBC aimed to use parts of the MMB material for radio programmes, while “the British Library’s oral history curatorial team saw an opportunity to create an oral history collection of unprecedented scope” (Gallwey 2013: 40). According to the Oral History Society (2015), oral history “is the recording of people’s memories, experiences and opinions”. The MMB interviews, thus, centre around memories, experiences and opinions of the subjects, loosely embedded into groups of themes such as

⁸Moreover, my frequency-based analyses with the RWC later on found the corpus to be comparable with Heli Paulasto’s databases on present-day Welsh English. Although the interviews in the Llandybïe Corpus were not to be broadcast on the radio, the token frequencies of non-standard habitual progressives, focus fronting and embedded inversions there were similar to the frequencies in the RWC (cf. Meriläinen and Paulasto 2014; Paulasto 2006).

⁹Not transcribed were the nine Welsh-language interviews (cf. British Library 2015).

“Where we live”, “Growing up” and “Beliefs and fears” (Perks 2001: 98). The fact that the focus of the interviews was not on linguistic topics may have led to a reduced attention of the informants to their manner of speaking. Many of the themes being quite emotional might add to the “speaker’s attention [being] [...] genuinely on what was being said, rather than on how it was being said” (Anderwald and Wagner 2006: 38). Still, to make sure that the material was worth consulting for analyses of Welsh English grammar, despite the fact that the informants knew parts of the conversations would be broadcast on the radio, I consulted the grammatical commentaries by Jonnie Robinson on selected MMB interviews from Wales (parts of which can be found in the Sounds Familiar section of the British Library’s homepage: <http://www.bl.uk/learning/langlit/sounds/>). Since these commentaries mention a range of non-standard grammatical features in the data, I felt the MMB could be a fruitful source to explore grammatical variation in Welsh English.

The British Library (at <http://cadensa.bl.uk/>) lists 196 MMB recordings from Wales with 280 interviewees. I transcribed parts of these recordings totalling nine hours of speech. Transcribing all MMB interviews from Wales would have been beyond the scope of this work, especially since the full-length recordings can only be heard at the British Library in London (where I spent a research stay). The parts of the MMB I transcribed involve 22 speakers, 10 of them female and 12 male. The majority of the conversations include one interviewee only (which is attributable to the fact that the MMB recordings from Wales generally mainly consist of single interviews; 85 per cent of the MMB Wales interviews are single interviews compared to 15 per cent group conversations). The informants in my MMB transcripts can be assigned to two age groups – 25 to 35 and 60 to 70 years of age at the time of recording.¹⁰ An initial idea was to include speakers from all decades of birth, i.e. from the 1890s to the 1990s. However, given the limited time I had to transcribe interviews, I decided to focus on two age groups only in order to increase the comparability of the transcripts in this relatively small dataset (instead of having, for example, only one speaker born in the 1890s, one born in the 1900s and so on). The speakers in the MMB transcripts further-

¹⁰In addition, the transcripts include two speakers who were aged 73 and 78 at the time of recording. Their interview was chosen for several reasons. For example, it takes place in the North of Wales, while the majority of MMB Wales interviews are from southern parts of the country. Furthermore, the recording also features two speakers aged in their sixties who were from the area and had never moved far away.

more had rather modest educational backgrounds (most of them finished school at 14 to 17 years of age) and they had typically never moved far from where they grew up. The interviews took place in different locations across Wales. The majority of MMB interviews transcribed were conducted by Anita Morgan, a Welshwoman, who had also conducted most of the BBC Voices interviews in Wales. I assumed that this increased homogeneity between the BBC Voices and the MMB recordings.

Compiling the corpus

The interviews were orthographically transcribed, based on the transcription conventions of the Freiburg English Dialect Corpus (FRED, cf. Hernández 2006). Detailed information about the markup can be found in Appendix B. As for the transcription software, the programme f5 was used, which facilitates relatively fast transcriptions by providing shortcuts for frequently-used words or phrases (cf. Audiotranskription 2015). Another advantage of the programme is that it allows for slowing down the speech tempo in audio files, which can be a helpful method to understand utterances that are unintelligible at a normal speech rate.

As regards access to the audio files, all full-length BBC Voices recordings were (and are) available online on the British Library's homepage (<http://sounds.bl.uk/Accents-and-dialects/BBC-Voices>). The MMB recordings, by contrast, are not accessible online, apart from some short extracts of interviews to be found on the British Library's homepage (<http://sounds.bl.uk/Accents-and-dialects/Millennium-memory-bank>).¹¹ Since I assumed the MMB to be a highly valuable source for research on Welsh English grammar, and since no other comparable open access interview collection could be found, I spent four weeks at the British Library in London in order to transcribe MMB interviews on site.

Following the transcription of the the BBC Voices and the MMB recordings, the exact word counts for each interview were determined. By means of R software (cf. The R Foundation 2015), the markup was deleted from the interviews. More specifically,

¹¹According to Perks (2001: 100), one reason for not providing full access to the interviews online was the absence of funding. “[C]asualties of insufficient funding were the wholesale transcription of the recordings and a state-of-the-art digital mass storage system, which would have held the actual minidisc interviews (and information about them), in a permanent web-accessible form” (Perks 2001: 100).

metalinguistic annotations such as “(overlap)”, which were not meant to enter into the word count, were removed so that only the plain text remained. The RWC in total encompasses 268,843 words excluding interviewer utterances (327,463 words including the interviewers). The BBC Voices transcripts add up to 191,077 words excluding the interviewers (231,184 words including the interviewers) while the MMB transcripts involve 77,766 words excluding interviewer utterances (96,279 words including interviewer utterances).

As well as word counts for each interview, the numbers of words for different subgroups in the corpus were also determined in order to provide the opportunity to analyse frequency in relation to different social and regional parameters. In more concrete terms, the numbers of words were determined for female versus male speakers, for people from different regions in Wales and for three age groups (people aged up to 30 years of age, people between 31 and 60 and people aged 61 or older). 161,084 words in the RWC were spoken by female speakers, while male utterances account for 107,759 words. 41,406 words were counted for the youngest age group, 120,016 for the speakers between 31 and 60 and 103,832 words for the oldest age group. The numbers of words per dialect area in Wales are displayed in Table 2.

Region	Number of words
South East (urban)	30,089
South East (valleys)	105,490
South West	13,540
Mid Wales	27,676
North East	30,559
North West	53,539
Sum	260,893

Table 2: Numbers of spoken words in recordings from different dialect areas of Wales in the RWC. The areas are based on Garrett, Coupland and Williams (1999: 325). One interview was excluded from the count since it was unclear whether it was conducted in the North West or the North East of Wales

The RWC comprises 132 speakers in total. 59.85 per cent of them are female and 40.15 per cent are male. As can be seen in Figure 5, the speakers were born between the 1920s and the 1980s. The interviews took place in various locations all over Wales,

which is illustrated in Figure 6.¹²

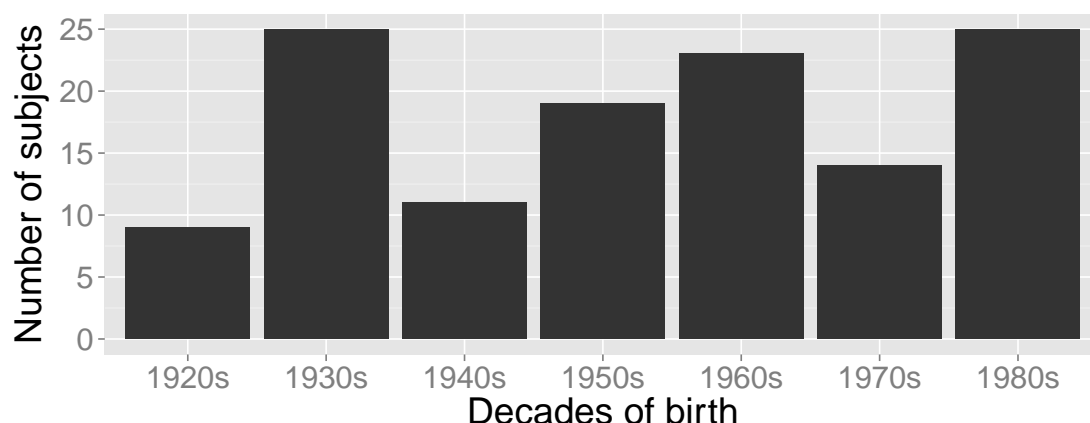


Figure 5: Numbers of speakers in the RWC according to their decades of birth. The graph excludes six speakers whose decade of birth is unknown

Before moving on to further methodological steps in the context of the present work, I would like to address the issue of corpus size. At 268,843 words, one might wonder whether the Radio Wales Corpus is large enough to investigate grammatical phenomena in Welsh English from a quantitative angle. Several sources indicate that this is indeed the case. Paulasto (2006) worked with corpora of similar sizes – the Llandybïe Corpus contains 257,000 words and the North Wales Corpus consists of 120,000 words. For both corpora, Paulasto lists a range of instances of focus fronting and non-standard habitual progressives. Apart from providing general frequency counts of the features, Paulasto’s data allow for subcategorising the occurrences of the features. For example, she provides frequencies of different sentence elements that are involved in focus fronting (such as subject complements and verb phrases), and she is able to categorise focus fronting according to different discourse functions, as responsive or emphatic (cf. Paulasto 2006: 204). Furthermore, Robinson’s grammatical commentaries on BBC Voices and MMB interviews from Wales point to a variety of non-standard grammatical features in the data (cf. British Library 2015; Robinson, Herring and Gilbert 2009–2012). For example, occurrences of the invariant tag question *isn’t it* are noted for several BBC Voices recording locations (such as Bon-y-maen, Flint, Risca).

¹²The map was created by myself using ArcGIS® software by Esri. ArcGIS® and ArcMap™ are the intellectual property of Esri and are used herein under license. Copyright © Esri. All rights reserved. For more information about Esri® software, please visit www.esri.com.

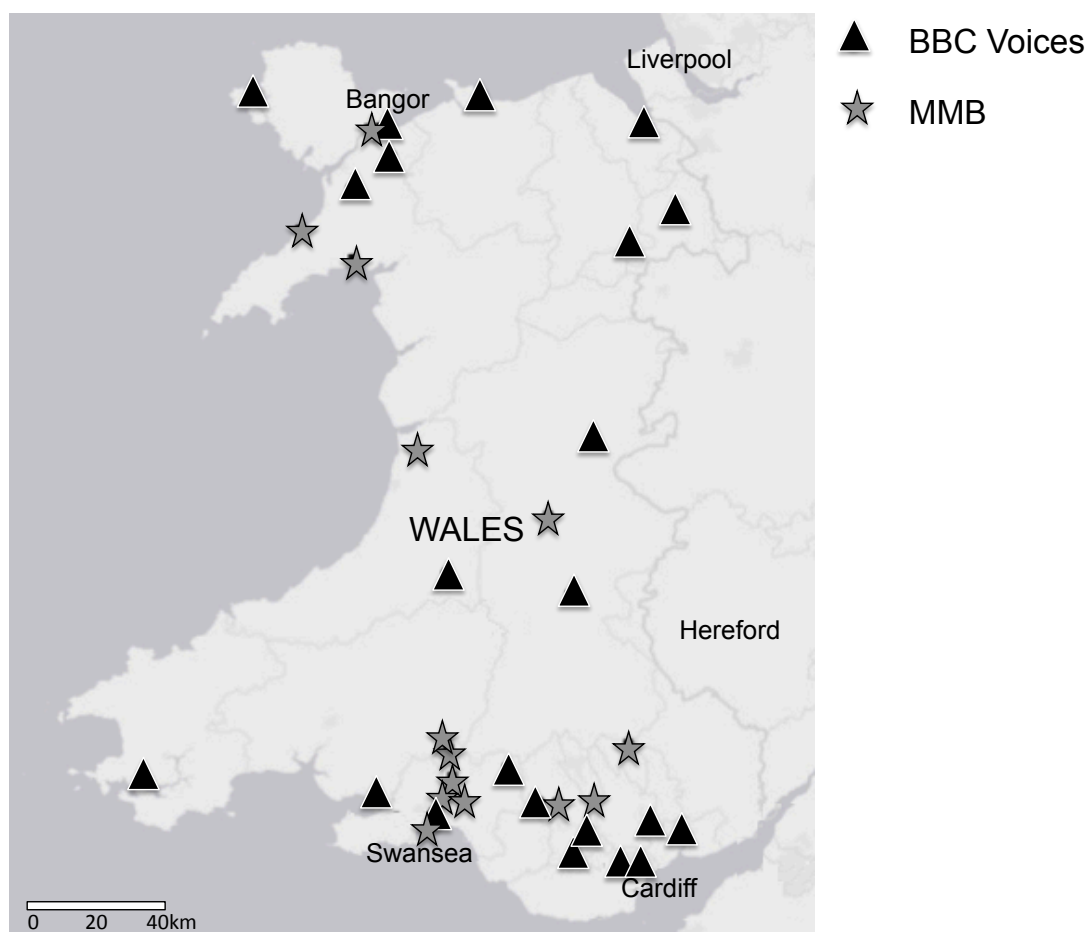


Figure 6: Map showing the places where the interviews in the RWC were conducted

Apart from Welsh English, corpora of similar sizes as the RWC have been used in quantitative approaches to other varieties of English, as in Lange's (2012) study of the syntax of spoken Indian English. Lange worked with the private face-to-face conversation data (218,531 words) from the Indian component of the International Corpus of English (ICE). In this study, Lange is able to provide frequency-based accounts of different pragmatic functions of the invariant tag questions *isn't it/is it* in Indian English (for example, confirmatory, facilitating and attitudinal). Needless to say, it would still have been preferable to work with a larger database on Welsh English in order to add more reliability to the findings (along the lines of Meyer (2002: 33): "In general, the lengthier the corpus, the better"). While that was not possible within the scope of the present work, I hope that this work raises the awareness of English spoken in Wales as

an interesting target of quantitative dialectological research, and inspires future studies which could entail the compilation of larger corpora of spoken Welsh English.

5.3.2 Linguistic Innovators corpus

Since one aim of the present study was to compare frequencies of grammatical features in Welsh English with their corresponding frequencies in London dialects of English, a corpus of spoken London English was needed. The corpus Linguistic Innovators: The English of Adolescents in London (cf. Kerswill et al. 2007) was chosen. According to my calculations, the Linguistic Innovators corpus (LI) comprises 1,112,501 words. It involves 114 speakers from two London boroughs – inner-city Hackney in London’s multicultural East End and outer-city Havering with a much lower immigration rate and located on the eastern margin of London (cf. Cheshire et al. 2011: 157).

There are several parallels between the RWC and the LI, based on which the latter can be assumed to be a suitable reference corpus for comparative linguistic analyses. (1) The interviews in the LI were conducted between 2004 and 2005, hence in the same years as the BBC Voices recordings. (2) As with the RWC, the LI recordings include group and single interviews, with a majority of group interviews (two interviewees or more). (3) Similar to the interview styles in the BBC Voices and the MMB data, the LI interviews were intended to be “relaxed conversation-like interviews” (Kerswill et al. 2007: 3). (4) From a linguistic viewpoint, the LI transcripts, like the RWC transcripts, offer specifications of the realisation of <isn’t it> – for example as <innit>.¹³

However, the LI differs from the RWC in that about 80 per cent of the words in the LI were spoken by adolescents between 16 and 19 years of age and only 20 per cent by older people born between 1914 and 1943, while the age groups are more evenly distributed in the RWC. To make the LI data more comparable to the RWC, 100 random samples of 2,000 words each were drawn from the LI, and only those samples were used for the grammatical analyses. The extent of this subset from the LI corresponds roughly to the size of the Radio Wales Corpus. 50 random samples were drawn from the younger and 50 from the older age group in the LI, with 25 samples each from

¹³Furthermore, the analyses of non-standard versus standard occurrences of features in Section 6.2 show that the RWC and the LI are sufficiently comparable regarding the grammatical features under investigation.

Hackney and from Havering. This final subset of the LI comprises 102 speakers and parts of 66 interviews.¹⁴ 48.04 per cent of the speakers are female and 51.96 per cent male.

To be sure that the subset offers stable frequency values, the frequencies of one feature – the invariant tag question *isn't it/innit* – were determined both in the random samples and in the whole LI and compared with each other. In the random samples of the younger age group, invariant *isn't it/innit* occurred with a mean frequency of 1.92 per 2,000 words and a standard error of the mean of $\sigma_{\bar{x}} = 0.48$. The mean frequency of the feature in the younger age group in the whole corpus (1.48 per 2,000 words) falls within the boundaries of the standard error of the mean. In other words, the random samples are able to predict the range in which the frequency value for the whole corpus lies. Similarly, for the older age group, the mean frequency of the invariant tag question *isn't it/innit* in the random samples (0.1 per 2,000 words) is able to predict the range in which the feature's mean frequency in the whole corpus (0.064 per 2,000 words) is located ($\sigma_{\bar{x}} = 0.043$). Based on the calculations, it can be assumed that the random samples offer frequency values sufficiently representative of the whole corpus.

5.4 Further studies

5.4.1 Analyses of TV clips

The Radio Wales Corpus was used in the present study to represent everyday spoken Welsh English. However, it is possible that not all Londoners in the salience survey were familiar with everyday spoken Welsh English, and that their perceptions of English in Wales were instead based on (staged representations of) Welsh English in the media. As discussed in Section 6.2, the use of Welsh English in the media may differ from Welsh English in everyday life, for example, in that actors and comedians use linguistic features in stylised or stereotypical ways (cf. Coupland 2001). Stylisation and stereotyping may go hand in hand with increased frequencies of use as compared to everyday Welsh English. On the other hand, Welsh presenters might tend to avoid

¹⁴The fact that most of the speakers in the LI were covered by the random samples may be due to many speakers being part of several interviews; they were interviewed several times together with different constellations of friends.

particular features in order to sound less Welsh. Hence, I considered it necessary to examine usage frequencies of Welsh English features in the media, and to contrast them with usage frequencies in the RWC.

The first step in approaching Welsh English in the media was to ask the London informants in the salience questionnaire about their familiarity with Welsh TV/radio presenters, comedians and programmes (see Section 5.2.1). The three most frequently named presenters/comedians were Rob Brydon, Rhod Gilbert and Alex Jones, as visualised in Figure 7. A range of video (YouTube) clips from different shows hosted by these presenters were selected for analysis. The show formats included, for example, magazine programmes (*The One Show*, Alex Jones) chat shows (*The Rob Brydon Show*, Rob Brydon) and stand-up comedy (different programmes by Rhod Gilbert and Rob Brydon). Additionally, the first two seasons of the series *Gavin & Stacey* were analysed, the TV programme mentioned most frequently by the London subjects. The series depicts the love story of a young couple, Stacey from Barry in Wales (played by a Welsh actress, Joanna Page) and Gavin from Essex. The show makes extensive use of explicit references to and representations of Welshness, for example, when Gavin and Stacey's families meet.

Overall, the media data were compiled in an attempt to represent the diversity of Welsh English on TV. Chris Montgomery (p.c.) also suggested to analyse different shows (instead of e.g. one show per presenter), since this may reflect more accurately what people actually perceive. While all of the programmes exhibit scripted Welsh English to some extent, there may be more room for spontaneous, natural comments in some shows (for example, in Alex Jones's conversations with studio guests in *The One Show*) than in others (as in *Gavin & Stacey*). For perceptions of Welsh English by outsiders, both more and less scripted uses of the variety in the media may play a role and are thus taken into consideration here.

In total, twenty hours of TV clips were investigated, four hours and forty minutes per presenter and six hours from *Gavin & Stacey*. An overview of all videos analysed is provided in Appendix C. Since transcribing the videos would have been disproportionately onerous, the frequency-based analyses were made by listening to the clips and noting down every instance of the grammatical features under investigation.



Figure 7: Welsh presenters and comedians named by the London subjects in the salience questionnaire. The larger the font size, the more often the name was mentioned

5.4.2 eWAVE study

The corpora just described (including the media data) provide an idea of the use of non-standard grammatical features in Welsh English and London English. However, according to my usage-based line of argumentation, it can be assumed that the regional indexation of those features is also conditioned by their pervasiveness in other British Isles dialects. This means that in order for a feature to be associated with Wales rather than, say, Ireland, the feature needs to be comparatively more frequent in Wales than in Ireland. Ideally, this hypothesis would be tested by determining the frequencies of non-standard grammatical features in corpora of (dialects of) English in Ireland, Scotland and England. Since that would have been beyond the scope of this work, the Electronic World Atlas of Varieties of English (eWAVE) was used instead (cf. Kortmann and Lunkenheimer 2013). The open-access online database was compiled mainly from 2008 to 2011, with an updated and extended release in November 2013, at the University of Freiburg (Freiburg Institute for Advanced Studies (FRIAS) and English Department) in collaboration with the Max Planck Institute for Evolutionary Anthropology in Leipzig (cf. Kortmann and Lunkenheimer 2013). It contains information on the use of 235 morphosyntactic features around the world, namely in “50 varieties of English [...] and 26 English-based Pidgins and Creoles” (Kortmann and Lunkenheimer 2013). eWAVE, thus, aims to enable “large-scale typology-style comparisons of the morphosyntactic structures of the spontaneous spoken (nonstandard) Englishes around the world” (Kortmann and Lunkenheimer 2012: 1).

For each morphosyntactic feature, the atlas provides indices of pervasiveness in the different varieties of English, pidgins and creoles. Those indices range from “A” meaning that a feature is “pervasive or obligatory” (Kortmann and Lunkenheimer 2012: 5) in a particular variety to “D” denoting a feature’s “attested absence” (Kortmann and Lunkenheimer 2012: 5).¹⁵ In the present study, the indices of pervasiveness of non-standard grammatical features in Welsh English were compared with those in other British Isles varieties of English.

It should be stressed that the pervasiveness ratings *cannot* be equated with usage frequencies. The ratings go back to judgements of scholars (there were different scholars for the different varieties) and may be subjective to some extent. According to Kortmann and Lunkenheimer (2012: 6),

in many cases, the ratings provided are impressionistic judgments by the experts, based on their own data and their (specialist and often native-speaker) knowledge of the variety. Only in some cases were larger corpora available to back up these judgments, and even then it was not possible to operationalize all the WAVE features for a corpus search.

Thus, “[w]hat looks categorical [with regard to the eWAVE ratings] can hardly be more than an abstraction of and a rough approximation to linguistic and social reality” (Kortmann and Lunkenheimer 2012: 6). Having said that, eWAVE proved to be a valuable source in the present study for obtaining a general overview of a feature’s pervasiveness across the British Isles.

5.4.3 Questionnaire: Perception of deviations from Standard English

The previous sections presented the range of methods applied to determine whether the salience judgements in the questionnaire-based survey are connected to a feature’s frequency of use or its general pervasiveness. The salience judgements could, however, also be influenced by the fact that some features are seen as larger deviations from Standard English than others, with features clearly considered as “non-standard”

¹⁵Furthermore, “X” indicates that the use of this feature is “not applicable” (Kortmann and Lunkenheimer 2012: 5) to a regional variety and “?” means that there is “no information available” (Kortmann and Lunkenheimer 2012: 5) on the occurrence of a feature in a particular variety.

probably being perceived more consciously than features that more or less conform with Standard English in people's perceptions. To test for the features' perceived non-standardness, an additional questionnaire was conceived, which was again to be distributed in Wales and London.

The design of the questionnaire, which can be found in Appendix D, was roughly based on Elmentaler, Gessinger and Wirrer (2010). Participants were presented with 32 test sentences and were advised to mark all deviations from Standard English they encountered. Sixteen of the sentences included one non-standard grammatical feature each.¹⁶ Eight of these features are to be found in Welsh English (focus fronting, invariant *isn't it* (<isn't it>), non-standard habitual progressive, inverted word order in indirect questions, zero past of regular verbs, *that*-clause replaces infinitival *to*-clause, non-standard habitual *do* and *there's* instead of *how*). These test sentences were the same as in the salience questionnaire, since I aimed to correlate the results for perceived non-standardness with those for sociolinguistic salience. Apart from the Welsh English constructions, the non-standard features in the other eight test sentences reportedly occur in varieties of English in Ireland, Scotland and England, and they had also already formed part of the salience questionnaire. These sentences served as distractors in order not to put too obvious a focus on Welsh English. The remaining sixteen sentences did not contain any deviations from Standard English and also served as distractors. They were included to prevent the subjects from thinking that there is a non-standard feature in every sentence, and thus using a strategy of marking structures in every sentence (while some of these might not have become conspicuous otherwise). According to Schütze (1996: 184), the numbers of standard and non-standard sentences in grammaticality judgement tests should be roughly the same, since otherwise the subjects "will tend to get into a yea-saying or nay-saying mode or will come to expect deviance".

In addition to those 32 test sentences, the questionnaire asked the subjects where they thought someone using the invariant tag question *innit* (<innit>) comes from. When the questionnaire was conceived, some results from the salience survey had already

¹⁶According to Elmentaler, Gessinger and Wirrer (2010: 116), the choice of one instead of several non-standard features per sentence facilitates a better control of individual features (cf. also Hettler 2014: 76).

been obtained – for example, the invariant tag question *isn't it* was found to be associated with Wales more frequently than with London or the South of England in general. Corpus analyses then suggested that this might be connected to the Londoners preferring *innit* over *isn't it*, hence I wanted to test if this preference also became apparent in people's perceptions. A short personal information section was also included in the questionnaire, based on the salience survey, inquiring, for example, into the informants' age, language background and place(s) of residence (cf. Krug and Sell 2013: 23).

The survey was conducted in July 2014 with 25 informants from Wales (Bangor) and 25 informants from London. As with the salience survey, the participants were randomly selected members of the public. In the group from Wales, 14 people were male and 11 female, and they were aged 33 years on average. Three people indicated that Welsh was their native language, three people considered both English and Welsh as their native tongues, and 19 subjects stated that their native language was English. In the London cohort, 13 participants were male and 12 were female, they were aged 31 years on average, and all of them had English as their native language.

5.5 Statistical methods

Different statistical methods were used in this work to (1) underpin methodological considerations and (2) evaluate and interpret the findings obtained from the studies discussed above. This section outlines the choices of the statistical approaches to the data. All of these approaches were carried out using R software (cf. The R Foundation 2015).

As described above, the Linguistic Innovators corpus (LI) was not analysed in its entirety in this work; instead, random samples were used. To test whether these samples were representative of the whole corpus, the standard error of the mean of occurrences of invariant *isn't it/innit* tag questions in the random samples was calculated. It was then checked whether this standard error was able to predict the range in which the mean frequency of *isn't it/innit* in the whole corpus lay. The standard error of the mean ($\sigma_{\bar{x}}$) can be defined as the standard deviation of mean values of same-sized samples of a population (cf. Bortz 2005: 90). According to Bortz (2005: 90), it is an important measure of dispersion for the estimation of population parameters (in the present study,

“population” refers to the entirety of the Linguistic Innovators corpus). The standard error has been used as a tool in corpus linguistics by other scholars. For example, Biber (1993) employs it in determining appropriate sizes of corpus samples.

Evaluating the salience test, I found that focus fronting and the invariant tag question *isn't it* were associated with Wales much more frequently than the other Welsh English features. However, what does “much more frequently” tell us? A log-likelihood test was used to statistically underpin the fact that the two features were significantly more salient than the rest of the Welsh English constructions. Similarly, log-likelihood ratios were calculated to confirm that focus fronting was significantly more frequent in Welsh English than (1) the other features in the RWC and than (2) focus fronting in London English. A log-likelihood test is a tool to determine frequency differences in the occurrence of features or combinations of features. Similar to chi-squared tests, log-likelihood measures are commonly used in corpus linguistics, for instance, in comparing frequencies of words or phrases across corpora (cf. Rayson, Berridge and Francis 2004). Unlike chi-squared tests, however, log-likelihood tests do not require normally distributed data (cf. Dunning 1993). As the restricted number of corpus hits and variables in the salience survey do not provide sufficient evidence for normally distributed data, log-likelihood ratios were preferred over chi-squared measures in the present study.

This work aims to explore to what extent salience correlates with different factors (e.g. frequency) for insiders and outsiders. To compare the two groups, correlation calculations were carried out. In a correlation calculation, the strength of a linear relation between two parameters (for example, usage frequencies from a corpus and salience values) is expressed by the correlation coefficient r , which can take a value between -1 and 1 (cf. Bortz 2005; Fischer 2010). If, for instance, frequency and salience correlate with a positive r value, this generally means that an increase of the frequency value goes hand in hand with an increase of the salience value (cf. Fischer 2010: 50). This method was chosen since it gives a relatively straightforward indication of the strength of relation between two variables. According to the R documentation (cf. The R Foundation 2015), the dataset is large enough for determining correlations using the `cor.test` function (cf. also Wenning 2014). It would still be useful to have follow-up studies in the future, validating my results with additional data for more and/or

other linguistic features.

Section 6.5 analyses how the number of Welsh English features successfully identified by informants from Wales and from London interacts with a range of social factors, such as age, language background and attitudes to linguistic diversity. Since the dataset is fairly large, involving 300 questionnaires, it was possible to work with generalised linear models.¹⁷ The models predicted the value of the dependent variable (number of Welsh features identified) based on independent variables (among others: age, sex, educational background) (cf. Gries 2009: 141). A generalised linear model gives an indication of which independent variable, as compared to the others, has the largest effect (cf. Fischer 2010: 50). This statistical method was thus a helpful tool in determining which social factors were most important regarding perceptions of Welsh English. Regression models, such as generalised linear models, have been increasingly used in recent quantitative (corpus-driven) sociolinguistic approaches (cf. Glynn 2010: 21–24). For example, Szmrecsanyi (2010) makes use of regression analyses to explore to what extent the use of genitive alternation in English is determined by language-external factors (geography, real time and text type).

5.6 Summary

This chapter has discussed the methodological steps taken in the context of the present study regarding the analysis of salience, frequency and further potential factors. Sociolinguistic salience was approached from an intralectal (people from Wales) and an interlectal (people from London) angle. By means of a questionnaire-based survey in Wales and London with 300 participants, it was determined which non-standard grammatical features are consciously associated with Welsh English. The questionnaire involved two salience tests to put the findings on a solid ground plus a personal information section, inquiring factors like age, language background and attitudes to dialectal diversity. Frequency was operationalised using corpus analyses. Due to the lack of suitable existing corpus material, I compiled a corpus of Welsh English, christened the Radio Wales Corpus (RWC). The RWC encompasses 268,843 words and contains interviews recorded in various locations across Wales between 1999 and 2005. The

¹⁷Unfortunately, it was not possible to explore the relation between salience and corpus frequencies with such regression analyses as the dataset was too small (cf. Castelloe and O'Brien 2000).

Linguistic Innovators corpus (LI, cf. Kerswill, Cheshire et al. 2007) was used as a reference corpus of London English. Of this corpus, which contains interviews conducted in 2004 and 2005, random samples totalling 200,000 words were analysed. Moreover, 20 hours of TV shows with Welsh presenters, comedians and actors were scrutinised to find out about differences between Welsh English as used in the media and in everyday life (as approximated by the RWC). To explore the pervasiveness of grammatical features in the whole of the British Isles, the Electronic World Atlas of Varieties of English (eWAVE) was consulted (cf. Kortmann and Lunkenheimer 2013). An additional questionnaire-based survey investigated to what extent different grammatical features are perceived as deviations from Standard English. This survey, involving 50 participants from Wales and London, also explored where in the British Isles subjects position a speaker using the invariant tag question *innit*. Several statistical methods were applied in underpinning methodological steps and in evaluating the results of the studies. Among these methods were (1) the calculation of the standard error of the mean to determine whether the random samples from the LI were representative of the whole corpus, (2) log-likelihood tests to find out if the salience and frequency values of focus fronting and the invariant tag question *isn't it* were significantly higher than those of the other features, (3) correlation calculations to determine relations between salience and potential factors for insiders and outsiders and (4) generalised linear models to find out about the effects of different social variables on the perception of Welsh English grammar.

Chapter 6

Results

This chapter presents and discusses the findings from the analyses carried out in the present study. While Section 6.1 deals with the results from the salience survey, depicting how the grammatical features under investigation were perceived by the Welsh and London informants, Sections 6.2 to 6.5 examine potential determinants of salience. These are (1) frequency, (2) regional dispersion, (3) perceived degree of non-standardness and (4) social factors.

Note that the term “salience” in this chapter always refers to the *regional indexation* of linguistic features, i.e. the degree to which features are consciously associated with a particular regional language variety (for example, Welsh English; see Chapter 2). In case discussions in the following sections also concern other forms of salience, such as cognitive salience, this is made clear in the text. Moreover, if not specified differently, the notation *isn't it* refers to the phonological realisation /'ɪzənt ɪt/, and the notation *innit* to /'ɪnɪt/ (see Chapter 3). In the context of the salience survey, *isn't it* refers to the orthographic notation <isn't it>.

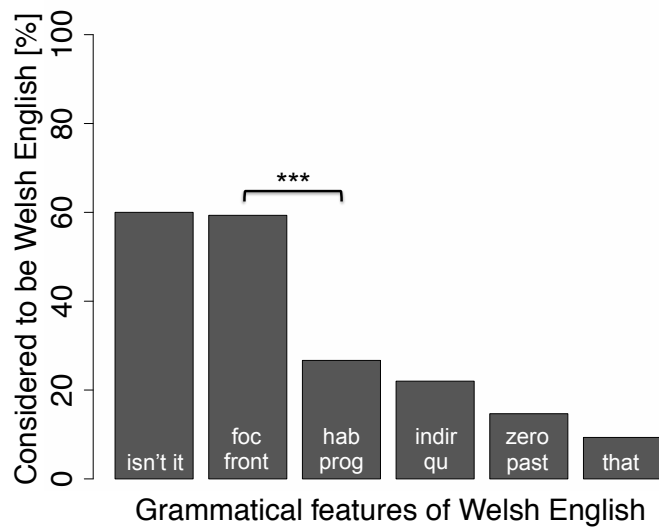
6.1 Salience in Welsh English grammar

The questionnaire-based salience survey, as described in detail in Section 5.2, involved a main salience test (test I) plus a smaller additional test to support the results (test II). The following findings were obtained from test I, where subjects (150 from Wales and 150 from London) had to assign written sentences with non-standard grammatical features to different English-speaking areas. For the informants from Wales, the

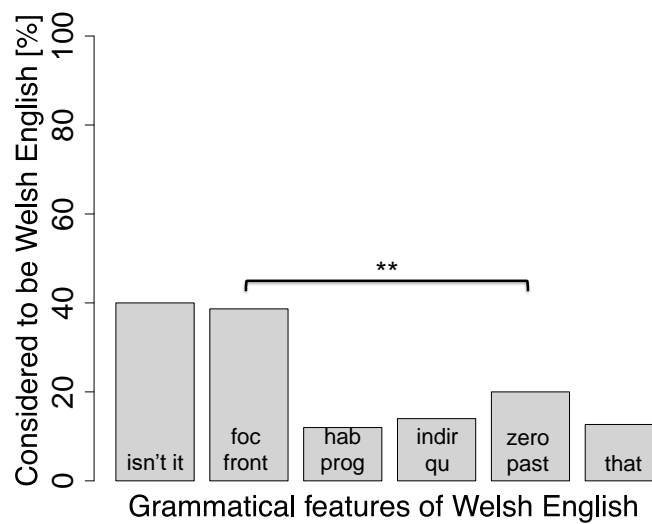
most salient of the six Welsh English features explored were the invariant tag question *isn't it* and focus fronting. These two, as can be seen in Figure 8, were considered to be Welsh English much more often than the other features. The difference in salience between these highly salient features and the others is statistically significant ($p < 0.001$), according to a log-likelihood test.

For the informants from London, similarly, the most salient features of Welsh English were the invariant tag question *isn't it* and focus fronting. The salience ratings for these two features in the London group were lower than in the Wales group, which is no surprise from a frequency-based point of view: the Londoners were exposed to Welsh English to a considerably lower degree in their everyday lives, thus the features relevant can be assumed to be less mentally entrenched. This is in line with Williams, Garrett and Coupland (1999: 357), who suggest that lower recognition rates are connected to less experience with the target dialect. “[L]ess access to dialect speakers, face-to-face or in the broadcast media” may result in “less accurate or less detailed cognitive templates” (Williams, Garrett and Coupland 1999: 357) of a dialect. Nevertheless, also for the London informants, the salience values for the invariant *isn't it* and focus fronting were significantly higher ($p < 0.01$) than for the other features under investigation.

To strengthen these results of test I, salience test II was carried out, using a different test method and different test sentences (for a detailed description of test II, see Section 5.2). The subjects were presented with several sentence pairs and had to choose the one sentence in each pair which was used by a speaker from Wales (while in fact both sentences could be used in Welsh English). The results of this test confirmed central tendencies of test I. For the informants from Wales, the invariant tag question *isn't it* and focus fronting were again the most salient of the six features. Once more, there was a statistically significant difference between the salience of these constructions and the others ($p < 0.05$). Table 3 contrasts the salience values for the judges from Wales in both tests. The figures in the table refer to the percentages of subjects that assigned a feature to Welsh English (for example, focus fronting was considered to be a feature of Welsh English by 59.33 per cent of subjects in test I and by 61.95 per cent in test II). The fact that the figures were generally higher in test II was no surprise since there every feature had a 50 per cent chance of being selected as “Welsh English”, while



(a) Informants from Wales (n = 150)



(b) Informants from London (n = 150)

Figure 8: Salience values of Welsh English features, obtained in the first test of the questionnaire-based survey

in test I it was not required to consider any of the features to be “Welsh English”. In the group of Londoners, test II also found focus fronting and the invariant *isn't it* to be more conspicuous than the other four constructions (however, the difference was not statistically significant). All in all, the results from test II strengthen the findings of test I. The parallels between the test outcomes suggest that the salience values in test I are not (necessarily) tied to the particular method applied or the test sentences involved.

	Test I	Test II
Invariant tag question <i>isn't it</i>	60	65.49
Focus fronting	59.33	61.95
Non-standard habitual progressive	26.67	38.94
Inverted word order in indirect questions	22	40.71
Zero past tense of regular verbs	14.67	20.35
<i>that</i> -clause instead of infinitival subclause	9.33	23.89

Table 3: Salience values in per cent for subjects from Wales in tests I and II

The fact that the *isn't it* construction emerged as being such a salient feature of Welsh English in both tests is striking, since the feature is reported to occur in London varieties of English, too (typically realised as *innit*, see Chapter 3). However, in salience test I, the Londoners assigned the feature to Wales to a higher degree than to the South of England or to London in particular. In general, Wales was the location named most regarding *isn't it* – both by the subjects from Wales and by the subjects from London. These results for the subjects from London are displayed in Figure 9.

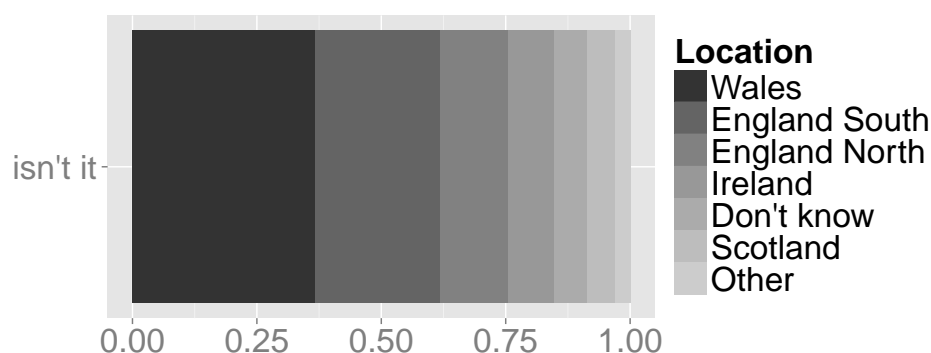


Figure 9: Degrees to which the invariant tag question *isn't it* was assigned to different regions in the British Isles by subjects from London

A potential explanation for the “Welshness” of the *isn’t it* construction, and of focus fronting, could lie in their frequencies of use. As discussed before, in the framework of Exemplar Theory, we can assume that listeners are sensitive to the statistics of their environment. It is plausible that if some Welsh English features are highly frequent, this multitude of exemplars is registered, and the features become mentally entrenched. Moreover, the constructions could be stored alongside some contextual information – such as a particular group of dialect speakers who typically use the construction. The following sections test the potential and limits of a frequency-based approach to explaining saliency. For respective calculations, the saliency values of test I are used. These values proved more conclusive than the results from test II, since in the former test no indications were given that any of the non-standard features could potentially be Welsh English. Test II was more restricted, only offering the options to consider a feature Welsh English or not, which might have led to the higher saliency values for all features. (Still, the differences in the preferences of features and the parallels between the two studies suggest that the subjects’ choices in test II were not random.)

6.2 Saliency and frequency

In the present study, frequency was approached from several perspectives. First, it was explored to what extent high token frequencies of non-standard features in Welsh English went hand in hand with high degrees of saliency. Second, the features’ frequencies in Welsh English were compared with those in London English to find out if frequency differences between the two varieties had affected the judgements of Londoners. Third, the features’ frequencies in Welsh English were contrasted with frequencies of functionally equivalent Standard English constructions as a basis for drawing more reliable frequency-related conclusions across different kinds of corpora. These corpus-based studies were followed, fourth, by an analysis of grammatical features used in TV programmes by Welsh presenters, comedians and actors. This was based on the assumption that not all of my London informants had face-to-face contact with Welsh people, and that all they knew about the English spoken in Wales may have been based on listening to Welshmen and -women (or performances/blatant stagings of Welshness) in the media. This fourth study, then, was an attempt to draw a more accurate picture of people’s actual perceptions of language structures. All in all, these four studies

aimed to show that salience has a quantitative side.

Throughout this chapter, the following terminology is employed for different forms of frequency (cf. Gries 2010): *Absolute frequency* refers to a feature's raw frequency of occurrence, i.e. its number of hits in a corpus. *Relative frequency* denotes normalised frequencies. These can be obtained by dividing a feature's number of hits (absolute frequency) by the total number of words in a corpus and then multiplying the result by e.g. 10,000 or 1,000,000. In the context of the present study, the corpus frequencies were normalised per 10,000 words. Since such relative frequencies facilitate comparisons between corpora of different sizes, they were the preferred frequency measure in the analyses presented below.

6.2.1 Frequency in Welsh English

In Chapter 2, the assumption was formulated that a linguistic form can be salient, i.e. at the centre of attention, when it is highly frequent. A grammatical construction in Welsh English may become entrenched in people's minds because they have often encountered the feature in their everyday lives. The impression that a feature is encountered "often" is necessarily connected to the fact that other features are felt to occur less often. As is the case with salience, the perception of frequencies depends on the context. Thus, in determining the token frequencies of Welsh English features in the present study, the individual frequencies were not rated as being "high" or "low" in general. Instead, the frequencies were related to each other and considered as "higher" or "lower" than the other frequency values. This seemed to be particularly important since all features explored were *relatively infrequent* in comparison with other English grammatical features, such as the simple past of regular verbs indicated by the word-final *-ed*. Thus, when talking about features of a "high frequency" in this study, this always means that the frequencies are high as compared to the other features under investigation.

The Radio Wales Corpus (RWC) was consulted to determine the token frequencies of the six Welsh English grammatical features. It should be noted that these corpus frequencies can only be an approximation of the *real* usage frequencies in Welsh English; frequency values will clearly vary between corpora. Still, the RWC was constructed in an attempt to be a good indicator of present-day Welsh English by covering a broad

range of regions and social groups in Wales.

By reading through the corpus texts, the raw frequencies of occurrence of the Welsh English grammatical features were determined. These absolute frequencies were then normalised per 10,000 words in order to facilitate frequency-based comparisons with other corpora.¹ The most frequent feature in the RWC was focus fronting with 4.58 tokens per 10,000 words (123 occurrences in total). The least frequent feature was the *that*-clause instead of the infinitival subclause, which went undetected in the RWC. Table 4 summarises the absolute numbers of occurrence and the frequencies normalised per 10,000 words for all constructions under investigation.²

	Frequency absolute	Frequency per 10,000 words
Focus fronting	123	4.58
Non-standard habitual progressive	71	2.64
Invariant tag question <i>isn't it/innit</i>	12	0.45
Inverted word order in indirect questions	11	0.41
Zero past tense of regular verbs	4	0.15
<i>that</i> -clause instead of infinitival subclause	0	0

Table 4: Absolute and normalised token frequencies of Welsh English features in the Radio Wales Corpus. The two most salient features, focus fronting and invariant *isn't it/innit*, are highlighted in bold print

Taking a look at the most salient features, a heterogeneous picture emerges. In line with my usage-based hypotheses, focus fronting is comparatively frequent. It is significantly more frequent than the other constructions under investigation, according to a log-likelihood test ($p < 0.001$). The invariant tag question *isn't it/innit* has a much lower probability of occurrence, though. Focus fronting appears ten times as often as the tag question in the RWC (123 versus 12 hits). The invariant *isn't it/innit* is also less frequent than the non-standard habitual progressive, although the latter received considerably lower salience values. The frequency difference between the two sali-

¹Other scholars such as Szmrecsanyi (2013) and Paulasto (2006) also work with frequencies per 10,000 words. The frequency values presented here can thus be directly compared with Szmrecsanyi's findings on occurrences of non-standard morphosyntactic features across the British Isles, and with the frequencies of grammatical Welsh English features as depicted by Paulasto.

²All but one instance of the invariant tag question in the RWC were realised as *isn't it*, while the remaining one was realised as *innit*.

ent features in Welsh English can be substantiated with data from the FRED corpus. As illustrated in Section 3.2, focus fronting is more than ten times as frequent in the FRED Wales data than invariant *isn't it/innit* tags (10.37 vs. 0.9 occurrences per 10,000 words).³ This means that high salience is mirrored in frequency for one feature, but not for the other. The finding suggests that while high token frequency may contribute to an item's salience, it is no general precondition for regional indexation in morpho-syntax.

These general tendencies seem to apply to perceptions of both insiders and outsiders. However, assuming that Welsh people have more experience with Welsh English, we may expect some usage-based differences in the relation between frequency and salience for Welsh people and Londoners when analysing the data in more detail. Figure 10 displays the frequency values of the non-standard features in the RWC, ordered by the features' salience for the subjects from (1) Wales and (2) London. The data indicate that the salience perceptions of the Welsh subjects are more in line with frequency than those of the Londoners. This becomes evident most prominently with regard to two features: non-standard habitual progressives and zero past tense forms of regular verbs. Non-standard progressives are the second most frequent feature in the RWC and the third most salient item for the insiders in the salience survey. In the group of the outsiders, however, the feature receives the lowest salience rating, even lower than that of the *that* replaces *to* construction, which is absent in the RWC. The rarity of the zero past tense in Welsh English is reflected in its rather low salience rating by the Welsh people, but less so in the Londoners' perceptions: For them, this non-standard past tense construction constitutes the third most salient feature.

Two correlation calculations helped underpin these observations quantitatively. The frequency values from the RWC were related to the salience values from the questionnaires. In the tests used, the strength of association between frequency and salience

³Furthermore, there are some parallels between the frequencies of non-standard features in the RWC and Paulasto's corpora. As discussed in Section 3.2, Paulasto (2006) found focus fronting to occur 3.17 times per 10,000 words in the North Wales Corpus (NWC) and 4.66 times per 10,000 words in the Llandybïe Corpus (LC). In the RWC, the construction emerged with a relative frequency of 4.58 per 10,000 words. Non-standard habitual progressives appeared in the NWC 1.33 and in the LC 2.6 times per 10,000 words, while their frequency per 10,000 words was 2.64 in the RWC. In the study by Meriläinen and Paulasto (2014) using the Llandybïe Corpus, the frequency per 10,000 words of the inverted word order in indirect questions was 0.62, compared to 0.41 in the RWC.

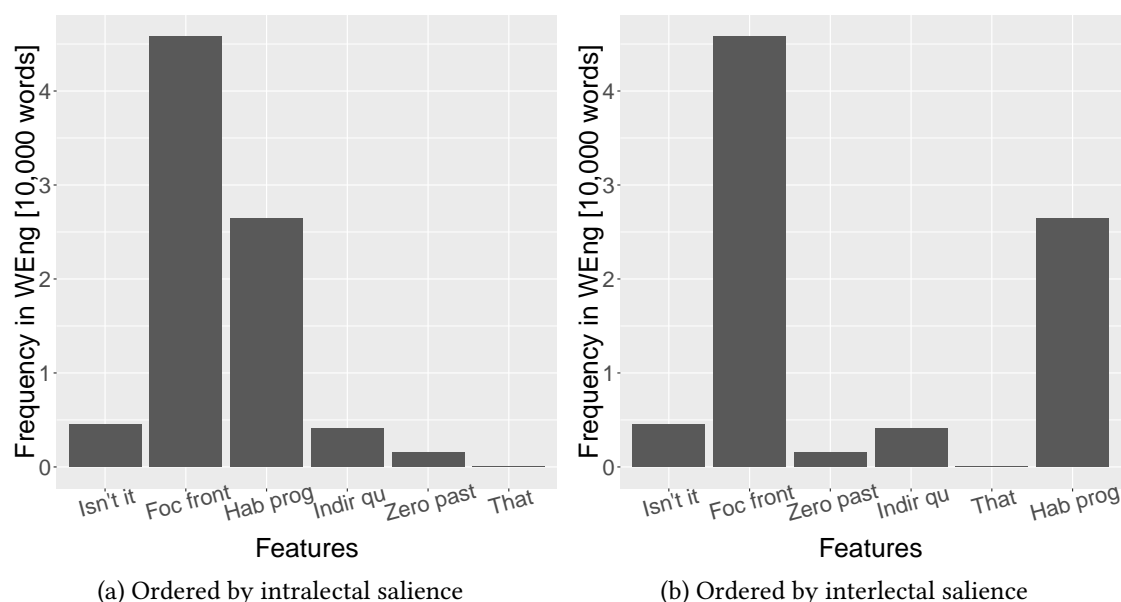


Figure 10: Token frequencies of Welsh English features in the RWC. The features are ordered by their saliency, from the most to the least salient item

is indicated by Pearson's product-moment correlation coefficient r (cf. Baayen 2008: 98). The calculations aimed to determine the correlation between frequency in Welsh English and saliency for (1) people from Wales and (2) people from London. The tests found a stronger correlation between frequency and saliency for the insiders ($r = 0.57$) than for the outsiders ($r = 0.39$). This means that the trend for more frequent features to be more salient is clearer in the group from Wales. This again indicates that experience is crucial: subjects who are more experienced with Welsh English (i.e. their native dialect that surrounds them in everyday life) have more accurate perceptions of regional grammatical features.

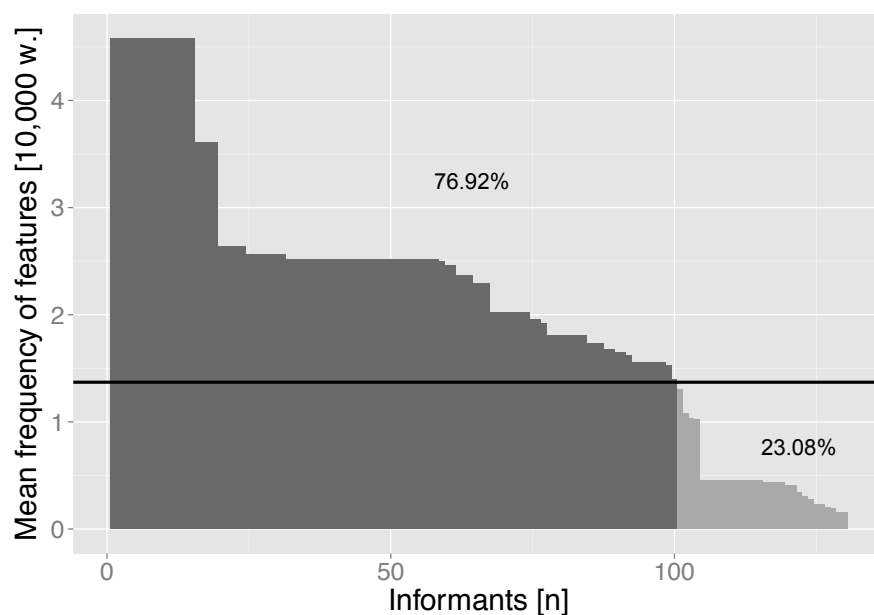
The calculations show how saliency perceptions of *average subjects* correlate with token frequency. They do not show, however, how many of the informants recognised the more frequent features rather than the less frequent ones. Thus, in a next step, frequency was related to the perceptions of the individual test subjects. The mean frequencies of the features that each individual informant had considered to be Welsh English were calculated. So, for example, if someone had assigned focus fronting and the invariant *isn't it* to Wales, which occurred 4.58 and 0.45 times per 10,000 words in the RWC, the mean frequency of these constructions would be $(4.58 + 0.45) / 2 = 2.52$.

If someone had assigned all of the six Welsh English features to Wales, the mean frequency per 10,000 words would be $(4.58 + 2.64 + 0.45 + 0.41 + 0.15 + 0) / 6 = 1.37$. This means that if a frequency effect holds, we should expect the majority of subjects to assign those features or combinations of features to Wales that have a mean frequency higher than 1.37. And indeed, this was the case in the present study. Of the 130 subjects from Wales that had at least considered one of the six Welsh English features as Welsh,⁴ 76.92 per cent chose features with a (mean) frequency higher than 1.37. The remaining 23.08 per cent recognised Welsh English features with (mean) frequencies of 1.37 or lower. This means the choices of the Welsh subjects were biased towards the more frequent features. As for the informants from London, the trend goes in the same direction, although being somewhat less pronounced. Of the 111 Londoners who had at least assigned one of the six features to Wales, 56.76 per cent characterised features with higher (mean) frequencies than 1.37 as Welsh English, while 43.24 per cent chose less frequent constructions.

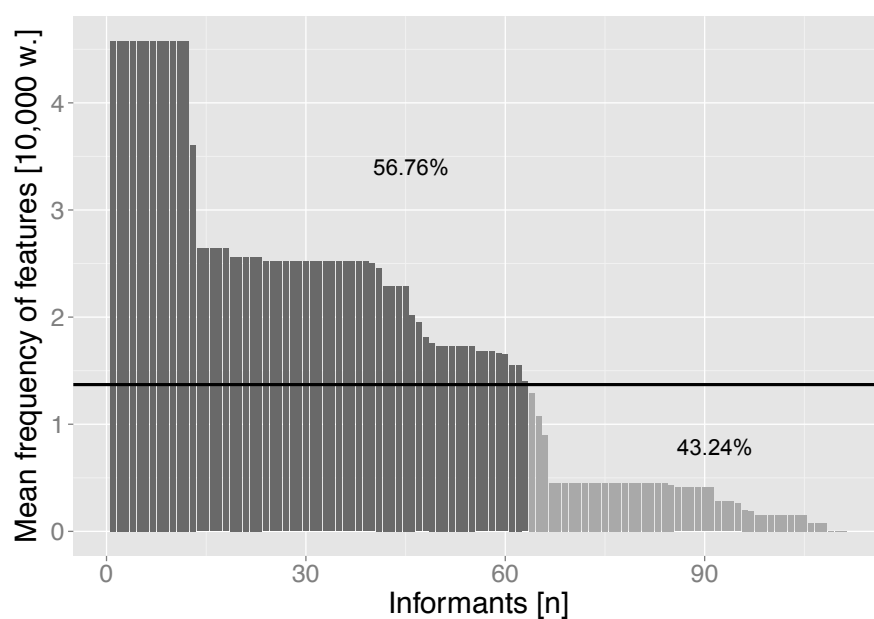
In sum, the gains from these subject-sensitive analyses, illustrated in Figure 11, are twofold. (1) They offer proof for a positive correlation between frequency in Welsh English and (intra- and interlectal) salience by showing that the majority of subjects recognised the more frequent Welsh English features. (2) However, the analyses also point to interindividual differences in the subjects' perceptions, which are not obvious from the general trends provided by the correlation calculations discussed above. In both groups, and particularly in the London cohort, there is variation regarding the features assigned to Wales, for example in that not all subjects recognised the frequent Welsh English features. As discussed in detail in Section 6.5, it is important for explanatory approaches to sociolinguistic salience to take such individual perceptions into consideration.

Although the studies described in previous paragraphs provide some evidence for the usage-based nature of dialect perceptions, there are limitations to the approach. Token frequencies in Welsh English alone cannot explain the conspicuousness of the invariant tag question *isn't it*. The feature was highly salient in the questionnaire survey. However, in comparison with the other features, it did not occur very frequently in the

⁴The remaining 20 subjects were excluded from these calculations, since assigning no feature to Wales does not offer any frequency-related information.



(a) Informants from Wales (n = 130)



(b) Informants from London (n = 111)

Figure 11: Percentages of subjects who recognised Welsh English features with mean frequencies higher than or equal to/lower than the mean frequency of 1.37 per 10,000 words (indicated by the black lines)

RWC. Focus fronting, for example, received about the same salience values as the *isn't it* construction, but was significantly more frequent in the corpus. Another limitation to merely working with frequencies in Welsh English is that the linguistic backgrounds of the Londoners get left out. One can assume that probabilities of occurrence in Welsh English have some influence on the perceptions of Londoners, but this may not be the full picture. Those perceptions could also be shaped by the Londoners' own uses and non-uses of particular features. To address these limitations, it is worthwhile to consult a database with present-day London English to find out to what extent frequency differences between Welsh English and London English have an impact on salience.

6.2.2 Frequency in Welsh English vs. London English

As discussed in Chapter 2, Rácz (2013) offers some empirical evidence for the influence of frequency differences between two varieties on salience. He found that the high salience of three phonetic-phonological features as perceived by people from outside the respective dialect areas or social classes went hand in hand with low usage frequencies of these features in the groups of the outsiders. The features were interlectally salient, according to Rácz, because they carried high degrees of “surprisal” (i.e. unexpectedness, low probabilities of occurrence in the outsiders' dialects). In the case of my subjects from London, this could mean that they find those features particularly conspicuous that are not very likely to be used in their own dialect.

To explore the impact of low probabilities of occurrence, the normalised frequencies of the six morphosyntactic features in the RWC were contrasted with the normalised frequencies of the same features in the Linguistic Innovators corpus (LI). The following results were obtained from the analysis of the LI, summed up in Table 5 and Figure 12. As expected from the literature, focus fronting was used to a much lower degree in the London corpus (0.95 per 10,000 words) than in the Welsh database (4.58 per 10,000 words). The difference in frequency values between the corpora is statistically significant, according to a log-likelihood test ($p < 0.001$). This offers an additional explanation for the high interlectal salience of the feature: it may be conspicuous for the Londoners as it differs from what they are used to in their own variety.⁵ The inverted word

⁵The fact that focus fronting occurs *at all* in the London data is not surprising. Paulasto (2006: 214) also records occurrences of focus fronting in the Survey of English Dialects (SED) material. However, in her Welsh corpora, the feature turns up more frequently. She concludes that what makes focus fronting

order in indirect questions, which had a relatively low saliency in the questionnaires, occurred with nearly the same frequencies in the LI and the RWC. The *that* construction, also of a low saliency in the questionnaires, could not be found at all in the LI, which was the same for the RWC. Non-standard habitual progressives, receiving the lowest interlectal saliency value, were even more frequent in the LI than in the RWC (the frequency difference between the corpora being statistically significant with $p < 0.001$). This higher probability of occurrence in London English might prevent the feature from becoming entrenched as a Welsh English item for outsiders. A similar effect may apply to zero past tense forms of regular verbs. They were also more frequent in the LI than in the RWC and considerably less salient than focus fronting and invariant tags to the Londoners.⁶ Overall, these findings offer support for the assumption that frequency differences between varieties can predict interlectal saliency to some extent.

	RWC (Wales) Frequency per 10,000 words	LI (London) Frequency per 10,000 words
Focus fronting	4.58	0.95
Invariant tag question <i>isn't it</i>	0.41	0
Invariant tag question <i>innit</i>	0.04	5.05
Inverted word order in indirect questions	0.41	0.4
<i>that</i> -clause instead of infinitival subclause	0	0
Zero past tense of regular verbs	0.15	1.05
Non-standard habitual progressive	2.64	4.6

Table 5: Normalised token frequencies of non-standard features in the Radio Wales Corpus and the Linguistic Innovators corpus. The two most salient features, focus fronting and invariant *isn't it*, are highlighted in bold print. The frequencies of the invariant tag question *innit* are listed separately

The situation for the invariant tag question *isn't it/innit* is more complex. Unsurprisingly, the construction could be found – quite frequently – in the London corpus. Nevertheless, the phonological realisation of the feature differed between the LI and the

so characteristic of Welsh English is not that it is categorically restricted to Wales, but that it is used more frequently and with more pragmatic functions there than in other British Isles varieties. “[F]requent use is another characteristic of the Welsh use of FF [focus fronting]” (Paulasto 2006: 214).

⁶Still, the interlectal saliency values for the zero past tense were higher than those for non-standard habitual progressives. The question to what extent these saliency differences are related to structural conspicuousness is approached in Section 6.4.

RWC. The Londoners in the LI pronounced all instances of the construction as *innit* (/ˈɪnɪt/). By contrast, *isn't it* (/ˈɪzənt ɪt/) was employed for all instances but one in the RWC. The data suggest that there are different preferences regarding the use of invariant tag questions in Welsh English and London English. This could be connected to salience in that *isn't it* (/ˈɪzənt ɪt/) is more characteristic of Welsh English, while the invariant tag *innit* (/ˈɪnɪt/) is a salient feature of London speech (or more generally of English in Southern England).

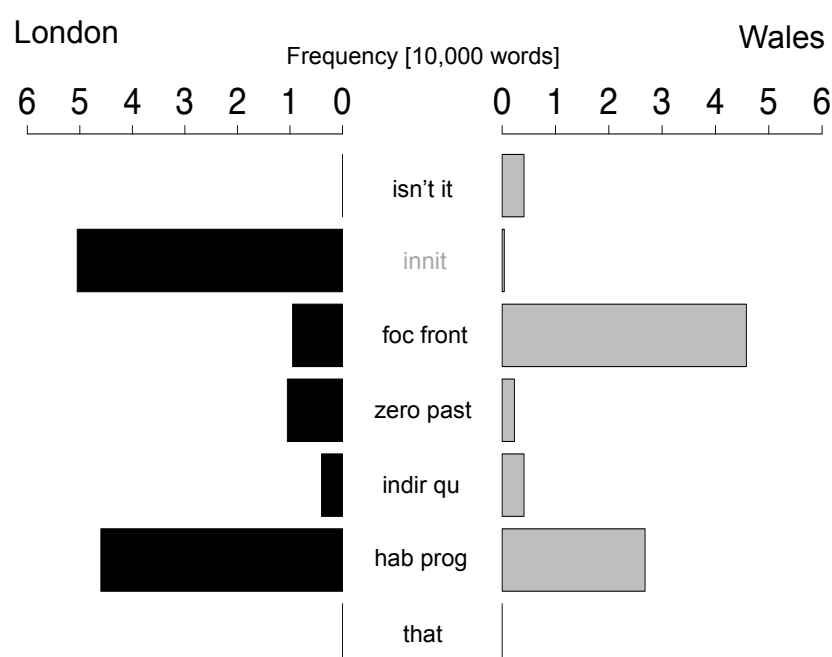


Figure 12: Token frequencies of morphosyntactic features in London English (Linguistic Innovators corpus) and Welsh English (Radio Wales Corpus)

This assumption is strengthened by data from an additional questionnaire-based survey conducted in the context of the present study with 25 subjects from London and 25 subjects from Wales (for a more detailed description of the questionnaire, see Section 5.4.3). The subjects were shown a sentence involving the invariant tag question *innit* and had to say where they thought the speaker came from. London was named by seventy per cent of the informants, while Wales was only referred to in two per cent of the cases. By contrast to that, as shown above, the most frequently named location regarding the invariant tag question *isn't it* in the salience questionnaire was Wales. It seems that the preferential differences of using *isn't it* and *innit* are reflected

in the subjects' perceptions of Welsh English and London English. In sum, the findings suggest that, despite the rather low frequencies in the RWC, the *isn't it* (/ˈɪzənt ɪt/) construction is a salient feature of Welsh English, as it occurs with higher frequencies in Welsh English than in London English. Strictly speaking, the regional indexation of the invariant *isn't it* as a Welsh English feature is not attributable to the feature's morphosyntactic characteristics, but to phonetic-phonological properties.

A correlation calculation was carried out to quantify the relation of frequency differences between varieties and interlectal saliency. It explored the strength of association between the saliency values for the Londoners and the frequency differences in Welsh English versus London English for the six morphosyntactic features (excluding realisations of *innit*). Relative frequency differences were determined by subtracting the frequencies (normalised per 10,000 words) in the LI from those in the RWC. For focus fronting, for example, the relative frequency difference would be $4.58 - 0.95 = 3.63$. The test found a positive correlation of relative frequency differences and interlectal saliency with $r = 0.72$. As this correlation is distinctly stronger than the one between frequencies in the RWC and interlectal saliency, the data suggest that for the group of Londoners, the frequency differences are a better predictor of saliency than frequencies in Welsh English alone.

The importance of frequency differences between varieties also becomes visible when relating the relative frequency differences to the individual perceptions of the subjects from London. The same subject-sensitive procedure as for frequency in the RWC was applied, determining the mean frequency difference of all features considered to be Welsh English by each individual informant. For example, if someone assigned focus fronting and the inverted word order in indirect questions to Wales, the mean frequency difference would be $(3.63 + 0.01) / 2 = 1.82$. In case all of the six features were considered to be Welsh English, the mean frequency difference would be $(3.63 + 0.41 + 0.01 + 0 - 0.9 - 1.96) / 6 = 0.2$. Consequently, if the informants' choices are generally tied to frequency, we should expect the majority of subjects to select features with (mean) frequency differences higher than 0.2. Indeed, 69.37 per cent of the Londoners selected Welsh English features with higher mean frequency differences than 0.2, while for 30.63 per cent, the frequency differences of the features amounted to 0.2 or less. The results, illustrated in Figure 13, are in line with the correlation calculation.

They indicate that frequency differences between Welsh English and London English can predict interlectal salience somewhat more accurately than frequencies in Welsh English alone (although the latter seem to play a role, too). Still, similar to Figure 11, the graph also points to between-subject variation in the recognition of different dialectal features.

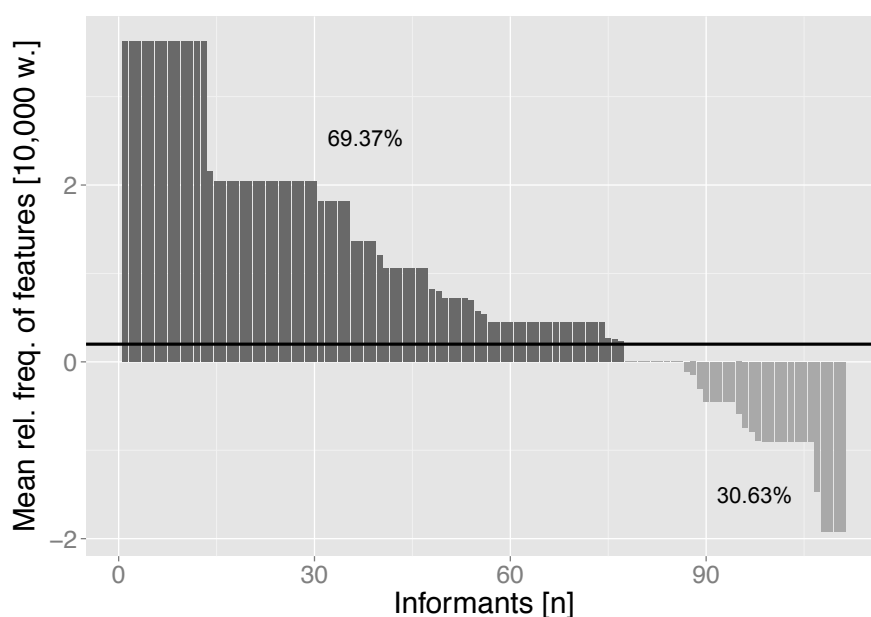


Figure 13: Percentages of subjects from London ($n = 111$) who recognised Welsh English features with mean frequency differences higher than or equal to/lower than the average value of 0.2 per 10,000 words (indicated by the black line)

In a larger sense, such frequency differences could also have an impact on the perceptions of people from Wales. The impression that a linguistic structure is common in their own dialect may be strengthened by the feeling that it is more frequent in their own dialect than somewhere else, for example, in the rest of the British Isles. The relation between salience and a feature's pervasiveness in the British Isles is investigated in Section 6.3. Before analysing other factors, though, it seems necessary to further scrutinise the corpus material of the RWC and the LI. A challenge in working with such databases is that the frequencies determined might be tied to the characteristics of the individual corpora. For example, if a past tense feature occurs very frequently, this might simply be because the people in the interviews talk about the past most of the time. To find out more about the uses of features *in context*, Section 6.2.3 contrasts the

probabilities of occurrence of the non-standard grammatical features under investigation with the frequencies of functionally equivalent Standard English constructions.

6.2.3 Frequency in Welsh English vs. Standard English

Dividing the investigated instances according to their standard or non-standard status in mainstream English is a method which has its advantages and disadvantages. Applied consistently, it is capable of demonstrating differences between corpora [...]. It does, however, also project the notion that this dividing line is somehow stable and absolute, which it is not. In all likelihood there are some other ways in which the [...] categorisation could have been carried out, yielding results that would have been equally justified (Paulasto 2006: 268).

In line with Paulasto's statement, I would like to stress that using the labels "standard" and "non-standard" in the following section is a simplification for the purpose of comprehensibility, but I am aware that (1) such categorisations are always somewhat subjective and that (2) there may be features that can neither clearly be characterised as "standard" nor as "non-standard".

As shown above, the invariant tag questions *isn't it/innit* occurred much more often in the Linguistic Innovators corpus than in the Radio Wales Corpus. One could wonder now if this was merely the case because tag questions are generally more frequent in the LI than in the RWC. I thus counted the frequencies of standard tag questions with the subject and the operator of the tag mirroring the subject and the operator of the main clause, as in *She is nice, isn't she?* (cf. Quirk et al. 1985: 810). In the LI, such standard tags occurred with a normalised frequency of 10.3 per 10,000 words, while the invariant *innit* (no occurrences of the invariant *isn't it*) appeared 5.05 times per 10,000 words. In the RWC, standard tags were slightly less frequent with 8.18 occurrences per 10,000 words. However, since the invariant tags *isn't it/innit* only occurred 0.45 times per 10,000 words in the RWC, they were clearly less frequent than *innit* in the LI in comparison with the standard tags. The results are thus in line with the tendencies of the non-standard features' frequencies.

In a second step, four types of tag questions involving *isn't it* or *innit* were contrasted: (1) use of *isn't it* in accord with the main clause (*It's nice, isn't it?*), (2) use of *isn't it* not in accord with the main clause (*She likes him, isn't it?*), (3) use of *innit* in accord

with the main clause (*It's nice, innit?*) and (4) use of *innit* not in accord with the main clause (*She likes him, innit?*). In the Radio Wales Corpus, *isn't it* or *innit* tag questions in accord with the main clause were clearly more frequent with 4.17 occurrences per 10,000 words than their counterparts not in accord with the main clause (0.45 occurrences per 10,000 words). By contrast to that, in the LI, the tags not in accord with the main clause (always *innit*) were more frequent than *isn't it* or *innit* in accord with the main clause, the former occurring 5.05 times, the latter 2.9 times per 10,000 words. This further strengthens the impression that the invariant *innit* is very frequent in the London data. What the analysis also showed was that for the tags in accord with the main clause, *isn't it* was clearly preferred over *innit* in the RWC (3.79 versus 0.37 occurrences per 10,000 words), while it was the other way around in the LI (*innit* occurring 2.45 times, *isn't it* only 0.45 times per 10,000 words). These results, displayed in Figure 14, underpin the above finding that speakers of Welsh English seem to prefer using *isn't it* in tag questions, while Londoners favour *innit*. Thus, the high salience of the invariant *isn't it* in Welsh English might be reinforced by the fact that Welsh people generally prefer to employ *isn't it* in tag questions, while *innit* is the more common way of realisation in London speech.

Focus fronting was much more frequent in the RWC than in the LI. Moreover, among the six features under investigation, focus fronting had the highest frequency value in the Radio Wales Corpus (4.58 per 10,000 words). This fits together with the feature's high salience in Welsh English, both in the questionnaires from London and in those from Wales. However, up to this point, it is not clear whether focusing devices are generally used more frequently in the RWC than the LI, which would relativise the pervasiveness of focus fronting in the Welsh data. Since there is no direct equivalent of focus fronting in Standard English, I decided to compare the feature with clefts (*It was an apple he got*) and pseudo-clefts (*What he got was an apple*). As stated in Chapter 3, clefts and pseudo-clefts are considered to be functionally equivalent to (cf. Thomas 1997: 78) or at least to have many characteristics in common with (cf. Paulasto 2006: 82) focus fronting.

The tendencies of the frequencies of focus fronting in the RWC and the LI could be confirmed when comparing these frequencies to occurrences of clefts and pseudo-clefts in the two corpora. In the Radio Wales Corpus, the standard constructions appeared 5.84

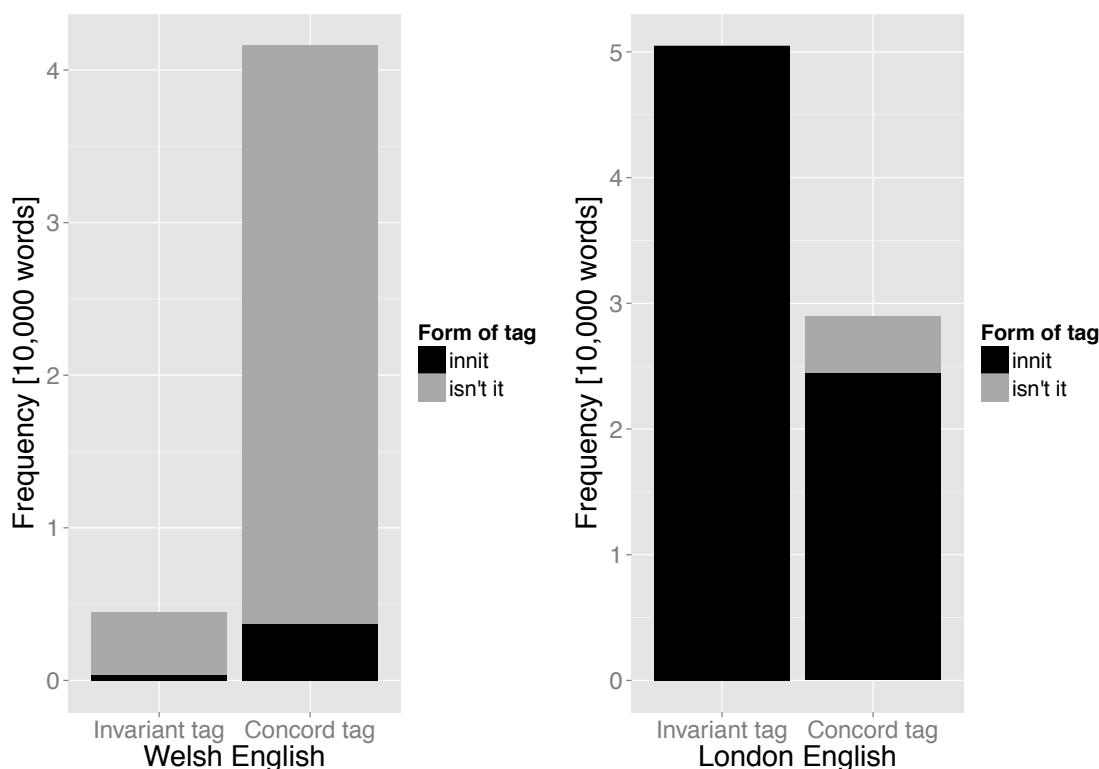


Figure 14: Invariant tag questions and concord tag questions (subject and verb of the tag are consistent with subject and verb of the main clause) involving either *isn't it* or *innit* in the Radio Wales Corpus and the Linguistic Innovators corpus

times per 10,000 words, focus fronting being somewhat less frequent with 4.58 times per 10,000 words. In the LI, the frequency per 10,000 words of clefts and pseudo-clefts was 6.15, which makes the frequency of focus fronting (0.95) appear even smaller. In relation to these standard constructions, focus fronting hence was a lot more frequent in the Welsh corpus than in the London data.

As for the non-standard habitual progressive, the frequency values showed a reversed tendency to that of focus fronting. Non-standard habitual progressives were more frequent in the LI (4.6 occurrences per 10,000 words) than in the RWC (2.64 occurrences per 10,000 words), possibly a reason for the feature's low salience in Welsh English. It seems crucial, though, to view those frequencies against the backdrop of other habitual markers, since habituality may be evoked more by some interview styles than by others. For example, one part of the RWC consists of oral history interviews, where

the subjects reminisce about past experiences (cf. Roller 2015). Many interviewees elaborate, for instance, on routines and habits during their childhood – stories which probably promote the use of habitual markers. To get an idea of past habituality in the corpora, I analysed occurrences of *used to*, *would* and the simple past indicating habitual actions. As for present habituality, the frequencies of respective simple present constructions were determined (see Chapter 3).

In the London corpus, the standard habitual markers were used 44.85 times per 10,000 words, while the non-standard habitual progressive occurred 4.6 times per 10,000 words. Counting together the frequencies of these standard and non-standard features, the latter make up 9.3 per cent of the whole. In the Welsh corpus, 33.44 occurrences per 10,000 words of the standard habitual markers oppose 2.64 occurrences per 10,000 words of the non-standard ones. Counting together the standard and non-standard habituals' frequencies, the latter amount to 7.32 per cent of the whole. That means, non-standard habitual progressives are (slightly) more frequent in relation to respective standard constructions in the LI than in the RWC, which confirms the tendencies observed previously. What is more, the higher frequency of non-standard habituals in the LI was both found when looking at past tense constructions in isolation and when only taking present tense features into consideration.

Apart from habitual markers, non-standard habitual progressives might be perceived as more or less frequent depending on the whole amount of *-ing* forms heard. The following forms were, thus, counted in the LI and the RWC:

- *-ing* participles in verb phrases, occurring, for example, in the progressive aspect succeeding BE: e.g. *I'm cooking now* (cf. Quirk et al. 1985: 96–97)
- Participial adjectives: e.g. *Her laughing husband* (cf. Quirk et al. 1985: 413)
- Gerunds, i.e. *-ing* forms occurring in “nominal *-ing* participle clauses” (Quirk et al. 1985: 1063): e.g. *He enjoys playing squash* (cf. Quirk et al. 1985: 1063)

Of all these *-ing* forms in the Linguistic Innovators corpus, non-standard habitual progressives make up 2.68 per cent, while they amount to 1.81 per cent in the Radio Wales Corpus. In this analysis, then, non-standard habituals are (slightly) more frequent in the London data than in the Welsh material.

The inverted word order in indirect questions – a feature found to be not particularly salient in the context of Welsh English – appeared in the RWC and the LI with almost the same normalised frequencies, namely with 0.41 and 0.4 occurrences per 10,000 words. To relate these findings to the overall pervasiveness of indirect questions in the corpora, instances of standard indirect questions with the verb following the subject were counted (as in *He asked me if I liked it*; see Chapter 3). It was then determined how frequent the non-standard indirect questions were in relation to these standard constructions. In the RWC, non-standard indirect questions make up 5.01 per cent of all indirect questions, and thus are a little bit more frequent than the non-standard forms in the London corpus (4.6 per cent). This is in line with the tendencies provided by the frequency values for the inverted word order in indirect questions in the two corpora.

The last feature that was tested against the backdrop of a functionally equivalent Standard English construction was the zero past tense of regular verbs.⁷ The zero past tense, also not very salient, occurred with higher normalised frequencies in the London corpus than in the Welsh data. To find out if these tendencies can stand up to a comparison with the standard language, standard past tense forms of regular verbs (ending in *-ed*) were also counted in the two corpora. In the RWC, such *-ed* forms appeared 76.85 times per 10,000 words, compared to 0.15 times for the zero past tense forms. Interestingly, the standard past tense forms were visibly more frequent in the LI with 120.6 uses per 10,000 words. However, adding up the frequencies of standard and non-standard past tense forms returns a still slightly larger proportion of the zero past in the LI (0.86 per cent) than in the RWC (0.19 per cent).

In sum, although the Radio Wales Corpus and the Linguistic Innovators corpus surely have their differences from one another, the analyses in this section offer validations for the frequency values of the non-standard features explored in the two corpora. The fact that the studies confirmed the frequency tendencies for *all* grammatical features under investigation suggests that the RWC and the LI are not too different in nature and, thus, that it is appropriate to use them for comparative linguistic analyses.

⁷ *that*-clauses replacing infinitival subclauses were not explored here, since they could neither be found in the RWC nor in the LI, and comparing the frequencies of standard features with zero would not offer any conclusive results.

6.2.4 Frequency in the media

Contrasting non-standard with standard constructions appears to be a helpful procedure to generalise tendencies of frequency values beyond individual corpora and their specific interviews, interviewers and interviewees. But to what extent does the speech of such interviewees, namely (ideally) the everyday speech of “ordinary” Welshmen and -women, influence perceptions of Welsh English by Londoners at all? In the salience questionnaire, a range of subjects from London indicated that they did not have any Welsh friends and that they had never been to Wales. It is thus plausible that what they know about Welsh English is rather (or at least also) connected to listening to Welsh people and/or performed Welshness in the media (cf. Williams, Garrett and Coupland 1999: 357). (Staged) Welsh English in the media is interesting for my study inasmuch as it may differ from ordinary everyday speech. For example, some Welsh TV presenters or comedians might stress on their Welsh identity through stylisation, “playfully and creatively select[ing] from a pre-existing repertoire of culturally significant Welsh dialect forms of English” (Coupland 2001: 347). This could result in overusing some stereotypical linguistic features. Other presenters, by contrast, might try not to sound too Welsh, which could find expression in avoiding particular (salient) Welsh English forms. Stylisation could be especially interesting regarding the invariant tag question *isn’t it*, which occurred with rather low token frequencies in the Radio Wales Corpus. Even though the frequency differences in comparison with London English offer an explanation for the feature’s high salience in Welsh English, it may also be that this invariant tag question is used frequently, and in a stereotypical way, in the media, which may affect the Londoner’s perceptions. Generally, it seems that “[t]aking media language and discourse into consideration is necessary because today’s mediascapes feature a vast range of sociolinguistic styles and representations of speakerhood for audiences to respond to” (Androutsopoulos 2014: 243).

Welsh English in the media was approached by exploring video material from (1) a range of TV shows hosted by Rhod Gilbert, Rob Brydon and Alex Jones – the three Welsh comedians/presenters named most frequently by the subjects from London in the salience questionnaire – and (2) seasons one and two of the series *Gavin & Stacey*, the TV programme from Wales the London subjects were most familiar with. As explained in detail in Chapter 5, 20 hours of video clips were analysed, counting the frequencies of the six non-standard grammatical features discussed above. The follow-

ing results, displayed in Table 6, were obtained from the analysis. Note that only the features' raw numbers of occurrence but no normalised frequencies are provided. The videos were not transcribed, thus the exact number of words of the “media corpus” is unclear. This should be unproblematic, though, as the table merely aims to show whether frequency distributions *between* the features differ in the TV clips and the RWC.

	TV clips Frequency absolute	RWC Frequency absolute
Focus fronting	22	123
Non-standard habitual progressive	3	71
Invariant tag question <i>isn't it/innit</i>	3	12
Inverted word order in indirect questions	0	11
Zero past tense of regular verbs	1	4
<i>that</i> -clause instead of infinitival subclause	0	0

Table 6: Absolute token frequencies of Welsh English features in the media (TV clips) and the Radio Wales Corpus. The two most salient features, focus fronting and invariant *isn't it*, are highlighted in bold print

Unsurprisingly, the two datasets differ from each other in some respects regarding the features' frequency distributions. In the TV clips, there are no instances of the inverted word order in indirect questions, and the habitual progressive is as frequent as the invariant tag question *isn't it/innit*, while the former is more frequent than the latter in the RWC. But there are also some general tendencies that apply to both datasets. Focus fronting is clearly more frequent than the rest of the features in the RWC and in the media data. Taking a closer look at the media data, this tendency emerges in different parts of the corpus: Both in (1) the clips with Welsh presenters/comedians and in (2) *Gavin & Stacey*, focus fronting is the most frequent feature. *that*-clauses replacing infinitival subclauses could neither be identified in the RWC nor in the TV clips. The invariant tag question *isn't it/innit* is considerably less frequent than focus fronting. As in the RWC, the invariant tag does not possess a higher frequency value than the less salient constructions, such as the habitual progressive, in the media data. Thus, these frequency counts do not offer any direct quantitative evidence for the invariant *isn't it* being a stereotype.

Taking a look at individual occurrences of the invariant tag does not provide any qualitative evidence of stereotypical or stylised uses of the feature, either (cf. Coupland 2001; Labov 1972a). The occurrences in the presenters' clips come from Alex Jones, and there is no hint that they were produced in an attempt to portray a Welsh person. "[...] we were saying earlier isn't it" seems to be a spontaneous side comment by Alex to her co-host Matt. "Well, we better had Jeff isn't it" turns up during a conversation with musician Jeff Lynne about him performing again on stage in the future, and about Alex Jones better checking if he is still able to do it. In *Gavin & Stacey* (seasons one and two), only one instance of the feature was found. At Gavin and Stacey's wedding (season 1, episode 6), Stacey's Welsh friend Louise uses it when asked where her boyfriend is. She explains that she did not let him come to the reception since he refused to go to church service and made fun of her because of it. Louise concludes with: "That was out of order though, innit?" Louise is a minor character almost never appearing in the series, thus it does not become clear whether she is meant to represent a stereotypically Welsh person. Looking at it the other way around, the main characters expressing Welshness quite strongly, such as Stacey's uncle Bryn or her friend Nessa, do not make use of the invariant tags in any episode analysed, while employing focus fronting quite regularly. Consequently, these findings cannot explain the high inter-lectal salience of the invariant *isn't it/innit* on the basis of stereotypical language use in the media.

In sum, the media analysis supports general tendencies observed in the frequencies in the Radio Wales Corpus. Focus fronting is considerably more frequent than the other features under investigation, including the invariant tag question *isn't it/innit*. The fact that there are no larger discrepancies regarding the presence of features in the media language analysed and in the RWC further safeguards my approach of predicting salience by means of corpus frequencies. Even though not all London participants in the present study may have been familiar with everyday Welsh English grammar, the representations of Welsh English they encounter on TV, at least in the shows analysed, seem to not differ significantly from ordinary everyday speech.

6.2.5 Variation of frequency in Welsh English

Before moving on to other potential explanatory factors of salience, it should be stressed that the frequencies of Welsh English features discussed in this chapter are generalisa-

tions across different speakers. They represent probabilities in the use of grammatical structures by an *average user* of Welsh English, thus necessarily missing interindividual differences. This section takes a closer look at internal variation in Welsh English, as represented by the Radio Wales Corpus, by analysing the factors age, region in Wales and sex. The findings are not directly aimed at explaining salience; the aim is rather to (further) reduce simplification by making the distribution of grammatical features more plastic. Still, interindividual differences in the *use* of the features might help to explain interindividual differences in the *perception* of the features, as discussed in Section 6.5.

Due to the limited number of hits in the RWC (four of the six features occurring no more often than a dozen times in the whole corpus), frequency distributions of individual features will not be dealt with here. Variation is approached on a more general level, by comparing the overall frequencies of *all* Welsh English features investigated in different speaker groups. For example, for male speakers, the absolute frequency of all non-standard features (i.e. focus fronting, non-standard habitual progressives, invariant *isn't it/innit*, non-standard indirect questions and zero past⁸) in the RWC amounts to 99. Normalising this value by 10,000 words, taking into account the number of all words by male speakers in the corpus (107,759), we arrive at a relative frequency of 9.19.

In order to test for age-related variation, the speakers in the RWC were divided into three groups: people aged up to 30 years of age, people between 31 and 60 and people aged 61 or older. For each group, it was determined how frequently the five non-standard features were used. As shown in Table 7, there is an apparent-time decrease in the occurrence of the features from the older to the younger generations. The usage rate of the 61 plus group amounts to 11.27 per 10,000 words, the rate of the youngest cohort not even to half that value.

A log-likelihood test reveals that the features' normalised frequency in the oldest age group is significantly higher than the values in the other two groups ($p < 0.001$). By contrast to that, the difference in relative frequencies between the two younger co-

⁸ *that* replacing infinitival *to* was not analysed, since the feature could not be found at all in the Radio Wales Corpus.

Age group	Frequency of features per 10,000 words	Frequency of features absolute
0–30	4.58	19
31–60	6.57	81
61+	11.27	117

Table 7: Token frequencies summed up for all Welsh English features in different age groups in the Radio Wales Corpus

horts is not statistically significant. As far as the range of features is concerned, all five constructions occur in the 61 plus and the 31 to 60 group. In the youngest cohort, only focus fronting, the invariant *isn't it/innit* and non-standard habitual progressives could be found, but no instances of the inverted word order in indirect questions and the zero past tense of regular verbs. In the RWC, the youngest speakers of Welsh English thus show a more restricted use of non-standard features (at least of those investigated here) than older generations, with regard to both token and type frequency (i.e. the range of features). This apparent-time decrease in frequencies of non-standard features could be an indicator of dialect levelling in Wales. In line with that, Paulasto (2006: 214–215) argues that “a significant amount of quantitative levelling has taken place in WE dialects over the past fifty years”. Specifically, she records declines in usage rates of focus fronting and non-standard uses of the progressive form. A question arising from these tendencies is whether decreases in usage frequencies for younger speakers affect their perceptions of features. Do young speakers still readily associate grammatical structures with Wales if they are not commonly used by themselves and their peers? Section 6.5 will address age-related differences in perceptions of Welsh English in detail.

As well as age-related variation, Chapter 3 suggests that there is regional variation in Wales, with English dialects in the North West and South West of the country exhibiting the clearest influences from the Welsh language, while Eastern Welsh English dialects are also affected by linguistic influences from bordering English English areas (cf. e.g. Awbery 1997; Collins and Mees 1999; Garrett, Coupland and Williams 1999; Penhallurick 2008b; Thomas 1997). To explore regional variation, the interviews in the Radio Wales Corpus were assigned to the Welsh dialect areas discussed in Garrett, Coupland and Williams (1999: 325). These are: South East (urban), South East (valleys), South West, Mid Wales, North West and North East. For each area, the frequencies of

the five grammatical features under discussion were determined.

Counting together the occurrences of all features, most non-standard constructions could be found in the data from the South West (11.81 per 10,000 words) and the North West (10.65 per 10,000 words), illustrated in Figure 15. The subcorpus from the North East exhibits the lowest rate of features (4.91 per 10,000 words). These findings are in line with the literature stating that Welsh-language influences on English – potentially at work regarding focus fronting, the invariant *isn't it*, the non-standard habitual progressive and the non-standard indirect questions – are most pronounced in the western parts of the country (cf. e.g. Coupland and Thomas 1990). This is due to the fact that the highest percentages of native Welsh speakers are to be found in these western parts of Wales.

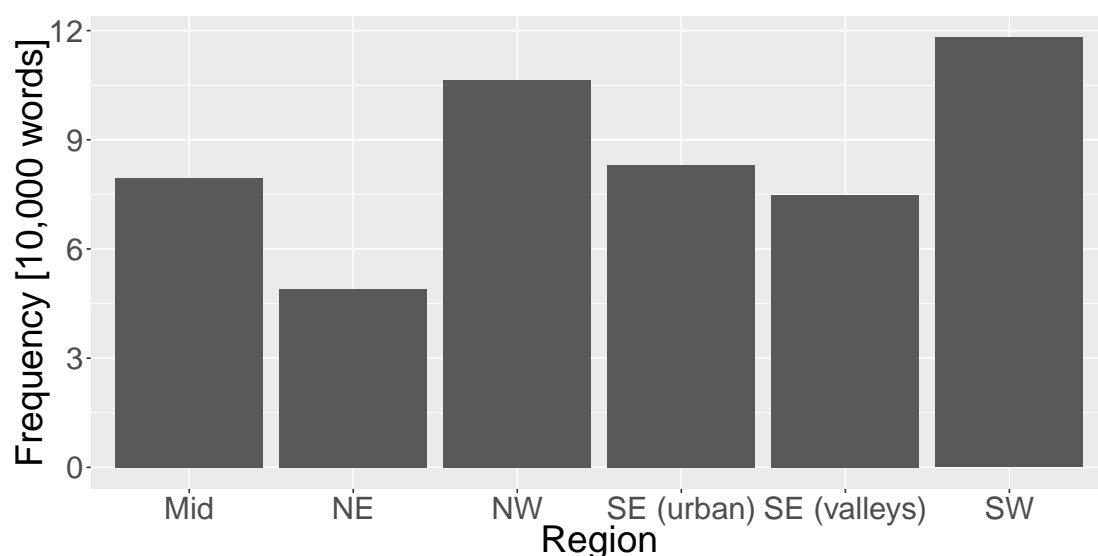


Figure 15: Normalised frequencies of Welsh English grammatical features in different Welsh regions in the Radio Wales Corpus

Despite these frequency differences, the features are generally widely spread across Wales, according to the RWC. Each of them was found in at least one northern and one southern area. Focus frontings and non-standard habitual progressives – the two most frequent features in the corpus – were detected in all RWC regions. Invariant tag questions *isn't it/innit* and the inverted word order in indirect questions turned up in five of the six dialect areas. The former was absent in the South West, the latter in the North East. The zero past tense form of regular verbs could only be attested for

two regions: the South West and the North West. With regard to the dialect regions, this can be summed up as follows: All of the five features occurred in the North West, four out of five in the South West, South East (urban), South East (valleys) and in Mid Wales, and three out of five in the North East. In sum, the features seem to be not limited to specific areas in Wales, potentially increasing their chances of being perceived as “Welsh” by people from all over the country. Still, these perceptions may be most pronounced in the western parts of the country due to the comparatively high token frequencies of the non-standard features there. The relation between listeners’ regional affiliation within Wales and their perception of Welsh English grammar will be scrutinised in Section 6.5.

Finally, variation in Welsh English was explored in relation to the variable sex. As discussed in Section 5.2.1, a range of studies found that women use (or at least claim to use) more standard language than men (cf. Edwards 2009; Garrett 2010). To get an idea of whether such sex-related differences in usage frequencies are visible in the Radio Wales Corpus, the speakers were divided into men and women. For each of the two groups, the frequencies of the five non-standard features were determined, listed in Table 8.

Sex	Frequency of features per 10,000 words	Frequency of features absolute
Female	8.06	130
Male	9.19	99

Table 8: Token frequencies summed up for all Welsh English features for female and male speakers in the Radio Wales Corpus

As can be seen in the table, there are no larger differences in usage frequencies of male and female speakers. Normalised by 10,000 words, the frequencies in the former group amount to 9.19, in the latter group to 8.06. Male speakers in the RWC use the non-standard features slightly more frequently, but the difference to the female group is not statistically significant. Note that the absolute frequencies are higher in the female cohort as the subcorpus is larger. All of the five features were found in both groups. Based on these rather marginal differences, it can be hypothesised that male and female speakers from Wales do not differ from each other significantly in their

regional indexation of the non-standard features under investigation. The assumption will be tested in Section 6.5, exploring potential effects of sex on saliency perceptions.

6.2.6 Interim summary: Frequency

The studies in this section suggest that saliency in Welsh English grammar, to a certain extent, is frequency-based. Insiders' saliency ratings were found to correlate better with token frequencies in Welsh English than ratings of outsiders. This indicates that the quantity of experience with a regiolect affects the accuracy of regiolect perception. Moreover, the subject-sensitive perspective found features with higher token frequencies to be generally more salient to insiders and outsiders. While (comparatively) high token frequency in the target variety may accompany some salient features, as focus fronting, it is no categorical requirement for regional indexation in morphosyntax, though. The invariant tag question *isn't it* appears to be a salient grammatical feature in Welsh English despite its rather low token frequency in the RWC. However, frequency differences between Welsh English and London English could offer an explanation for the feature's conspicuousness: it is more frequent in Welsh English than in London English – in the realisation /'ɪzənt ɪt/, with the Londoners preferring *innit* – /'ɪnɪt/. Moreover, a positive correlation between interlectal saliency and frequency differences in the RWC and the LI was found, which was distinctly stronger than the correlation with token frequencies in Welsh English only. It seems that for people from outside a particular dialect area (here: Londoners), frequencies in the target dialect alone (here: Welsh English) are not the most important experiential factor; frequency differences between one's native and the target dialect have a larger effect on how the target dialect is perceived.

To rule out the possibility that these findings are tied to particularities of the corpora used, several analyses were carried out. The tendencies of the frequencies of the non-standard features were found to be confirmed when relating them to the frequencies of functionally equivalent Standard English constructions in the RWC and the LI. An investigation into Welsh English grammar as used in TV shows, moreover, found that the frequencies of the non-standard features in the RWC, representing “ordinary” everyday Welsh English, and the frequencies in the TV clips analysed show the same tendencies. Thus, for Londoners' perceptions of Welsh English it may not make a significant difference whether those perceptions are based on everyday Welsh English or

Welsh English as heard in the media (at least in the form of the shows explored here). Finally, the analyses with the RWC point to age-related, sex-related and regional grammatical variation in Welsh English. Several features occurred in all age groups, among female and male speakers and in different regions of Wales, but the findings also underline how usage frequencies vary between interviewees, and suggest that usage-based linguistic knowledge is unique to every single speaker.

6.3 Salience and pervasiveness in the British Isles

Frequencies in Welsh English and frequency differences in Welsh English versus London English seem to contribute to how Welsh English features are perceived by insiders and outsiders. But up to this point, frequency can only explain why some features are seen as more characteristic of Welsh English than of English in London. It remains unclear why features such as the invariant tag question *isn't it* were associated with Wales much more frequently in the salience questionnaire than with, say, Ireland or Scotland. It thus seems that in studying the regional indexation of variables in the British Isles, i.e. the question why features are considered as typical of a specific area, one needs to take a look at a feature's pervasiveness across the *whole of the British Isles*.

I hypothesise that a feature's pervasiveness in the British Isles influences perceptions of both outsiders and insiders. As discussed in Chapter 2, if a non-standard feature is considered as typical of Welsh English by a Londoner, this may not only be due to a higher usage frequency of the feature in Wales than in London. Additionally, the feature may generally be more pervasive in Wales than in other regions of the British Isles, leading the Londoner to assign the feature to Wales instead of, for example, Ireland or a part of England. As for insiders, i.e. people from Wales, a high frequency of a feature in Welsh English may be coupled with the item being less frequent in other (e.g. bordering) dialects, making Welsh people realise even more that the feature is distinctive for their own area. In sum, I assume that inter- and intralectally salient features are generally more pervasive in Wales than in the rest of the British Isles.

Based on the previous sections, the probably most straightforward way of approaching the pervasiveness of non-standard features across the British Isles would be via corpus analyses. Frequencies of the features could be determined in corpora of (varieties of)

English in Ireland, Scotland and England and contrasted with the frequencies obtained for Welsh English. Since this was beyond the scope of the present study, though, the Electronic World Atlas of Varieties of English (eWAVE) was consulted instead (cf. Kortmann and Lunkenheimer 2013). It needs to be stressed that eWAVE does *not* provide frequency counts. As stated in Section 5.4.2, the atlas categorises occurrences of non-standard grammatical features in different regional varieties of English on the basis of indices of pervasiveness. These indices are displayed in Table 9. To give an example of a feature, the non-standard habitual progressive has a “B” rating in Welsh English, that means it is considered to be “neither pervasive nor extremely rare” (Kortmann and Lunkenheimer 2012: 5) there.

eWAVE rating	Indicates
A	Feature is pervasive or obligatory
B	Feature is neither pervasive nor extremely rare
C	Feature exists, but is extremely rare
D	Attested absence
X	Not applicable
?	No information available

Table 9: Indices of pervasiveness in eWAVE. The definitions are taken from Kortmann and Lunkenheimer (2012: 5)

These indices of pervasiveness – based on judgements by (typically native-speaker) linguistic experts – can only be “an abstraction of and a rough approximation to linguistic and social reality” (Kortmann and Lunkenheimer 2012: 6; see also the discussion in Section 5.4.2). One of the reasons is that there is “some variation [...] in how features were interpreted by contributors, [and] how A, B and C ratings were assigned to features not categorically present or absent in a variety” (Kortmann and Lunkenheimer 2012: 6). Also, the eWAVE ratings do not account for individual properties of features in specific varieties, such as a unique intonation or non-standard pragmatics, which may increase the conspicuousness of a construction in a particular variety. (However, more information about variety-specific properties of features can be found in the accompanying articles describing the varieties in the atlas in more detail.) This means that the findings presented in this section are not directly comparable to the frequency-based analyses in Section 6.2. Nevertheless, by providing a general overview of the pervasiveness of features across the British Isles, the eWAVE data help to view the results for

Welsh English and London English in the context of the bigger picture.

The analyses carried out on the basis of eWAVE included the following working steps:

- (1) Compiling an overview of the six non-standard features under investigation and their eWAVE ratings of pervasiveness in Wales as well as in the rest of the British Isles.
- (2) Exploring potential correlations between a feature's salience and its overall pervasiveness, i.e. the ratio between the pervasiveness in Wales and the pervasiveness in the rest of the British Isles.
- (3) Analysing the role of the number of regions where a feature occurs – to what extent are highly salient features restricted to small-scale geographical areas?

Regarding (1), the eWAVE ratings of pervasiveness of the six non-standard features in Wales and in other British Isles regions are displayed in Table 10. Note that as for “Other possibilities for fronting than StE”, the aspect of “focus” is not mentioned in the feature name. However, the eWAVE examples provided for Welsh English clearly fall in the category of focus fronting, for example ‘*worn out*’ *I got* (Penhallurick 2012: 67). Similarly, the example of the feature provided for Irish English in eWAVE is an instance of focus fronting: *A story now he told me* (Filppula 2012: 46). “Invariant non-concord tags” in eWAVE do not only cover the invariant *isn't it* (plus variations such as *innit* or *ain't it*) but also *eh*. However, *eh* in tag questions is not common in Welsh English (cf. e.g. Parry 1999; Penhallurick 2008a). According to Penhallurick (2012: 62), who compiled the eWAVE ratings for Wales, the rating of this feature in Welsh English refers to “generalized *isn't it/innit* as a confirmatory interrogative tag”. By contrast to that, the eWAVE ratings of two other British Isles regions, namely Scotland (cf. Miller 2008; Smith 2012) and the Channel Islands (cf. Rosen 2012, 2014), rather refer to *e(h)* than to *isn't it/innit*.⁹ These two regions were thus excluded from the analyses, since they might have distorted the results.

⁹Miller (2008) discusses a range of examples of *e* in negative and positive declarative clauses, such as *...we know him quite well by now e?* Rosen (2012: 102), who provided the eWAVE ratings for the Channel Islands, states that “even though the invariant non-concord tag *eh* [...] occurs in all spoken English varieties, its frequency of use, pragmatic functions and distribution across all age groups of the Channel Islands speech communities make it quite a distinct ChIsE feature”.

Feature name in eWAVE	Feature number	Pervasiveness in Wales	Pervasiveness in other Br. Isles areas
Invariant non-concord tags	F165	A	B: Southeast of England B: Southwest of England C: North of England
Other possibilities for fronting than StE	F224	A	B: Ireland C: Southwest of England
Wider range of uses of progressive be + V-ing than in StE: extension to habitual contexts	F89	B	A: Isle of Man B: Ireland C: Orkney and Shetland
Inverted word order in indirect questions	F227	B	A: Ireland B: North of England B: Southwest of England C: Isle of Man C: Orkney and Shetland
Substitution of that-clause for infinitival subclause	F207	B	/ (No other areas)
Zero past tense forms of regular verbs	F132	C	/ (No other areas)

Table 10: The non-standard features under investigation in the present study as described in eWAVE. The data are taken from Kortmann and Lunkenheimer (2013). The two most salient features, non-standard fronting and invariant tag questions, are highlighted in bold print

From a first qualitative glance at the data, two aspects become noticeable that – in this combination – only apply to the salient invariant tag questions and focus fronting: (a) the features have an A rating in Wales and (b) the ratings for the other regions are lower than those for Wales. Both aspects do not apply to the (less salient) non-standard habitual progressive and the inverted word order in indirect questions. The features only have B ratings in Wales and occur with A and B ratings in other areas. The *that* construction and the zero past tense of regular verbs do not occur in other British Isles areas in eWAVE, hence they are more pervasive in Wales than outside. But they only have a B and a C rating in Wales. In sum, the salient dialectal features in this dataset are quite pervasive in the target area (here: Wales) and are more pervasive in the target area than in other (close-by) dialect regions.

To explore the relation between a feature's pervasiveness in the target area and outside more closely, the eWAVE indices of pervasiveness D–A were converted into the

numbers 0–4. In more concrete terms, D received the value 0, C was converted into 1, B into 2 and A into 3. That means, the non-standard habitual progressive, for example, received the value 2 for Welsh English, since it has a B rating there. In a next step, the features' mean pervasiveness outside Wales (but in the British Isles) was determined. This was done by counting together the pervasiveness values for all regions and dividing the result through the number of regions. For example, non-standard habitual progressives in eWAVE received an A rating in Manx English, a B in Irish English and a C in Orkney and Shetland. In numbers, this would be $3 + 2 + 1 = 6$. Dividing this result through the number of regions, namely 3, yields the mean pervasiveness of 2 per region.

Using these pervasiveness ratings from Wales and from the rest of the British Isles, I then calculated the features' overall pervasiveness value. *Overall pervasiveness* is defined in the context of the present study as the difference of a feature's pervasiveness in Wales and the feature's mean pervasiveness in the rest of the British Isles. In what follows, overall pervasiveness is abbreviated as OPV. The calculation can be illustrated as follows:

Pervasiveness in Wales	—	Mean pervasiveness in rest of British Isles	=	Overall perva- siveness (OPV)
E.g. 2	-	1.8	=	0.2

A positive OPV value indicates that the feature is more pervasive in Wales than on average in the rest of the British Isles. For example, in Wales, the inverted word order in indirect questions is rated as a B feature, thus it receives the value 2. The feature's mean pervasiveness outside Wales is 1.8. Subtracting 1.8 from 2 yields an OPV value of 0.2. Thus, in eWAVE, this feature is slightly more pervasive in Wales than on average in the rest of the British Isles. Table 11 summarises the information about the pervasiveness of the non-standard features (a) in Wales and (b) in the rest of the British Isles, as well as (c) the features' OPV.

As stated above, a central hypothesis in this section is that (inter- and intralectally) salient features are more pervasive in Wales than (on average) in the rest of the British

Feature	Pervasiveness Wales	Pervasiveness rest of Br. Isles	OPV
Substitution of that-clause for infinitival subclause	2	0	2
Other possibilities for fronting than StE	3	1.5	1.5
Invariant non-concord tags	3	1.67	1.33
Zero past tense forms of regular verbs	1	0	1
Inverted word order in indirect questions	2	1.8	0.2
Wider range of uses of progressive be + V-ing than in StE: extension to habitual contexts	2	2	0

Table 11: eWAVE-based pervasiveness ratings of non-standard features in Wales, in the rest of the British Isles and in Wales vs. the rest of the British Isles (OPV). The features are ordered by their OPV values. The two most salient features, non-standard fronting and invariant tag questions, are highlighted in bold print

Isles. In other words, features with a higher OPV are more salient. As illustrated in Table 11, this does only partially apply to the data. Non-standard frontings and invariant tags are positioned in the upper half of the table, with higher OPVs than three less salient features. The *that* clause, however, exhibits the highest OPV and the lowest salience rating. Comparing insiders and outsiders, the OPV values for zero past tense forms and habitual progressives are more consistent with the salience ratings of the outsiders. The observation is confirmed by correlating OPV and salience, yielding a lower value for the Welsh subjects ($r = 0.16$) than for the Londoners ($r = 0.42$).

If a feature's OPV plays a larger role for the judgements of outsiders than of insiders, how can this be accounted for? It might be that for intralectal judgements, a feature's pervasiveness inside versus outside the own area is just not as important (even though it might play a role to some extent) as the frequencies in one's own dialect, particularly for people who have not yet been in contact with a lot of other regional dialects. By contrast to that, Section 6.2 showed that interlectal salience is more strongly related to frequency differences between Welsh English and London English than frequencies in Welsh English only. Generally, the social indexation of linguistic variables through outsiders seems to rely a lot on comparative probabilities (cf. Rácz 2013). If someone

assigns a feature to Wales, this might be because he or she is more familiar with the feature from Welsh English than from other varieties of English. Thus, it makes sense that the pervasiveness of features in the British Isles has a somewhat higher predictive power regarding inter- than intralectal salience.

These tendencies could be confirmed when relating OPV to the salience ratings of the individual Welsh and London informants (in same way as was done with frequency in Section 6.2). If an informant had assigned all of the six Welsh English features to Wales, the mean OPV of these features would amount to $(1.33 + 1.5 + 0 + 0.2 + 2 + 1) / 6 = 1.01$. Of the 130 Welsh informants who had at least identified one of the Welsh English features, 55.38 per cent selected Welsh English features with (mean) OPV values higher than 1.01. That means, a slight majority of subjects chose the features more pervasive overall, but the percentages point to a great deal of interindividual variation in the data. Of the 111 London subjects who had assigned at least one of the features to Wales, 63.06 per cent chose constructions with above-average OPV values. That means the choices of the informants from London were more in line with OPV than those of the Welsh informants. Still, the data also display interindividual variation in the London group.

To substantiate these findings for Welsh English, correlation calculations were carried out with the data from all of the 24 non-standard features in the salience questionnaire. As described in Chapter 5, the questionnaire also included six test sentences each with non-standard features from varieties of English in Ireland, Scotland and England. Apart from serving as distractors disguising that the survey actually focused on Welsh English, these constructions also offered the opportunity to be analysed in the light of pervasiveness in the British Isles, since all of them are covered by eWAVE. Table 12 provides an overview of the features of Irish English, Scottish English and English varieties of English and their (interlectal) salience values in the questionnaire.¹⁰

¹⁰Note that “Target area” in Table 12 refers to the dialect areas provided as options the salience questionnaire, which were somewhat broader than those provided by eWAVE. While features from the South of England could simply be assigned to “England (South)” in the questionnaire, eWAVE further distinguishes between “English dialects in the Southeast of England”, “English dialects in the Southwest of England” and “East Anglian English” (Kortmann and Lunkenheimer 2013). The subjects in the questionnaire, however, also had the opportunity to tick “Other” and write down a more specific region in England (or in another British Isles area) where they would locate a particular feature.

Feature name in eWAVE	Feature number	Example	Target area	Saliency (%)
After-perfect	F98	<i>She's after buying a house.</i> 'She has just bought a house.'	Ireland	35.67
Is for am/will with 1st person singular	F118	<i>I's not doing that.</i> 'I am not doing/won't do that.'	England (North)	27.33
Use of definite article where StE favours zero	F64	<i>She was appointed CEO at the Google, UK.</i>	Ireland	25
Was – weren't split	F163	<i>The girls was laughing, but Henry weren't.</i>	England (South)	20.67
Go-based future markers	F114	<i>He gon pick me up.</i>	Scotland	19.67
Epistemic mustn't	F122	<i>You mustn't</i> ['can't'] <i>be right.</i>	England (North)	18.67
Amn't in tag questions	F164	<i>I'm home, amn't I?</i>	Scotland	18
Other non-standard habitual markers: analytic	F93	<i>She do be depressed a lot.</i>	Ireland	17.67
Want/need + past participle	F124	<i>The dog wants petted.</i>	Scotland	16.67
Subject pronoun drop: referential pronouns	F43	A: <i>You got the tickets?</i> – B: <i>Yes, bought already.</i> 'Yes, I've bought them already.'	Scotland	16.33
Clause-final but = 'though'	F211	<i>I'm happy to help you but.</i>	England (North)	16
Too; too much; very much 'very' as qualifier	F222	<i>That's too</i> ['very'] <i>heavy.</i>	Ireland	15.33
Variant forms of dummy subject there in existential clauses	F173	<i>They is a lot of food left.</i>	Ireland	15.33
Plural forms of interrogative pronouns: using additional elements	F39	<i>Who-all was going to come here?</i>	Scotland	15.33
Benefactive "personal dative" construction	F9	<i>They got them some new pairs of jeans.</i>	Ireland	15

Use of indefinite article where StE has definite article	F61	<i>A sun is rising.</i>	England (South)	13.33
Double modals	F121	<i>We might should get a new car.</i>	Scotland	13
Alternative forms/ phrases for dummy it	F4	<i>Thass rainen</i> ‘It is raining’	England (South)	4.67

Table 12: Grammatical features of (varieties of) Irish English, Scottish English and English in England in eWAVE (cf. Kortmann and Lunkenheimer 2013), ordered by their salience values in the questionnaire (perceptions of outsiders)

Note that the following analysis had to concentrate on *interlectal salience*, since I only had data on how outsiders perceived the structures from Ireland, Scotland and England, while lacking informants from inside all of these regions (except those from Southern England). Thus, to get a homogenous “outsider perspective” on all features in the questionnaire, I only used the outsider ratings of the Welsh features (namely those by the Londoners) and of the English features (namely those by the Welsh people) for the following calculations. For the Irish and Scottish features, I worked with the salience ratings of both Welsh and London subjects (since they were both outsiders). For each feature, the OPV value was determined. Then the OPV values of the 24 features (the Welsh ones included) were correlated with the salience values. Interlectal salience and OPV correlated with $r = 0.42$ in this dataset. The value was the same as for the Welsh features only and points to a positive correlation of an intermediate strength. That means, features with higher OPV values, i.e. which were more pervasive in the target area than outside, were generally more salient to outsiders. This relation between OPV and interlectal salience is visualised in Figure 16.

Two things should be noted regarding the findings of the OPV study. First, apart from the overall positive trend, Figure 16 also points to variation in the data. Several features were more and some less salient than predicted by the OPV values. For example, the “is for am/will with 1st person singular” (F118) in the North of England was one of the most (interlectally) salient features in the salience survey and had an OPV value of 1. The “substitution of that-clause for infinitival subclause” (F207) in Welsh English was considerably less salient but had an OPV value of 2. The fact that OPV can predict salience only to a limited extent in this study strengthens the (quite obvious but important)

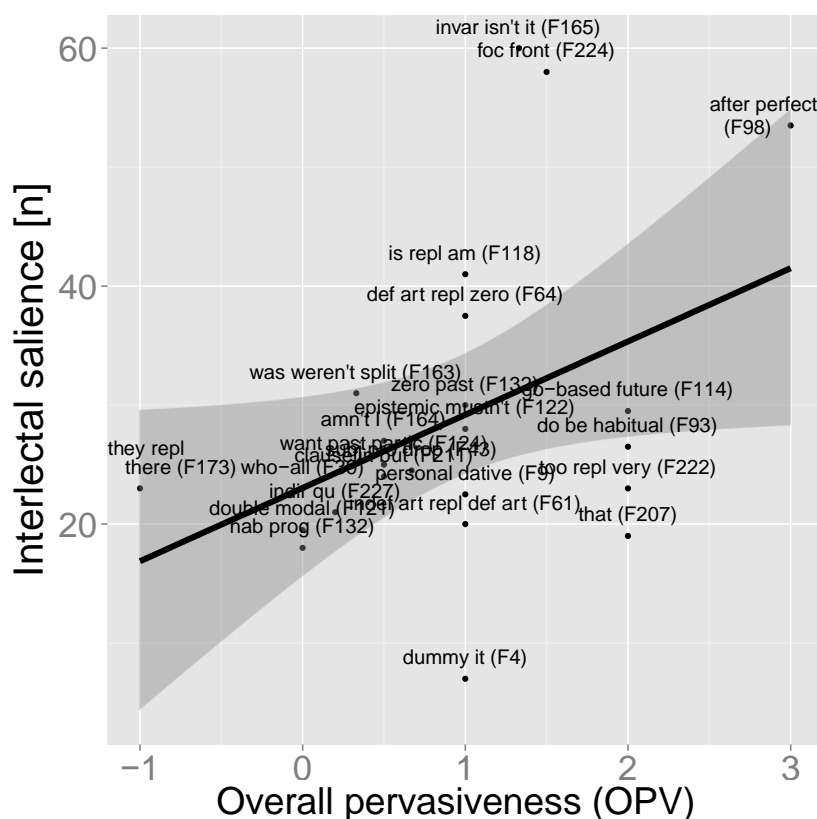


Figure 16: The relation between interlectal salience and overall pervasiveness (OPV) in eWAVE for all 24 non-standard grammatical features explored in the salience survey

assumption that the regional indexation of linguistic variables is determined by an interplay of several factors. Other determinants such as frequency (within a variety as well as ratios between varieties), and, as discussed further down, structural-linguistic and social factors should also be taken into consideration when explaining the socio-linguistic salience of regional linguistic features. Second, the findings regarding OPV may only be restrictedly representative of actual language use. As stated above, the pervasiveness ratings provided by different scholars for different language varieties based on different datasets may be subjective “impressionistic judgments” (Kortmann and Lunkenheimer 2012: 6) to some extent.¹¹ Therefore, it would be very interesting to have future corpus studies determining the features’ frequencies in varieties of English in England, Ireland and Scotland, and contrasting them with the frequencies in Welsh

¹¹Indeed, my frequency counts from Welsh English show that the pervasiveness ratings in eWAVE do not always match the tendencies of frequencies. For example, the *that*-clause replacing an infinitival subclause received a B rating for Welsh English in eWAVE, but could not be found at all in the Radio Wales Corpus.

English.

Apart from a feature's pervasiveness/frequency in other (neighbouring) dialect areas, a construction's sociolinguistic salience may also be influenced by its geographical spread. As stated in Chapter 2, the literature suggests that salient regional linguistic features occur in small-scale geographical areas (cf. Lenz 2003, 2010). An item that is geographically restricted to Wales may carry strong local signs (cf. Kortmann 2012). A hypothesis following from this would be that salient regional features occur only in few or no other close-by dialect areas.

Recalling Table 9, it becomes clear that geographical spread is no distinctive factor for salience in the present study. *that* replacing infinitival *to* clauses and zero past tense forms of regular verbs are *rarissima*, i.e. features restricted to one area, in this case Wales (cf. Kortmann 2012: 695). This is not reflected in high salience values, though. Invariant tags, on the other hand, occur in several British Isles regions. They are no more regionally restricted than non-standard habitual progressives, and especially invariant *innits* are currently spreading across the British Isles (see Section 3.2). Correlating the features' numbers of regions, obtained from eWAVE, with salience supports this finding. Interlectal salience and the number of regions correlate with $r = 0.05$, which points to a very low statistical correlation. Intralectal salience and the number of regions correlate with $r = 0.34$, which even indicates that the more salient features are geographically more spread out. To consolidate the results with additional data, the numbers of regions outside the target area for all 24 features in the questionnaire were correlated with their (interlectal) salience values. The calculation yielded $r = 0.01$, again indicating a very low, almost non-existing statistical correlation. These results thus do not offer any proof for the assumption that a low number of regions where a feature occurs outside the target area correlates with high salience.

A potential explanation for that finding would be that the number of regions alone cannot account for salience; it is important how pervasive a feature is in the target area as well. Features which occur in small-scale geographical areas only and which are rather infrequent *within* these areas might be just too rare in speech to be consciously noted by people and to carry local signals. An example for such a feature would be the zero past tense of regular verbs, a *rarissimum* according to eWAVE. It occurs with

quite low token frequencies in the Radio Wales Corpus (see Section 6.2) and only has a C rating for Welsh English in eWAVE. The feature's low sociolinguistic saliency may be a consequence of its rarity in Wales and outside. On the other hand, a rarissimum which is quite pervasive in the target area may carry strong regional signs. An potential candidate would be the *after*-perfect in Irish English. This rarissimum was one of the most salient features in the present saliency study and has an A rating for Irish English in eWAVE.

The findings in this section can be summarised as follows. (1) Overall pervasiveness plays a larger role for the perceptions of outsiders than of insiders. A reason for that could be that intralectal dialect perceptions are influenced more strongly by token frequencies in one's native dialect, especially for speakers who have not yet been much in contact with speakers of other dialects. (2) According to the subject-sensitive analyses, (especially interlectally) salient features are somewhat more pervasive in the target area than outside. This is also suggested by the correlation calculation based on the 24 features from the saliency questionnaire. The result complements the findings from Section 6.2, adding that the high saliency of linguistic regionalisms to outsiders is not only determined by low probabilities of occurrence in *their own* variety, but also in *the rest of the British Isles*. (3) A feature's geographical spread is not per se a good indicator of regional indexation. Geographically more restricted structures can be, but are not necessarily salient. A decisive factor here could be a feature's pervasiveness/frequency in the target area. If a feature is restricted to a specific area and pervasive/frequent there, this might increase its conspicuousness to outsiders and highlight its regionality. If the feature is rare in the target area, it may not cross the threshold to consciousness and, thus, and not be subjected to processes of regional labelling.

Of course, overall pervasiveness cannot completely account for the variation in saliency ratings. It is hence necessary to analyse further potential determinants of regional indexation. Maybe the saliency of some not very pervasive features is increased due to structural prominence, while other, more frequent features might be less structurally obtrusive. Thus, the next section analyses whether and how the saliency of the Welsh English features under investigation is connected to structural conspicuousness.

6.4 Salience and perceived deviation from the standard language

As discussed in Section 2.3.2, a range of works dealing with the causes of (socio-) linguistic salience mention structural language-internal factors, such as the categorical (versus the gradient) character of a phonological variable or a feature's occurrence in a prosodically prominent position in a sentence (cf. e.g. Auer 2014; Auer, Barden and Grosskopf 1998; Kerswill and Williams 2000; Rácz 2013). Generally, such structural factors seem to be at work on the level of *cognitive salience*. The mere structure of a feature does not necessarily evoke associations with a specific social group or region¹². But structurally conspicuous features may be perceived more consciously than features which are structurally unobtrusive, with only the conspicuous – the cognitively salient – features being able to cross the threshold to sociolinguistic salience.

The following study explores to what extent the Welsh English features under investigation stand out structurally in the sense that they are perceived as deviations from Standard English. The study proceeds in three steps: (1) conducting a questionnaire-based survey in Wales and London, asking people to mark all deviations from Standard English they can find in a set of sentences, (2) relating the indices of non-standardness to salience and (3) in case perceived non-standardness plays a role, trying to find structural generalisations for salient and less salient features.

As for (1), questionnaires were gathered from 25 Welsh people and 25 Londoners (for a more detailed description of the questionnaire, see Section 5.4.3). As described above, the subjects were presented with a range of test sentences and had to mark every deviation from Standard English that they encountered. On the basis of these grammaticality judgements, the *perceived non-standardness* of each feature to Welsh people and Londoners was determined. For example, if a feature was considered to be a deviation from the standard by all subjects from Wales, the perceived non-standardness of this feature would be 100 per cent.¹³ Tables 13 and 14 display the values of perceived

¹²Although it *can* do so, for example, when a construction is similar to another one that is known to be characteristic of a particular speaker group. Then analogy formation may lead to associating the first feature with the same group.

¹³Note that this study did *not* test for degrees of perceived non-standardness in the sense that one feature is seen as simply wrong, while another one may be acceptable in this form in some contexts

non-standardness for the Welsh subjects and the London informants.

	isn't it	zero past	that	indir qu	foc front	hab prog
Non-standard (%)	92	88	76	44	36	28
Salience (%)	60	15	9	22	59	27

Table 13: Percentages of the features' perceived non-standardness and their salience values for the Welsh subjects. The features are ordered by perceived non-standardness. The two most salient features, focus fronting and invariant *isn't it*, are highlighted in bold print

In the group of insiders, there is no apparent systemacity in that the more salient features were generally perceived as larger deviations from Standard English. While the invariant *isn't it* received the highest percentage of perceived non-standardness (92%), the percentage for focus fronting was significantly lower (36%). The majority of Welsh informants did not consider the latter construction to be a deviation from Standard English. By contrast to that, the features with the lowest salience values, the zero past tense of regular verbs and the *that* construction, were identified as deviations from the standard by most of the subjects. This suggests that for insiders, the regional indexation of linguistic constructions is not necessarily related to the perceived non-standardness of the constructions.

However, the situation might look different with regard to the perceptions of outsiders. If linguistic features have low probabilities of occurrence in London speech, they might sound unusual and be probably perceived as deviations from Standard English by a Londoner, while they might sound rather common and natural for a Welsh person who uses the forms *him-* or *herself*. Indeed, the values of perceived non-standardness in the present study conform better with interlectal salience. Correlating salience and perceived non-standardness yields $r = -0.09$ for the insiders and $r = 0.41$ for the out-

or functions. It is plausible that such degrees of non-standardness (which could be tested in future studies) affect perceptions. They can probably explain the high perceived non-standardness of the *that* replaces *to* construction in the present study. This form does not exist in Standard English, and thus may have been perceived by the subjects as simply wrong or ungrammatical. The non-standard habitual progressive, on the other hand, may be positioned in a grammatical grey area in folk-linguistic judgements, since speakers are familiar with various Standard English *-ing* forms, for example, as in the present progressive.

siders.

	zero past	that	isn't it	foc front	hab prog	indir qu
Non-standard (%)	100	100	96	88	64	60
Saliency (%)	20	13	40	39	12	14

Table 14: Percentages of the features' perceived non-standardness and their saliency values for the London subjects. The features are ordered by perceived non-standardness. The two most salient features, focus fronting and invariant *isn't it*, are highlighted in bold print

Table 14 illustrates that deviations from the standard play a larger role in perceptions of Welsh English by outsiders than insiders. In the London group, all of the six features were identified as non-standard forms by the majority of subjects. By contrast to that, only half of the features were considered as deviations from the standard by the majority of the Welsh informants. The vast majority of London subjects marked the invariant *isn't it* and focus fronting as deviations from the standard. The less salient inverted word order in indirect questions and the non-standard habitual progressive received lower non-standardness ratings by the Londoners. However, the remaining two features with rather low saliency ratings, the zero past and the *that* construction, were labelled as non-standard by *all* London subjects. It seems that to the informants from London, and also to those from Wales, these two constructions constituted clear deviations from Standard English, however, the informants had difficulties in assigning them to a particular British Isles region (although eWAVE suggests that they only occur in Wales).

Further proof for that impression is provided when taking a look at the saliency questionnaire again. It was determined how many informants ticked the “don't know” option instead of assigning a feature to (a) region(s). Of the six Welsh English features, the zero past tense and the *that* construction received the highest percentages of “don't know” – in both the London and the Welsh cohort. 42.67 per cent of the Londoners indicated that they did not know where to locate the *that* construction in the British Isles, 28.67 used the “don't know” option for the zero past tense of regular verbs. Of the Welsh informants, as can be seen in Figure 17, 49.33 per cent ticked “don't know” for the *that* construction, and 43.33 per cent did not know where to locate the zero past. By contrast to that, focus fronting and the invariant *isn't it* received the lowest rates

of “don’t knows”. Focus fronting could not be assigned to any region by 11.33 per cent of the Welsh participants and 6.67 per cent of the Londoners. Regarding *isn’t it*, 12.67 per cent of the subjects from Wales and 7.33 of the Londoners chose the “don’t know” option. In sum, it seems that the perceived non-standardness of a feature does not necessarily go hand in hand with regional indexation. Some features may be *cognitively salient* in that they are perceived as deviations from the standard language, but they do not carry any regional signals. A decisive factor here could, again, be frequency, in that low frequencies of occurrence (in a variety or in relation to another) inhibit processes of social/regional labelling.

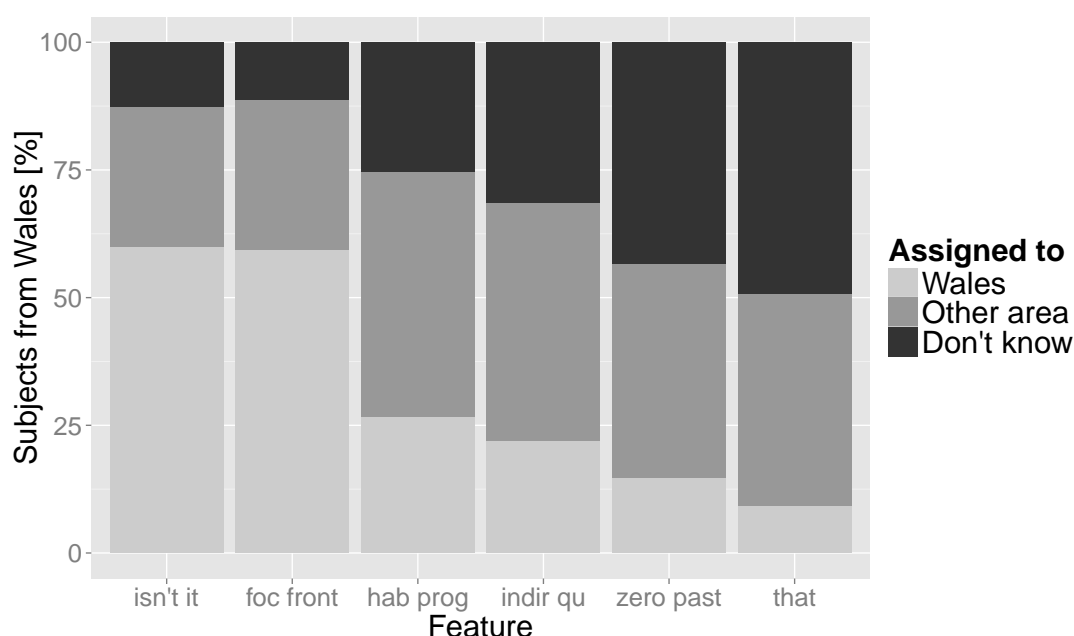


Figure 17: Degrees to which the grammatical features explored in the salience survey were assigned to Wales, to other areas or nowhere (“Don’t know”) by the informants from Wales

Subject-sensitive analyses confirmed that the perceived deviation from the standard language is a more important factor for outsiders’ than insiders’ perceptions. If a Welsh informant had allocated all of the six features to Wales, their mean percentage of perceived non-standardness would be $(92 + 36 + 28 + 44 + 88 + 76) / 6 = 60.67$. If higher perceived non-standardness leads to higher salience, we should expect the majority of informants to assign features to Wales with mean non-standardness values higher than 60.67. However, only 50 per cent of the Welsh participants did so; the other 50

per cent associated features with Welsh English that had perceived non-standardness values of 60.67 or lower. By contrast, two thirds (72.97 per cent) of the London informants recognised Welsh English features with higher perceived non-standardness values.

To substantiate these findings with additional data, I analysed the grammaticality judgments of the other eight non-standard features in the questionnaire which are reportedly used in Ireland, Scotland or varieties of English in England. Again, as in the analyses of overall pervasiveness in the British Isles, the calculations had to concentrate on interlectal salience, since I lacked questionnaires from people from Ireland, Scotland and different parts of England. Thus, the following analysis only includes the outsider ratings of the Welsh features (namely those by Londoners) and of the English English features (namely those by Welshmen and -women). For the Irish and Scottish constructions, the ratings of perceived non-standardness from the Welsh and the London informants were used, as both groups were outsiders. For each feature, the percentage to which it was considered to be a deviation from the standard was determined. The values were then entered into a correlation calculation, involving the variables (interlectal) salience and perceived non-standardness. The correlation calculation yielded $r = 0.24$, indicating a positive correlation of an intermediate strength. This means that also in the combined data with non-standard features from Wales, Ireland, Scotland and England, there is a (slight) general tendency for sociolinguistically salient features to be seen as larger deviations from Standard English than less salient ones, at least by outsiders. Now, why is that?

To find out about potential structural influences on a feature's perceived non-standardness, some mainly qualitative analyses were carried out. In a first step, the features were classified regarding their grammatical domains in eWAVE. The following domains are covered by the atlas: noun phrase, pronouns, tense and aspect, modal verbs, verb morphology, agreement, negation, complementation, relativisation, adverbial subordination, adverbs and prepositions, discourse organisation and word order (cf. Kortmann and Lunkenheimer 2012: 5). Comparing the features and their grammatical domains, it became noticeable to me that all structures that fall in the category of negation received relatively high percentages of perceived non-standardness. The structures are the use of *amn't I* in tag questions, the invariant tag question *isn't it* and

I'm here, amn't I?
The boys was interested but Mary weren't
It is so nice today because a sun is shining
I wanted that I should get leave
 I'm going to the cinema every week **They is something bad**
The cat wants petted **wrong with her**
 Sort of a student he was She got her a new car
I'm wondering is it Frank or not
It was very cold last night so I walk home
very quickly I'm after selling the boat
She lives just across the hallway, isn't it?
It's an exhausting job. I'm happy to do it but

Figure 18: Test sentences in the questionnaire dealing with deviations from Standard English. The degree of a feature's perceived non-standardness is expressed through colouring (the more "non-standard", the darker the colour) and font size (the more "non-standard", the larger the font size)

the *was* – *weren't* split (as in *The boys was interested but Mary weren't*). Interestingly, Cheshire (1996: 5–6) considers negative clauses to be "prominent contexts" which may be particularly salient to people because of their interactional significance.¹⁴ Negation in discourse is, for instance, often employed for the purpose of implicit or explicit denials, or in order to make sure that speakers' "addressees have the same orientation to the topic as they have themselves, in terms of either shared background knowledge or their personal stance" (Cheshire 1996: 6). For the latter case, Cheshire (1996: 6) provides the following example where the addressee (Jacky) corrects the speaker (Jenny) in her assumption that both of her parents are alive:

- (28) *Jenny*: Who is it who tells you off in your family..your mother or your father?
Jacky: Well my mum 'cos I haven't got a dad now..

¹⁴It should be noted that Cheshire's argumentations refer to situational salience (see Section 2.3.1). So, interactional prominence can help explain why a particular feature stands out in a specific conversation. However, I argue that interactional prominence can also have an effect on extra-situational salience, as investigated in the present study. When a feature regularly occurs in interactionally prominent positions in conversations, people might generally become aware of it and recall it consciously outside specific circumstances.

What is more, the *amn't I* and the *isn't it* constructions are tag questions, which Cheshire (1996: 9) considers “explicitly interactive, focussing on the relationship between speakers”, for instance, by expressing uncertainty or sarcasm.

Generally, Cheshire (1996: 8) concludes that “those syntactic environments that require speaker-addressee collaboration in the production of discourse are presumably important to speakers, serving to create involvement, and it is not surprising, therefore, if these salient environments tend to regulate variation” (for example, in that more non-standard forms are used and retained there than elsewhere).

Another factor connected to structural prominence in discourse may be an item’s position at the beginning or end of a linguistic unit (cf. Auer 2014: 13). This could be an additional reason for why some non-standard grammatical features in the present study were more structurally conspicuous than others. Among these features with high rates of perceived non-standardness were, for example, the negation structures just discussed and focus fronting. Focus fronting might be particularly prominent in that context, since the salient item is placed sentence- or clause-initially *for the purpose of focus*.¹⁵

A further structural factor that could influence a feature’s cognitive prominence, as described by Auer (2014: 13–14), is the fact that the feature involves content words rather than function words. Auer (2014: 13) assumes that speakers of a language variety are more aware of linguistic items that carry meaning (thus, lexical items are generally more salient than syntactic ones) or at least distinguish meaning (hence, phonemic features are typically more salient than subphonemic ones). One could argue here that from a cognitive linguistics perspective, grammar is indeed meaningful, instead of being an entirely formal system, and that it “has much to tell us about both meaning and cognition” (Langacker 2013: 5). But as I understand it, Auer’s comment refers to judgments of non-linguists, to whom some structures might in fact sound more meaningful

¹⁵ Another potential factor connected to an item’s position in a sentence or clause is prosodical prominence. Features in prosodically prominent positions might stand out acoustically (cf. Kerswill and Williams 2000: 69). Since the perception of non-standard grammar in the present study was tested by means of written stimuli, prosodical prominence is not analysed in detail here. It could be an interesting object of study for future approaches to morphosyntactic salience, though, especially in the context of situational salience (i.e. exploring why particular features stand out in specific situations).

than others.

As for the features in the present study, several of those with a high degree of perceived non-standardness involved content words, as the zero past tense of *walk*, *wants petted* in the want/need + past participle construction or *Sort of a student* in the context of focus fronting. Some features with a lower degree of non-standardness involved function words, for example, *they* replacing *there*, the indefinite *a* instead of the definite *the* and *her* replacing *herself*.

Explanations of saliency based on grammatical domains and content versus function words proceed from linguistic categorisations of language. While linguists are able to classify and describe different non-standard and standard uses of features in a fine-grained way, the dividing line between what is “standard” and what “non-standard” may be not that clear in laymen’s perceptions. An additional reason for the low cognitive saliency of non-standard habitual progressives and the inverted word order in indirect questions may be that in these cases the distinction between standard and non-standard functions is too subtle for non-linguists. As pointed out before, there are various uses of *-ing* forms in Standard English, including, for example, habitual progressives. What is special about Welsh English habitual progressives is that they are used in contexts without a clear temporal limitation – a function much less common in Standard English (see Section 3.2). While dialectologists are aware of this functional nuance, it might not be relevant for folk-linguistic perceptions. The *-ing* form may be registered on a more general level, with listeners being unaware of non-standard functions and regional signals, as progressives are perceived to be present all over the British Isles.

Similarly, the inverted word order in indirect questions might not be seen as a clear deviation from the standard language since the dividing line between direct and indirect speech is rather fluid in spoken language. According to Günther (1997), syntactic constructions of indirect and direct speech manifest themselves in a continuum rather than a strict dichotomy. There are many hybrid forms in everyday speech, with speakers, for example, evaluating and interfering while using direct speech. The inverted word order in indirect questions also is an example of a construction where characteristics of indirect and direct questions merge (structurally). It might be hard for laymen to

clearly identify the feature as “non-standard” given the range of other mixed forms and the blurred distinction between direct and indirect speech in general.

To sum up, the results in this section suggest that the perception of deviations from the standard language contributes to salience for outsiders to some extent (while probably being less decisive for insiders). Why some features are perceived as larger deviations from the standard than others can be explained on the basis of structural language-internal factors. Among these factors may be the interactional prominence of grammatical domains, a feature’s position within a linguistic unit, the use of content versus function words and the degree to which a feature is distinguishable from Standard English structures with similar forms and/or functions. The qualitative analyses cannot fully determine which of these structural factors are the most influential ones, but I hope that they inspire future research, for example, in the form of perception experiments with large amounts of non-standard features which are systematically balanced according to several formal criteria.¹⁶

On a final note, I would like to bring frequency in again. Although it can be assumed that a feature’s perception as deviation from the standard language is, to some extent, determined by its structure, it may also have a connection to frequency. To illustrate this point, let us briefly return to the questionnaire discussed in this section. For *every* Welsh English feature investigated, there were more people in the London group considering it a deviation from the standard than in the Welsh group. This means that more Welsh people than Londoners assumed the structures investigated to actually conform with Standard English. This is presumably connected to usage frequencies in that the Welsh informants, especially those who are non-mobile and whose social networks

¹⁶An example of such a formal criterion would be a feature’s position in a sentence/clause. A future experiment could test whether non-standard features are perceived more consciously when they are placed sentence-initially or -finally than when they occur in sentence-medial position. For such an experiment, a range of features could be selected that have the potential to occur in different positions in sentences/clauses (e.g. beginning, middle, end). The experiment could involve test sentences as used in the present study, including one non-standard feature each, where subjects have to state after each sentence whether they have perceived a deviation from the standard language (cf. also Elmentaler, Gessinger and Wirrer 2010). Each feature would occur in three sentences: one time each sentence-initially, sentence-medially and sentence-finally. It could then be analysed whether the features’ rates of conscious perception differ with regard to their position in the sentence. Note that a necessary preliminary step would consist in conducting a rating study to make sure that only those features are included in the experiment that occur *naturally* in different positions. Some constructions might have the potential to occur in all positions, but in fact they appear in just one position most of the time.

mainly consist of people speaking the local dialect, usually only use the non-standard forms in everyday life, not knowing that they constitute deviations from the standard. By contrast, regionally mobile Welshmen and -women who know people speaking other dialects of English and who have witnessed that some features are used less frequently outside of Wales than inside might be more aware of the non-standardness of such features. Thus, although the figures and discussions in this section offer interesting general tendencies, we have to expect “[e]xperience-driven differences” (Kuperman and Van Dyke 2013: 802) in the familiarity with linguistic structures. Since “salience, however defined and however caused, will be different for different social groups” (Kerswill and Williams 2000: 86), the next section aims to shed light on several social factors and their potential impact on salience.

6.5 Salience and social factors

The empirical approach to frequency presented in this work parallels the so-called From-Corpus-to-Cognition Principle, introduced by Schmid (2000). The central idea of the principle is that “observed frequencies in the actual use of a language correlate with degrees of preferences in the linguistic system” (Schmid 2000: 39). Cognitive representations of language, according to the principle, can be predicted on the basis of usage frequencies in corpora. This means that “statistical generalizations over the collective performance of a linguistic community as attested in large corpora reflect the linguistic competence of individual speakers” (Blumenthal-Dramé 2012: 30). Although a range of studies show that the statistics of language use shape our perception of language (see Chapter 2), corpus data can only go as far as being “a yardstick for language representation in the brain of an average language user (which may, in turn, be rather weakly representative of actual brains)” (Blumenthal-Dramé 2012: 65). People differ in their degree of familiarity with particular linguistic structures (cf. Kuperman and Van Dyke 2013: 802), and, in fact, it is important to take this into consideration when working within a usage-based framework. Assuming that usage frequencies affect our mental representations of language, these representations clearly have to be unique for everyone, since we differ in the quantity and quality of linguistic information we absorb every day (cf. Taylor 2012: 16). Thus, in addition to documenting general statistical tendencies – which can be helpful sources for theorising sociolinguistic salience and its causes – the present study also examines interindividual differences in percep-

tions of dialect grammar.

As described in Section 5.2.1, the salience questionnaire included a range of questions concerning the subjects' linguistic and social backgrounds. Some of these questions, exploring factors such as age or educational qualifications, were identical for the informants from Wales and those from London. In the following, these factors are analysed from a comparative angle to see to what extent they affect intra- and interlectal salience. After that, several factors are scrutinised that are particularly interesting in the context of the Welsh informants (e.g. native language Welsh versus English) or the subjects from London (e.g. friends from Wales).

6.5.1 Subjects from Wales vs. subjects from London

The factors analysed in the context of both intra- and interlectal salience were: sex, year of birth, educational background, attitudes to dialectal diversity and friends who speak (an)other dialect(s) than one's own, detailed in Table 15. The analysis of social factors and their potential influences on salience was carried out by means of generalised linear models, using the `glm` function in R (cf. Baayen 2008: 197). The models contrasted the effects of different independent variables (social factors) on the dependent variable (salience; see Chapter 5). The significance of an effect is expressed by the p-value. The data, taken from the salience questionnaire, encompassed 300 informants; 150 from Wales and 150 from London. Salience was operationalised as the number of Welsh English features each person correctly assigned to Wales.

The first model predicted intralectal salience on the basis of the social factors just described. According to the model¹⁷, displayed in Table 16, the variables `sex`, `highest.qualif` and `friends.other.dialect` do not have any significant effects on intralectal salience. This means that the results do not differ significantly between female and male participants, between people with different educational qualifications and between those who have or do not have friends speaking (an)other dia-

¹⁷As for the factor `highest.qualif`, the level "A-Levels" is not included in Table 16, since it functions as the intercept with which the other levels (such as "PhD") are compared. The model was rerun several times with each of the other levels as intercepts, too, using the function `reorder` in R (cf. Gries 2009: 85). The tendencies after releveilling remained the same with no significant effects for `highest.qualif`.

Factor	Factor label in the model	Levels
Sex	<code>sex</code>	Female, male
Year of birth	<code>year.birth</code>	Welsh subjects: 1922–1998 London subjects: 1932–1996
Highest educational qualification	<code>highest.-qualif</code>	None, GCSEs, A-Levels, Bachelor's, Master's, PhD, other
Stances to the statement "I like hearing a range of dialects"	<code>hearing.-dialects</code>	1–7 (1 denotes "strongly disagree", 7 "strongly agree")
Having friends who speak a dialect of English other than one's own	<code>friends.-other.dialect</code>	Yes, no

Table 15: Social factors explored in the context of both intralectal and interlectal salience

lect(s). The remaining two factors, though, show significant effects on salience, namely `hearing.dialects` ($p < 0.05$) and `year.birth` ($p < 0.01$). In more concrete terms, those Welsh people who were more open towards linguistic diversity were overall better at identifying the Welsh English features. This finding strengthens Garrett's (2010) assumption that stances towards linguistic diversity are a powerful factor in people's evaluations of language (see Section 5.2.1).

The effect of the informants' years of birth on salience implies that the older informants recognised more of the Welsh English features than the younger informants did. As discussed in Section 5.2.1, there are two possible explanations for this finding. First, it could be that the older speakers were generally more experienced and had a deeper linguistic knowledge of Welsh English. Second, it may be that some of the features were still used by the older informants while they were not common any more amongst the younger population. Paulasto (2006) offers some evidence for that claim, finding that focus fronting and non-standard progressive forms are currently undergoing processes of dialect levelling and are used less frequently by the younger than by the older generations in her datasets. The Radio Wales Corpus supports these observations. As discussed in Section 6.2, the corpus material provides apparent-time evidence for a diachronic decline in the frequency and range of non-standard features in Welsh English. Thus, the fact that older people were better at assigning the Welsh features to Wales in

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	30.296831	11.922898	2.541	0.01105	*
hearing.dialects	-0.140697	0.062668	-2.245	0.02476	*
sexmale	-0.257942	0.199386	-1.294	0.19577	
year.birth	-0.015556	0.005977	-2.603	0.00925	**
highest.qualifBachelor	-0.120868	0.267718	-0.451	0.65165	
highest.qualifGCSEs	-0.271535	0.484396	-0.561	0.57510	
highest.qualifMaster	-0.094122	0.339727	-0.277	0.78174	
highest.qualifNone	-0.056677	0.681083	-0.083	0.93368	
highest.qualifOther	-0.341044	0.386423	-0.883	0.37747	
highest.qualifPhD	-0.044451	0.446448	-0.100	0.92069	
friends.other.dialectyes	0.125530	0.314335	0.399	0.68963	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Table 16: Results of the generalised linear model with intralectal salience as the dependent variable and several social factors as independent variables

the questionnaire may be connected to them using those features to a greater extent in everyday life than the younger subjects.

One might wonder now whether there are age-related differences in the recognition of individual features. To test this idea, I implemented several versions of the statistical model just discussed. Each version had one non-standard grammatical feature as the dependent variable (instead of the number of Welsh English features identified). So, for example, it was determined to what extent focus fronting was recognised rather by the older than by the younger informants. According to the models, the following non-standard features are influenced significantly by `year.birth` in that they were more often assigned to Wales by older than by younger subjects: the non-standard habitual progressive ($p < 0.05$), the inverted word order in indirect questions ($p < 0.05$) and the invariant tag question *isn't it* ($p < 0.01$). This implies that the older the subjects were, the stronger their mental connection was between these features and Welsh English. By contrast to that, no age-related effects were found for focus fronting, the zero past tense of regular verbs and the *that* construction. A speculative explanation involves frequency. Focus fronting, although less frequently used by younger generations, generally still occurs with quite high token frequencies in present-day Welsh English (in relation to the other features investigated and as attested in the Radio Wales Corpus). This might conserve the construction's prominence among younger genera-

tions. By contrast, the zero past tense of regular verbs and the *that* construction had the lowest frequency values in the RWC. These low probabilities of occurrence might lead to the constructions neither being consciously noticed by many younger nor by many older people. In addition to processes of dialect levelling, the effect of age on the recognition of the invariant *isn't it* may be related to increased uses of *innit* tag questions in the younger generations in Wales. It could be that the younger Welsh subjects, similar to the London subjects, identify with the use of *innit* rather than *isn't it*.

To test influences of social factors on interlectal saliency, a model was run on the basis of the London informant data. The independent variables, again, were `sex`, `year.birth`, `highest.qualif`, `hearing.dialects` and `friends.other.dialect`. The only independent variable that had a significant effect on saliency in this model was `hearing.dialects` ($p < 0.05$). Thus, in the London group, too, informants with more positive attitudes towards linguistic diversity were better at identifying the Welsh features. Figure 19 displays the linear effects of `hearing.dialects` on intra- and interlectal saliency. By contrast to the Welsh informants, the factor `year.birth` did not make a difference in the group from London, neither did the factors `sex`, `highest.qualif` and `friends.other.dialect`. A hypothesis would be that the effect of `year.birth` is more pronounced in combination with speech production (in the case of older Welsh informants, who may actively use the features) than speech perception (in the case of older and younger London informants, who might rather know the features from hearing them, but not so much from their own language use.)

6.5.2 Subjects from Wales: Native language, region in Wales

As described in Chapter 3, Welsh is an official language of Wales and still spoken by about 20 per cent of the population. In Chapter 5, it was mentioned that the native language of 11 per cent of the Welsh informants in the saliency questionnaire was Welsh, 15 per cent considered both Welsh and English as their native tongues and 74 per cent of the Welsh subjects indicated that their native language was English. Chapter 5 also brought forward the hypothesis that native speakers of Welsh are better at identifying some Welsh English features (tracing back to the Welsh substrate) because they also

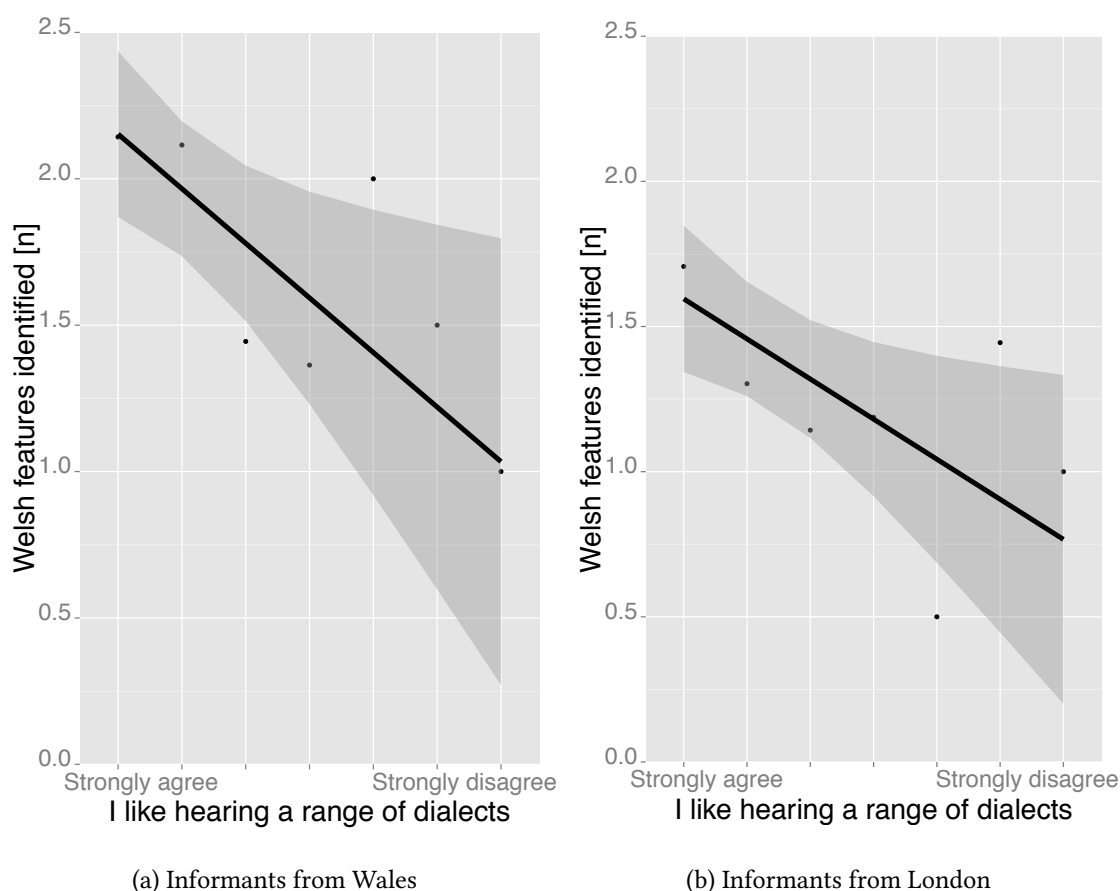


Figure 19: The relation between the subjects' openness towards linguistic diversity and the number of Welsh English features assigned to Wales

know the structures from the Welsh language.

Native language (*native.lang*) was integrated in the model discussed earlier in this section with intralectal salience as the dependent variable (see Table 16). According to the model, *native.lang* does not have a significant effect on salience. The test results for native speakers of (1) Welsh, (2) English and Welsh and (3) English did not differ from each other significantly. One possible explanation is that the Welsh English features are used in Wales both by Welsh- and by English-speaking people. Taking a look at the Radio Wales Corpus confirms this assumption. Focus fronting, the invariant *isn't it/innit*, the non-standard habitual progressive, the inverted word order in indirect questions and the zero past tense of regular verbs were all produced both by speakers who spoke and by speakers who did not speak Welsh.

Next, I looked more closely at the different regions in Wales where participants lived. As discussed in Section 3.1.2, there is linguistic diversity within Wales, with the English of the western regions probably being the most susceptible to Welsh-language influences, while English dialects in the eastern border areas are (also) shaped by neighbouring English English dialects. To find out if this linguistic diversity manifests itself in perceptions of dialect grammar, the Welsh informants were assigned to groups according to their places of living. The groups were based on the dialect areas of Wales described by Garrett, Coupland and Williams (1999: 325): South East (industrial), South East (valleys), South West, Mid Wales and North West. (The North East was excluded from the calculations since the “group” would only have consisted of one speaker.)

The factor *region* was integrated into the model predicting intralectal salience (see Table 16). The model could not find a significant effect for *region*. In other words, the salience ratings did not differ significantly between test subjects from different parts of Wales. This result can be substantiated with findings from the Radio Wales Corpus. As discussed in Section 6.2, the five non-standard features (the *that* construction excluded) all occur in several regions of Wales, each at least in one northern and one southern area. It may well be that the features’ prevalence across Wales leads to similar test results for people from different parts of the country.

6.5.3 **Subjects from London: Contact with Welsh English**

As for the London informants, I assumed that the intensity of contact they had with Welsh English had an effect on how they perceived Welsh English grammar. According to my frequency-based line of argumentation, Londoners who hear Welsh English frequently recognise particular Welsh English constructions more readily than Londoners without any direct contact to English dialects in Wales. As discussed in sections 5.2, 5.4 and 6.2, contact with Welsh English does not necessarily mean knowing someone from Wales personally, but it can also refer to being familiar with the speech of Welshmen and -women in the media. My analyses found that the frequencies of occurrence of several non-standard grammatical features show similar tendencies in everyday speech (Radio Wales Corpus) and in the media language of the presenters Rob Brydon, Rhod Gilbert and Alex Jones as well as in the series *Gavin & Stacey* (see Section 6.2). This prompts the hypothesis that perceptions of Welsh English grammar of Londoners are similarly affected by having personal contacts from Wales and by

knowing Welsh TV presenters (in case the intensity of exposure to the two sources is the same).

Contact with Welsh English was operationalised in the salience questionnaire using the following factors (see Section 5.2.1): (1) knowing Welsh people (friends, family members, work colleagues); here, the Londoners both had to indicate whether they generally knew people from Wales and then more specifically whether they knew people who speak a Welsh dialect of English. As for the latter case, it was furthermore explored how frequently the London informants spoke with those people. Connected to knowing Welsh people was (2) the question whether the Londoners had ever been to Wales, and if so, how much time they had spent there. As well as personal contact with Wales and Welsh people, the questionnaire (3) investigated whether the London informants knew presenters or comedians from Wales, and, if so, who those presenters and comedians were. Moreover, subjects had to state (4) the Welsh TV programmes they knew (if there were any).

With regard to (1), the independent variable `friends.wales` was integrated into the generalised linear model used above (showing the effects of social factors on interlectal salience). According to the model, `friends.wales` has a significant positive effect on interlectal salience ($p < 0.05$). This means that London informants with Welsh friends, family members or work colleagues generally recognised more of the Welsh English features than subjects without any personal contacts from Wales. Integrating the factor `friends.welsh.english` yields similar results (although some people in the questionnaire indicated that they knew a person from Wales but that this person did not speak Welsh English). Generally, more Welsh English features were identified by subjects who knew people speaking a Welsh dialect of English, the effect again being statistically significant with $p < 0.05$. Next, the independent variable `how.often.speak` was integrated into the model, denoting how frequently the Londoners who knew speakers of Welsh English actually talked to them. The scale here ranged from “every day” to “never”. The model could not find a significant effect for the variable. Ironically, those who indicated speaking with their friends every day had the lowest recognition rates of the Welsh features, lower than the subjects stating that they actually never talked to their Welsh friends. I assume that subjectivity played a role here, both in the processing of language (some Londoners might be ex-

posed to Welsh English less than others and still register particular features) and in descriptions of linguistic activities (“never” talking to friends may really mean “never” to one person, while it denotes rare occurrences to another). It would be interesting to have this factor analysed in more depth in future studies, for example, by conducting interviews where individual linguistic preferences and activities can be documented in more detail.

The factor complex (2) explored the extent to which a London informant’s having spent time in Wales had an influence on his or her perception of Welsh English morphosyntax. First of all, the model just discussed was augmented with the factor *been.to.wales*, indicating whether or not the subjects had ever been to Wales. No significant effect could be attested for that factor. Londoners who had been to Wales already were slightly, but not significantly, better at recognising the Welsh English features in the questionnaire. Still, I wanted to find out if the length of time spent in Wales interacted with salience to some extent. Thus, next the independent variable *how.long.stay* was integrated into the model, indicating the length of the stay(s) in Wales on a scale from “up to a week” to “more than six months”. Again, no significant effect could be found for the variable. Although those who spent more than six months in Wales were overall the best at assigning the Welsh features to Wales, there was no significant difference between them and the informants who had spent shorter periods of time in Wales. The picture for those who had been to Wales for less than six months was quite mixed, for example, with those who only spent up to a week in Wales identifying more Welsh features than those who had been to Wales for up to three months. It seems that in order to be familiar with Welsh English grammar, it is not a necessary requirement to have spent a particular period of time in Wales. As shown above, familiarity with Welsh English can also be connected to knowing people from Wales. And, as argued in the next paragraph, even that might not be necessary, as long as outsiders are familiar with Welsh English as used in the media.

As explained in Section 5.2.1, the salience questionnaire asked the Londoners whether they knew Welsh TV and radio presenters. The results were integrated into the model discussed in previous paragraphs in the form of the independent variable *know.presenter*. No significant effect of *know.presenter* on interlectal salience could be found. The data show that those London subjects who knew a

presenter or comedian from Wales had overall higher recognition rates of the Welsh English features, but the difference to the other group was not statistically significant. Maybe this is because different presenters have different ways of speaking, some of which not being very representative of everyday Welsh English. A similar picture emerges for TV programmes. For the independent variable `know.programme`, no significant effect on salience could be found, indicating that Londoners with or without knowledge of Welsh TV programmes did not differ from each other significantly in the number of Welsh features identified. Clearly, not all TV shows and series connected to Wales express Welshness as explicitly and vividly as *Gavin & Stacey* does. *Doctor Who*, for example, was mentioned by several London subjects in the questionnaire, presumably because the show is filmed in Cardiff. However, according to a survey carried out in Wales by the University of Glamorgan, the links to Wales are not too obvious. Subjects found that *Doctor Who* “denied its Welsh location in all but a few episodes” and criticised “attempts at passing Cardiff off as London” (Blandford et al. 2010: 29). By contrast to that, *Gavin & Stacey* was perceived by subjects as a “credible [...] and authentic representation of a Wales they recognised” (Blandford et al. 2010: 30).

The analysis of video clips in Section 6.2 suggests that representations of Welsh English in shows by Rhod Gilbert, Rob Brydon and Alex Jones as well as in *Gavin & Stacey* are fairly representative of everyday Welsh English, at least regarding the morpho-syntactic features under investigation. Thus, one cannot rule out the possibility that somebody who is familiar with the language of these programmes is familiar with Welsh English morphosyntax to some extent. To find out more about that, the London informants were divided into two groups; people in the first group stated that they knew at least one of the three presenters or *Gavin & Stacey*, people in the second group neither indicated to know Brydon, nor Gilbert, nor Jones, nor *Gavin & Stacey*. Interestingly, adding this as an independent variable `know.presenter.GavStac` to my model returned a significant effect. Those informants that knew at least one of the presenters or the series were significantly better ($p < 0.05$) at recognising the Welsh English features. This means, and it sounds quite obvious in fact, knowing TV shows with a particular regional affiliation can, but does not necessarily affect knowledge of the respective regional dialect. Such a correlation may come about if the media figures

(e.g. presenters, actors) use the regional dialect themselves.

In sum, several social factors seem to interact with intra- and interlectal perceptions of dialect grammar and should thus be included in explanations of sociolinguistic saliency. Intralectal saliency appears to be influenced by attitudes to linguistic diversity and age in my dataset. Interlectal perceptions of Welsh English also correlate with attitudes to linguistic diversity, and moreover seem to be influenced by having Welsh friends and by knowing TV shows where the regional dialect is used. Table 17 summarises the factors that were found to affect intra- or interlectal saliency in the present study.

Intralectal saliency (Wales)	Interlectal saliency (London)
Attitudes to linguistic diversity	Attitudes to linguistic diversity
Age	Welsh friends
	Knowing TV show where regional dialect is used

Table 17: Social factors that the present study found to have statistically significant effects on intra- and/or interlectal saliency

Interestingly, all of these factors can, at least hypothetically, be related back to frequency. Being open to linguistic diversity might lead to dealing with different dialects more often in everyday life and thus being exposed to these dialects more frequently than people with negative attitudes towards linguistic diversity.¹⁸ As for age, people who have more experience with particular features and/or who use them more frequently themselves might also be better at recognising them. Londoners who have friends, work colleagues or family members from Wales might hear Welsh English more frequently than Londoners without such contacts. Finally, the frequency with which one is exposed to Welsh English in the media might also impact the strength of representations of Welsh English grammar in the mind.

¹⁸Note, however, that there could also be a frequency effect in the other direction: a lack of experience with different dialects may result in negative attitudes towards linguistic diversity. It would be interesting to learn more about the interaction of saliency, experience and attitudes in future studies.

6.5.4 The role of incomers

At the close of this section, I would like to discuss the role of incomers, that is to say people who grew up outside of Wales (elsewhere in the British Isles), and who subsequently moved to Wales. As stated at in Chapter 5, the cohort of Welsh informants in the present study involved both “true insiders” who grew up in Wales and incomers. The perceptions of incomers, I hypothesise, can be positioned somewhere between those of insiders and outsiders. Incomers may be more familiar with Welsh English than outsiders but probably less than people who have always lived in Wales.

For the analysis, the Welsh informants were divided into insiders and incomers. First of all, it was determined whether insiders and incomers differed from each other significantly in assigning Welsh English features to Wales. For that purpose, `insiders - incomers` was added as an independent variable to the model used above (predicting intralectal salience on the basis of social factors). This independent variable had the levels `insider` and `incomer`. According to the model, there was no significant difference between the judgements of insiders and incomers. This is convenient for the present study, since the salience values for all informants from Wales – incomers and insiders combined – seem to be sufficiently representative of judgements by “true insiders” from Wales. Still, the model stated that the number of Welsh features identified was slightly higher in the group of insiders than in the group of incomers.

These general tendencies indicate that insiders and incomers in the present study did not differ from each other significantly in the number of Welsh English features correctly assigned to Wales. From these tendencies, however, we cannot directly infer that there are no differences between insiders and incomers regarding individual features. Thus, I next calculated six versions of the `glm` described in the previous paragraph. In each of these models, the salience value (the number of people who assigned a feature to Wales) of one of the six grammatical features was the dependent variable. For example, one model determined whether there was a difference between insiders and incomers in assigning focus fronting to Wales. The models found no significant differences between insiders and incomers regarding any of the grammatical features.

Still, as can be seen in Figure 20, the numbers of informants that considered a feature to be Welsh English were higher among the insiders than among the incomers regarding

the invariant tag question *isn't it*, focus fronting, the inverted word order in indirect questions and the non-standard habitual progressive. The zero past tense of regular verbs and the *that* replaces *to* construction were the only features that received higher salience ratings by the incomers than by the insiders. Reconnecting these results to the frequencies of the features in the Radio Wales Corpus (with the zero past and *that* having the lowest frequency values), it seems that the insiders' salience values were somewhat *more in line with frequency* than the ratings of the incomers. What becomes also noticeable from the graph is that the salience ratings of the incomers were higher than those of the outsiders – for every feature. This fits my hypothesis that the perceptions of incomers can be positioned somewhere between those of insiders and outsiders, presumably closer to those of insiders. An additional glm with data from insiders, incomers and outsiders found a marginally significant difference between incomers and outsiders in the numbers of features assigned to Wales ($p < 0.1$), while, as stated above, no significant difference could be found between the results of insiders and incomers.

Concluding the analyses at this point would still be simplistic, since the group of incomers was heterogenous itself. It involved people who had moved to Wales fairly recently and others who had been living in the country for several decades. According to my experience-based line of argumentation, informants who have been living in Wales for a long time have more experience with Welsh English than those who have moved there recently, thus the former should be better at recognising the Welsh features. The incomers' perceptions could, moreover, be influenced by the areas where they grew up. For instance, if a non-standard feature is used in the local dialect there, too (in addition to Wales), it might be assigned to the area where one grew up and not to Wales (or to both areas).

To shed light on interindividual differences in the group of incomers, another glm was calculated. The dependent variable was *salience* for the incomers, denoting the number of Welsh features identified. The independent variables were *decades* – *in.wales*, indicating for how many decades the subjects had already lived in Wales, and *region.origin*. For the latter variable, the exact places where the informants had grown up were allotted to the larger regions South East of England, South West of

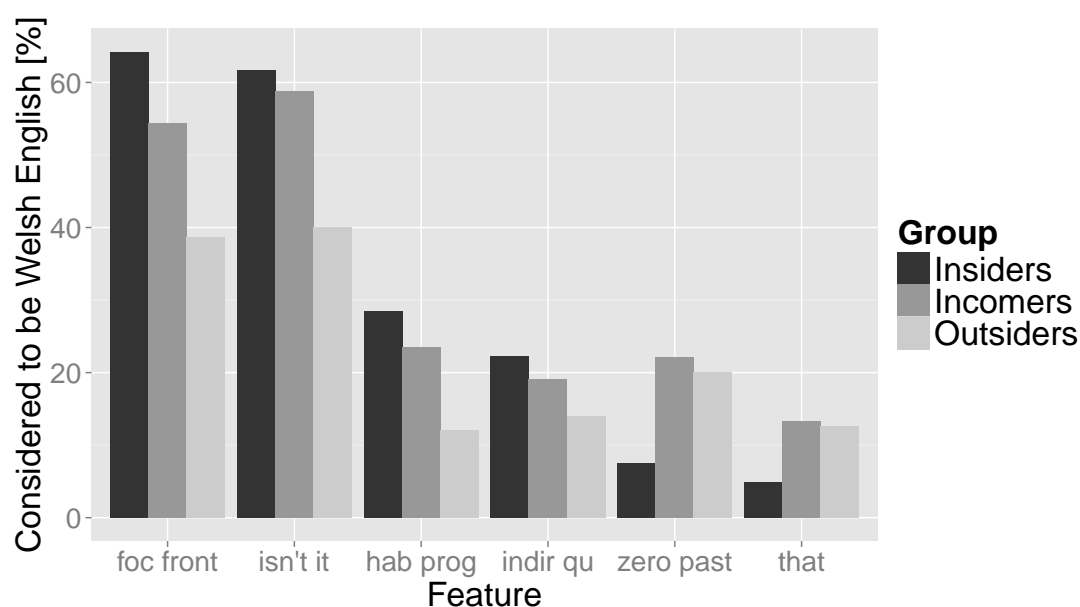


Figure 20: Salience values of non-standard grammatical features for insiders (grew up in Wales), incomers (now living in Wales) and outsiders (living in London)

England, West Midlands, North of England and Scotland.

The model returned no significant effects for `decades.in.wales` and `region.-origin`. This means there were no significant differences in the numbers of features correctly identified as Welsh English between people having lived in Wales for a shorter or for a longer period of time. Still, longer-standing residents of Wales were slightly better at identifying the Welsh English features than people who had moved there recently. As for `region.origin`, the salience ratings did not differ significantly between people who had grown up in different parts of the British Isles.

`decades.in.wales` and `region.origin` did not have statistically significant effects on the numbers of features correctly assigned to Wales in my dataset. However, they may still have affected the salience ratings of individual features. Specific linguistic features may be more or less obtrusive in contrast with one's own dialect. Furthermore, it is plausible that individual – perhaps less frequent – features are perceived as typical of Welsh English only after having lived in Wales for a particular period of time and having accommodated to Welsh English to some extent, while other features stand out right from the beginning. To find out more about that, six versions of the `glm` used in the previous paragraph were calculated, each time with the salience of

one feature (the number of people who assigned the feature to Wales) as the dependent variable. For example, one model explored the relation between the salience value of focus fronting and `decades.in.wales` as well as `region.origin`. In these models, however, none of the grammatical constructions was affected significantly by `region.origin` or by `decades.in.wales`.

To sum up, the salience ratings by insiders and incomers did not differ from each other significantly, both regarding the number of Welsh English features identified and regarding the salience values of individual features. Still, insiders were generally slightly better than incomers at assigning the Welsh English features to Wales. This could be connected to insiders being somewhat more familiar with the constructions than people who came to live in Wales at a later point of time. Incomers, on the other hand, were better at assigning features to Wales than outsiders. These findings suggest that perceptions of language are indeed usage-sensitive in that they are affected by exposure to regional speech with different frequencies.

6.6 Summary and discussion

This chapter has explored to what extent salience in dialect grammar can be explained by means of on a usage-based approach. The chapter began by presenting general results from the salience survey, finding that focus fronting and the invariant tag question *isn't it* were significantly more salient than the other grammatical items under investigation, for both insiders (from Wales) and outsiders (from London). The analysis of token frequencies in Welsh English found more frequent features to be – on a general level – seen as more characteristic of Welsh English. This applied to insiders and outsiders, but insiders' salience judgements correlated better with token frequency in Welsh English. This provides evidence for the usage-based character of dialect perceptions: Welsh people, who had more experience with the Welsh English features than Londoners, were more accurate in their recognition of the features. While (comparatively) high token frequency in the target variety accompanies some salient grammatical features (focus fronting), the study found that it is no necessary requirement for salience in morphosyntax. The salient invariant *isn't it* occurred with rather low token frequencies in the Radio Wales Corpus. The comparative analyses of RWC and LI found the feature to be more frequent in Welsh English than in London English,

though, which may explain its high salience for the Londoners. Overall, these relative frequency differences appear to be a good predictor of outsiders' salience judgements, being more reliable than frequencies in Welsh English only.

Before moving on to other potential determinants of salience, it was necessary to ascertain that the frequencies just mentioned were not tied to particularities of individual corpora, which would make the conclusions unreliable. Contrasting frequencies of non-standard and functionally equivalent standard constructions in the Radio Wales Corpus and the Linguistic Innovators corpus offered validations for the frequencies of the non-standard features. These studies also found the two corpora to be not too different in nature, suggesting that they may be appropriately complementary for comparative linguistic analyses. Another comparative linguistic analysis involved contrasting occurrences of non-standard features in the RWC with those in TV shows featuring Welsh presenters and actors. The rationale here was that some Londoners' perceptions of Welsh English might not so much be influenced by being familiar with "everyday" Welsh English (as approximated by the RWC), but rather by listening to Welsh people in the media, who might consciously use or avoid particular features for specific purposes. In the TV clips analysed, however, the frequencies of the non-standard grammatical features showed similar tendencies as in the Radio Wales Corpus. Thus, it seems that the RWC is a sufficient means to draw frequency-based conclusions on people's perceptions of Welsh English. It should be kept in mind, however, that such frequency-based conclusions refer to the language use of an average speaker of Welsh English. The more in-depth analyses with the RWC pointed to variation in Welsh English, for example, regarding age, sex and place of living in Wales.

Apart from usage frequencies in Welsh English and London English, the present study suggests that the degree to which a feature is seen as characteristic of Welsh English by outsiders is also influenced by the feature's pervasiveness across the whole of the British Isles. The eWAVE analysis found interlectally salient constructions to be generally more pervasive in Wales than in the rest of the British Isles. No correlation could be determined between salience and geographical spread, i.e. the number of regions where a feature occurs in the British Isles outside Wales. The analyses suggest that features restricted in use to the target area (e.g. Wales) do only carry regional signs

when they are quite frequent *within* the target area.

The degree to which a feature is considered as a deviation from the standard language, moreover, seems to correlate with salience, at least with regard to the perceptions of outsiders. While the perceived non-standardness of features did not seem to play a larger role for intralectal salience (usage frequencies in one's own variety probably being the more influential factor here), it seems to have some effect on interlectal perceptions. Potential structural factors leading to perceived non-standardness could be interactional prominence, the position at the beginning or end of a linguistic unit, the fact that a feature involves content rather than function words and the ease of distinguishing a non-standard item from formally and/or functionally similar features in the standard language. Frequency, however, probably also influences the perception as "non-standard", for instance, in the case of non-mobile people who only use and hear a non-standard form in their community, and who are not aware of the fact that it constitutes a deviation from the standard language.

The fact that people are aware of deviations from the standard to varying degrees points to interindividual differences in perceptions of dialect grammar. Salience is subjective, and the present study discusses several social and attitudinal factors that impact on how a grammatical construction is perceived. For example, attitudes to linguistic diversity seem to affect both intra- and interlectal salience in that people who are more open towards hearing different dialects are generally better at recognising dialectal features. Furthermore, intralectal salience seems to be related to age; the older informants in the salience study had higher recognition rates of the Welsh English features than the younger subjects, possibly because some features are used more by (and thus are more expectable to) the older than the younger generations. Someone's background as a true insider (grown up in Wales) or an incomer (moved to Wales later) did not have a significant effect on salience. Nevertheless, the findings suggest a (slightly) higher accuracy of assigning a feature to Wales when having grown up there than when having moved to Wales later. Interlectal salience, as I propose, can be influenced by having or not having friends or acquaintances from Wales. Furthermore, knowing TV shows where Welsh English is used (here: shows by presenters Rob Brydon, Alex Jones and Rhod Gilbert plus the series *Gavin & Stacey*) also seems

to have a positive effect on the recognition of dialectal features.

The findings prove that salience is indeed a complex construct, conditioned by the interaction of several determinants. Table 18 summarises the results for the individual features under investigation. For each determinant (e.g. token frequency in Welsh English), a “+” marks the two features with the highest values. For example, focus fronting and habitual progressives receive a “+” in the category token frequency in Welsh English. On the basis of these data, several conclusions on salience in dialect grammar can be drawn. First, the salient features focus fronting and invariant *isn’t it* both show high values in *several* categories. Focus fronting is among the features with the highest token frequencies in Welsh English, token frequencies in Welsh English versus London English and overall pervasiveness values. *Isn’t it* also receives a “+” with regard to frequency differences in Welsh English and London English and is one of the features with the highest perceived non-standardness values. The less salient constructions are either marked with just one “+” (habitual progressive, zero past, *that* replaces *to*) or none (inverted word order of indirect questions). This underlines that regional indexation in morphosyntax originates from the interplay of several factors and that the effect of just one determinant (e.g. high token frequency in target variety or structural prominence) may be not enough.

Feature	Token freq. WE	Token freq. WE vs. LE	Pervasiveness Br. Isles	Perceived as non-standard
Focus fronting	+	+	+	
Invariant <i>isn’t it</i>		+		+
Hab. progressive	+			
Zero past				+
<i>that</i> replaces <i>to</i>			+	
Indir. question				

Table 18: The feature-dependent determinants of salience explored in the present study and their relation to the features under investigation. A “+” marks the two features with the highest values in a category. The two most salient features, focus fronting and invariant *isn’t it*, are highlighted in bold print

Second, the determinant shared by focus fronting and invariant *isn’t it* is frequency differences in Welsh English versus London English. The two salient features show the highest relative frequencies in the Welsh corpus as compared to the data from Lon-

don. This indicates that token frequency differences between the target variety and the variety of the outsiders play a vital role for interlectal salience. More generally, it provides evidence for the hypothesis that sociolinguistic salience in dialect grammar is usage-based to some extent.

Third, if the criterion of high token frequency in the target variety is not met by a feature (invariant *isn't it*), structural prominence might play a similarly important role in promoting a feature's salience. The salience of focus fronting can be explained based on its probability of occurrence (token frequency in Welsh English, frequency differences between Wales and London, pervasiveness in the British Isles). The effect of these experiential factors seems to be weaker for *isn't it*. However, the feature was perceived as a deviation from the standard language by the vast majority of informants. I suggest that if a feature's usage-based characteristics are less distinct, this can be compensated for by a high degree of structural conspicuousness. Still, for regional indexation structural prominence needs to go hand in hand with frequency to some extent. This explains the rather low salience of the zero past tense of regular verbs as compared to the invariant *isn't it*: the zero past tense is also structurally obtrusive but occurs with low token frequencies in Welsh English and in Wales versus London. Structural prominence alone does not bring along regional signals.

In addition to these findings related to individual features, some conclusions on salience perceptions of insiders and outsiders can be drawn. Table 19 showcases factors analysed for intra- versus interlectal salience. These are subdivided into feature-dependent factors (as token frequency in Welsh English) and subject-dependent factors (as age). For each factor, the table indicates whether it correlated better with insiders' or outsiders' perceptions in the present study.

The salience of individual grammatical features seems to be more closely tied to token frequency in the target variety for insiders than outsiders. This may be connected to insiders being more familiar with features' probabilities of occurrence in their own dialect. People from Wales may thus recognise Welsh English constructions with sufficiently high token frequencies even if these constructions are not very obtrusive from a structural point of view. For outsiders, frequency differences between the target variety and their own variety seemingly have a larger effect on how non-standard fea-

	More relevant for insiders	More relevant for outsiders
Feature- dependent	Token frequency in WE	Frequency differences in WE vs. LE Pervasiveness in WE vs. rest of Br. Isles Perceived deviation from standard
Subject- dependent	Age	Contact with WE (personal, media)

Table 19: Determinants of salience explored in the present study and their relevance for intralectal versus interlectal salience. The findings are based on correlation calculations and linear regressions

tures are perceived. An item's pervasiveness in the target variety as compared to the rest of the British Isles is another helpful indicator of regional affiliation for outsiders. These relative probabilities may play some role for insiders' judgements, too, but probably only for mobile Welsh people who are actually aware of dialects other than their own. A feature's perceived non-standardness also appears to affect interlectal salience more strongly than intralectal salience. This could be attributable to frequency in that Londoners perceive features with low probabilities of occurrence in their dialect as deviations from the standard, while the features are more expectable and natural to Welsh people. With regard to subject-dependent factors, the study found that the age of informants has a larger effect on insiders' than outsiders' judgements. This may be due to insiders' perceptions being more closely linked to their own language use. And based on current processes of dialect levelling in Wales, older speakers are expected to use the non-standard forms investigated more extensively than younger ones. A significant factor for perceptions of outsiders is the degree to which they are in contact with the target dialect, be it through personal acquaintances or via the media. (Of course, contact with Welsh English also plays a role for insiders, but the very fact that they live in the target area suggests that they *all* are exposed to Welsh English on a very regular basis.)

While these factors affect insiders' and outsiders' judgements to different degrees, there are some fundamental commonalities between the groups. This work suggests that both intra- and interlectal salience are usage-based. In both cases, salience relies to some extent on listeners' experiences with their own way of speaking and their knowledge of probabilities of occurrence in the local dialect, be it Welsh English or

London English. Additionally, insiders' and outsiders' perceptions of regional features are shaped by their openness to dialectal diversity. Being interested in other ways of speaking may result in an increased knowledge of dialectal forms, rendering dialect perceptions more accurate.

How do feature-dependent and subject-dependent determinants of salience interact? I assume that feature-dependent factors provide necessary conditions for a variable to become sociolinguistically salient – for the *average* language user. To what extent these factors come into play for *individuals*, though, is determined by subject-dependent factors. So, for example, a construction may theoretically be a good candidate of becoming interlectally salient due to its frequency of occurrence (in the target variety and in relation to (an)other variety/ies) and its perception as a deviation from the standard language. However, it might not be consciously perceived by someone who has negative attitudes towards hearing dialects other than his or her own. These argumentations go in the direction of Purschke's (2014) distinction between the *salience potential* and the *salience perception* of a linguistic item. In his view, *salience potential* refers to the phenomenon-related quality of salient regionalisms as a precondition for perceptual conspicuousness. This does not, however, allow for direct conclusions about whether salient features are subjectively conspicuous to an individual listener and are part of listener-based interpretations (*salience perception*, cf. Purschke 2014: 31, 33).

In sum, I suggest that regional grammatical features have to pass through two filters to be perceived as salient by individuals: the feature-based filter, conditioned by factors as token frequency differences between varieties, and the subject-based filter, based on factors as attitudes to dialectal diversity. Importantly, the requirements for passing through the filters always depend to some extent on the individual feature and the individual listener in question. Salience is context-dependent and the contexts of using particular features and of using dialects in general are unique to everyone.

Chapter 7

Conclusion and outlook

This work has tested the potential and limits of a frequency-based approach to explaining the salience of regiolectal morphosyntactic features. The present chapter sums up the core findings and, through critical reflection, relates them back to the hypotheses formulated in Chapter 4. This is followed by discussing general gains and implications of the study for future research.

7.1 Relating the results to the hypotheses

Hypothesis 1: The sociolinguistic salience of a dialectal grammatical feature is affected by its frequency in language use.

The corpus-based analyses revealed that salience in Welsh English grammar can to some degree be predicted by frequencies of use. Across the different test subjects, more (intra- and interlectally) salient features generally occurred with higher token frequencies in the Radio Wales Corpus than less salient constructions. In addition to that, a positive correlation between interlectal salience and frequency differences in Welsh English versus London English was found. In other words, features with lower probabilities of occurrence in London speech as compared to English in Wales were more salient to the Londoners. This agrees with Rácz (2013), who also found frequency differences to be influential regarding dialectal and sociolectal perceptions of outsiders. Also, the fact that the judgements of insiders (raised in Wales) correlated better with frequencies in Welsh English than the judgements of incomers (moved to Wales later), which in turn correlated with frequencies better than the judgements of

outsiders (from London), illustrates the experiential basis of dialectal perceptions. In sum, the findings offer evidence for Hypothesis 1 by showing that linguistic perceptions are, to some extent, grounded in usage frequencies. The qualification “to some extent” is important, though, since the analyses also found that salience in morphosyntax is not *categorically* tied to high usage frequencies. The invariant tag question *isn't it* was highly salient in this study, but was not very frequent in the Radio Wales Corpus (as compared to the other features investigated). The *that* replacing infinitival *to* construction could not be found at all in the Radio Wales Corpus and, although it was the least salient feature in the present study, people had particular attitudes towards it and some assigned it to Wales even though it is probably not really present there. This suggests that a purely quantitative approach does not do justice to the complex construct of salience. Although the statistics offer some interesting general tendencies, it is necessary to take a detailed look at each individual feature since different determinants may be more or less distinct from case to case (cf. Hettler 2014: 86). This is in line with Rácz (2013: 155), who states that “whenever the salience of a variable becomes an issue, one should refrain from sweeping generalisations and proceed with the attention and caution the subject deserves”.

Hypothesis 2a: Both intra- and interlectal salience are influenced by token frequencies of dialectal grammatical features in the target variety and by the frequencies of those features across different (close-by) varieties.

Hypothesis 2b: Interlectal perceptions are affected by token frequency differences between one's own and the target variety.

The complexity of sociolinguistic salience also becomes apparent with regard to the regional origin of the listeners. As for Hypothesis 2a, the present study found positive correlations between frequencies in Welsh English and salience for both insiders from Wales and outsiders from London. However, two limitations to this finding need to be addressed. First, (comparatively) high token frequencies in the target variety are no categorical requirement for salience, as shown by the example of the invariant tag question *isn't it* above. Second, the correlation between token frequency in Welsh English and salience was weaker for the London group. It seems that, along the lines of Hypothesis 2b, frequency differences between Welsh English and London

English are the more powerful explanatory factor of interlectal salience. Furthermore, the eWAVE study suggests that interlectal salience is affected by the pervasiveness of a feature in the native area versus the rest of the British Isles. No visible relation between intralectal salience and pervasiveness in Wales versus outside could be determined, however. This could imply that insiders' perceptions of non-standard features are influenced more strongly by token frequencies in their own regional dialect than by probabilities of occurrence in other dialects. However, this cannot be more than a cautious assumption. It needs to be kept in mind that the indices of pervasiveness in eWAVE are no frequency counts and that they are only rough indicators of actual language use. Carrying out the same analyses on the basis of corpus frequencies might lead to different results.

Hypothesis 3: The sociolinguistic salience of a dialectal grammatical feature is influenced by its geographical spread, its linguistic structure and individual personal, social and attitudinal characteristics of the listeners.

The eWAVE study did not return a connection between the geographical spread of a feature and salience in the sense that more areally restricted features were generally more salient. The data suggest that features confined to a small-scale area carry regional signals only when they are quite pervasive within this area.

As far as structural factors are concerned, the analyses imply that they play a larger role in interlectal than in intralectal salience. To the informants from Wales, salient features of Welsh English did not necessarily constitute large deviations from Standard English. Focus fronting, for example, was only characterised as non-standard English by one third of the Welsh participants, while almost 90 per cent of the Londoners considered it a deviation from the standard. As to why some features are perceived as distinctly more non-standard than others, several structural-linguistic factors can be taken into consideration. Features may stand out, for example, because they are placed in an interactionally prominent position in a sentence or clause (cf. Cheshire 1996), because they are positioned at the beginning or end of a linguistic unit (cf. Auer 2014) and/or because they involve content rather than function words (cf. Auer 2014). A reason behind the low salience of some features may be their similarity (formal or

functional) to other structures in the standard language, with a difference too subtle for non-linguists to notice.

Both intra- and interlectal perceptions, so the present study revealed, can be influenced by personal, social and attitudinal parameters among listeners. For example, higher recognition rates of dialectal features correlated with more positive stances towards dialectal diversity. For outsiders specifically, the intensity of contact with Welsh English, for example, in the form of having friends from Wales, seems to have positive effects on the recognition of dialectal Welsh English features. All in all, these results underline the subjective character of salience and stress that individual salience perceptions (cf. Purschke 2014: 31) can differ considerably from the average tendencies as predicted by From-Corpus-to-Cognition approaches.

7.2 General gains and implications for future research

7.2.1 Usage-based linguistics

The present study was conducted in the context of the research training group (“Graduiertenkolleg”) GRK DFG 1624 *Frequency effects in language* at the University of Freiburg. Funded from 2009 to 2018, the group aims “to carry out empirically rich and methodologically co-ordinated research on frequency effects in language, with an empirical focus on standard and non-standard varieties of European languages” (GRK Frequenz 2016). As stated in previous chapters, a central claim is usage-based linguistics is that “linguistic structures emanate from usage events” (Behrens and Pfänder 2016: 3). Humans automatically extract frequency information from their surroundings and these frequencies affect performance (cf. Divjak and Caldwell-Harris 2015: 53). Stefan Pfänder and Heike Behrens, the speakers of the GRK DFG 1624, discuss central fields of research in the context of frequency effects in language in their introduction to the volume *Experience counts: Frequency effects in language* (2016). The following sections illustrate how my study contributes to two of those fields, thus highlighting the relevance of this work for the discussion of frequency effects.

Relation of frequency and salience

Behrens and Pfänder (2016: 8) address salience in the context of frequency in interaction with other processing factors, and state that morphosyntactic salience has been studied mainly in the areas of perceptual dialectology, dialect contact research and sociolinguistics. Most of the salience studies in these fields, however, have focused on phonetic/phonological variation, thus “features of dialect grammar are largely excluded” (Behrens and Pfänder 2016: 8).¹ Moreover, with the exception of Rácz (2013), “barely any systematic work has been done on either the relevance of (high or low) frequency in explaining perceptual salience (or non-salience) or the exact interaction of frequency and salience in various scenarios” (Behrens and Pfänder 2016: 8). The present study approaches these limitations by providing a systematic account of the relevance of frequency for explaining sociolinguistic salience in morphosyntax. Quantitative analyses were combined with studies of individual particularities of features (e.g. structural) and interindividual differences between speakers (e.g. age) affecting perception, in concert with Behrens and Pfänder (2016: 2–3): “The investigation of frequency effects requires a very fine-grained analysis of the usage-conditions of linguistic structures, as well as the application of sophisticated statistical methods.” In the following, central findings on the usage-based character of sociolinguistic salience are summarised:

- (Comparatively) high token frequency in the target variety plays a role but is no categorical requirement for salience in dialect grammar. The high salience of one feature in the present study, focus fronting, goes hand in hand with high usage rates in Welsh English. The example of the considerably less frequent but equally salient invariant *isn't it*, however, shows: features with different token frequencies can be entrenched to similar degrees. While it has been established that, generally, “[r]epeated encounter leads to entrenchment, the strengthening of memory traces” (Behrens and Pfänder 2016: 4), my work finds that for salient dialectal grammatical features the “much-quoted frequency effect on routinization and entrenchment” (Schmid 2014: 245) does not always hold. Other factors, such as perceived structural prominence, may compensate for low token frequencies and help strengthen an item’s mental representation.

¹Exceptions are e.g. Cheshire (1996) and Kerswill and Williams (2000, 2002), as discussed in previous chapters.

- The findings for insiders and outsiders provide evidence for the causality of frequency effects in dialect perception. As Divjak and Caldwell-Harris (2015: 57) state, “[t]o determine how increased usage itself may be responsible for frequency effects, researchers have tried to identify people who could reasonably be expected to have different usage histories”. For example, Caldwell-Harris, Berant and Edelman (2012) found that religious phrases were processed faster by religious Jews than secular Jews and that secular Jews showed weaker frequency effects. The authors conclude that

[f]requency effects that vary according to individual differences in language exposure are a natural outcome of usage-based theories of language and are thus a promising arena for testing specific predictions about how usage influences entrenchment (Caldwell-Harris, Berant and Edelman 2012: 165).

My study reveals that Welsh people’s perceptions of Welsh English grammar conform better with token frequency in Welsh English than Londoners’ judgements. In other words, the frequency effect (token frequency in Welsh English) is stronger for insiders than outsiders, in line with the findings by Caldwell-Harris, Berant and Edelman (2012: 165). The results indicate that “at least part of the frequency effect [in Welsh English] is due to language users’ actual experience with those words and phrases” (Divjak and Caldwell-Harris 2015: 57).

- The present study finds frequency effects among low-frequency constructions. As stated in Chapter 6, all features investigated are relatively infrequent in English compared to other grammatical structures such as the regular simple past tense form (*-ed*). And they are probably much less frequent than a range of salient phonetic, phonological and lexical items. Divjak and Caldwell-Harris (2015: 56) state that “[f]requency effects [...] have been attested for items across the low to high frequency range although less research exists on the former”. This work thus complements findings by Rácz (2013) on much more frequent phonetic/phonological constructions: both in his *and* in my study, frequency differences between the target dialect and the dialect of the outsiders appear to affect interlectal salience.
- My analyses, however, also point to significant differences between salience in phonology and morphosyntax. The grammatical features analysed seem to be

less overtly stigmatised and stereotyped in the Labovian sense than salient phonological features discussed in the literature (cf. e.g. Johnstone and Kiesling 2008; Rácz 2013). Such stigmatisation processes can result in reduced usage frequencies, with salient variants being avoided. The forms may become “increasingly divorced from forms that are actually used” (Johnstone and Kiesling 2008: 9). My corpus data do not provide any evidence for such a salience-induced reduction of frequency. The two most salient features are used continuously in Welsh English. Focus fronting is significantly more frequent than the other constructions investigated. Comparing the frequencies of focus fronting and the invariant *isn't it*, the frequency ratio has not changed between the 1970s (FRED) and the late 1990s/early 2000s (Radio Wales Corpus). In other words, *isn't it* was already considerably less frequent than focus fronting some decades ago. The media analysis, moreover, found similar frequency ratios as in the corpora, providing no evidence for features being stereotypically (and more frequently) used in the media than in real life.

Language change

London English was not in the focus of this project; nevertheless, the usage frequencies and perception data of the invariant tag question *innit* yield some interesting insights into its process of becoming established as an individual feature. Whether this process is an instance of lexicalisation or grammaticalisation is a matter of dispute among linguists. As argued by Brinton and Traugott (2005: 62), “[f]usion of syntagmatically free items into fixed phrases and sometimes further reduction by coalescence is typical both of certain types of lexicalization [...] and of grammaticalization”. Both processes go along with a reduction of compositionality regarding form (fusion) and meaning (idiomaticisation). An argument for *innit* as an instance of grammaticalisation would be that the meaning of grammaticalised items becomes bleached and that specific components of meaning get lost (cf. Brinton and Traugott 2005: 68). While the concord tag question *isn't it* refers to the subject and verb of the main clause, this reference is lost for the invariant *innit*, as it can be attached as a discourse marker to basically any clause. Krug (1998), by contrast, argues that the invariant *innit* is a product of lexicalisation as the former syntactic construction has shed syntactic functions and has become a lexico-grammatical pattern. Along these lines, the construction grammar approach by Trousdale (2008) conceptualises lexicalisation as a process whereby

constructions become less productive (or schematic), so that fewer items can fill constructional slots. The invariant *innit* seems to be developing into a less general and more substantive construction.

Whether it is lexicalisation or grammaticalisation is not decisive for the present discussion, as frequency can affect both processes in similar ways. For example, (1) high frequency can trigger the emergence of new constructions, and (2) help establish the new constructions as autonomous forms. As for (1), if independent items co-occur frequently, their morphological boundaries may become blurred and merge eventually, resulting in lexicalisation (cf. Krug 1998: 187) or grammaticalisation (cf. Behrens and Pfänder 2016: 5). The corpus-based approach by Krug (1998) provides evidence for this assumption, finding that *isn't it/innit* is the “most frequent tag of all operator-subject combinations in the English paradigm”. Moreover, the most frequent contraction in operators of negated tags in this study is *isn't*. Krug's frequency-based analyses also illustrate how far *innit* has spread in the British Isles, with regard to age groups, regions and social backgrounds. In sum, his work accounts for the emergence of *innit* as a new discourse marker.

The present study provides usage-based evidence for the further advance of the invariant *innit* as an autonomous item. Apparently, the feature is not simply perceived as a synonym for the invariant *isn't it*, but has developed individual characteristics. For example, *innit* and *isn't it* carry different regional signals. The fact that invariant *isn't it* and *innit* constitute two distinct features (a more “Welsh” one and a more “London” one) in people's minds can be explained based on token frequencies: while *isn't it* is considerably more common in Welsh English tag questions, non-standard and standard ones, *innit* is the dominant form in London English.

7.2.2 Perceptual dialectology

Another central goal of this work was to offer new insights for the fields of perceptual dialectology and language attitudes research. Whilst previous studies into laymen's dialect perceptions have mainly focused on phonetic-phonological and lexical features, I have here presented a systematic large-scale analysis of folk linguistic perceptions of grammatical variation in the British Isles. The study reveals that there are significant differences in the degrees to which Welsh English grammatical fea-

tures are associated with Wales, by insiders and outsiders. Note that the most salient grammatical features are probably still less salient than prominent (and more frequent) phonetic-phonological items, like the long monophthongs [e:] and [o:] – considered “strong stereotypes of Welshness in English” by Coupland (2001: 354). But the present study shows: social indexation is at work in dialect grammar. It thus complements the works by Coupland (2001) and Williams, Garrett and Coupland (1996) by demonstrating that laymen’s mental conceptions of Welsh English are not only shaped by the much-discussed phonological and lexical items, but also by morphosyntactic ones.

The findings on the distinctness of invariant *isn’t it* and invariant *innit* underline the importance of folk linguistic judgements for understanding developments in dialects. Based on the research literature, my salience questionnaire did not further distinguish between the two features, simply providing informants with written <isn’t it>. The salience ratings from the subjects revealed that their perceptions of the feature/s are more fine-grained, and that two phonologically different constructions need to be distinguished here. Unearthing the areal distributions of linguistic forms, a central goal of dialectologists, can thus profit from non-linguists’ perceptions of linguistic variation in geographical space (cf. Montgomery and Cramer 2016; Preston 1989).

In the context of language attitudes research, my results substantiate Garrett’s (2010) finding that people’s stances towards linguistic diversity are a crucial factor in dialect judgements. Whilst Garrett (2010) draws on data from an attitudes survey about (both British Isles and non-native) accents of English, the present work further shows that openness to linguistic diversity also plays an important role in the recognition of dialectal grammatical features.

7.2.3 Welsh English

Welsh English served as an object of study in this work. Since this variety of English is so far understudied, my third central goal was to contribute to research on English in Wales and thus to provide findings of interest to variationist sociolinguists and dialectologists. The corpus analyses to determine frequencies of grammatical features in Welsh English exhibit some parallels to works by Paulasto and colleagues (cf. Meriläinen and Paulasto 2014; Paulasto 2006). For three features they explore (focus fronting, non-standard habitual progressives and the inverted word order in indirect ques-

tions) the present approach offers a frequency-based support. Additionally, I provide new frequency data for three other grammatical features not analysed quantitatively in the context of Welsh English by Meriläinen and Paulasto (2014) and Paulasto (2006) or other recent works. For example, the constructions are not included in the catalogue of 57 features presented by Szmrecsanyi (2013), which he uses for his dialectometric FRED-based study of morphosyntactic variation in British English dialects. The results of the present study thus complement Szmrecsanyi's findings with regard to the pervasiveness of non-standard grammatical features in Wales and London.

Furthermore, by contrast to Paulasto's corpora and FRED, the interviews in the Radio Wales Corpus come from various locations all over Wales. The RWC is probably the first corpus since SAWD from the 1970s to cover Wales as a whole. The corpus provides many opportunities for investigating grammatical, lexical and, to a certain extent, also phonological variation in Welsh English (the BBC Voices recordings are freely available online, at <http://sounds.b1.uk/Accents-and-dialects/BBC-Voices>). The region-specific analyses with the RWC in this work constitute a starting point for such analyses, providing fresh data on grammatical variation in Wales.

The study of the tag questions *isn't it* and *innit* also contributes to dialectological research. As stated in previous sections, *isn't it* appears to be strongly associated with Welsh English while *innit* is a salient marker of London English. It could thus be useful for future accounts and classifications of grammatical variation in British Isles dialects of English – as in the format of eWAVE or the Survey of Anglo-Welsh dialects (SAWD) – to distinguish more clearly between the invariant tag questions *isn't it* and *innit*.

Finally, the systematic investigation of the relation between speech data and perception data is novel in research on English in Wales. While corpus hits can indicate which features are *present* in Welsh English today, perception data can reveal which of these features are *relevant* to speakers and seen as a vital part of their own dialect.

7.2.4 Outlook: Towards a usage-based dialectology

Not only does the present study contribute to each of those different linguistic disciplines, but it also forges a bridge between them, and thus helps to consolidate the emerging field of *cognitive/usage-based dialectology* (cf. Kristiansen 2006; Szelid and

Geeraerts 2008). According to Szelid and Geeraerts (2008: 23), “there is no standing tradition of dialectological research in Cognitive Linguistics. That is to be regretted, because the inspiration could well be mutual.” As illustrated in the present study, dialectology can profit from usage-based models in that the latter help explain perceptions of language-internal morphosyntactic variation by means of a cognitive, experience-based approach (cf. also Kortmann 2010: 840). Szelid and Geeraerts (2008: 24) additionally note that “the usage-based nature of Cognitive Linguistics challenges the traditional methodological focus of dialectology on language structure rather than language use.” *Natural* language use and perception data of non-standard dialects (as provided by the present corpus studies and perception tests), on the other hand, can support cognitive linguistic approaches in modelling the mental organisation and representation of language. According to Kristiansen (2006: 109), natural language has to form an essential foundation of usage-based linguistics: “we can only take the claim that Cognitive Linguistics is a usage-based approach seriously if the kind of language that we analyze is *real* language, language as it is actually used by real speakers in real situations” (emphasis in original).

Overall, I hope that this work lays the foundations for many studies to come exploring *real* language in terms of dialect perception, morphosyntactic variation and mental (usage-based) representation. Such studies may build upon the central conclusion from the present work that “experience counts” (Behrens and Pfänder 2016: 1), but that this is only one of several factors determining which linguistic features count to the individual.

Bibliography

- Aitchison, John and Harold Carter (1994). *A geography of the Welsh language 1961–1991*. Cardiff: University of Wales Press.
- Altendorf, Ulrike and Dominic Watt (2008). “The dialects in the South of England: Phonology”. In: *Varieties of English 1: The British Isles*. Ed. by Bernd Kortmann and Clive Upton. Berlin and New York: Mouton de Gruyter, 194–222.
- Anderwald, Lieselotte (2008). “The varieties of English spoken in the Southeast of England: Morphology and syntax”. In: *Varieties of English 1: The British Isles*. Ed. by Bernd Kortmann and Clive Upton. Berlin and New York: Mouton de Gruyter, 440–462.
- Anderwald, Lieselotte and Susanne Wagner (2006). “FRED – The Freiburg English Dialect Corpus: Applying corpus-linguistic research tools to the analysis of dialect data”. In: *Using unconventional digital language data. Vol. 1: Synchronic corpora*. Ed. by Joan C. Beal, Karen P. Corrigan and Herrmann L. Moisl. Basingstoke: Palgrave Macmillan, 35–53.
- Androutsopoulos, Jannis (2014). “Beyond ‘media influence’”. In: *Journal of Sociolinguistics* 18(2), 242–249.
- Audiotranskription (2015). *Transcribe faster with f5 (Mac)*. Url: [http : / / www . audiotranskription . de / english / f 5](http://www.audiotranskription.de/english/f5) (5 May 2015).
- Auer, Peter (2014). “Anmerkungen zum Salienzbegriff in der Soziolinguistik”. In: *Linguistik Online* 66(4), 7–20.
- Auer, Peter, Birgit Barden and Beate Grosskopf (1998). “Subjective and objective parameters determining ‘salience’ in long-term dialect accommodation”. In: *Journal of Sociolinguistics* 2(2), 163–187.
- Awbery, Gwenllian M. (1997). “The English language in Wales”. In: *The Celtic Englishes I*. Ed. by Hildegard L. C. Tristram. Heidelberg: Winter, 86–99.

- Awh, Edward, Artem V. Belopolsky and Jan Theeuwes (2012). "Top-down versus bottom-up attentional control: A failed theoretical dichotomy". In: *Trends in Cognitive Sciences* 16(8), 437–443.
- Baayen, R. Harald (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
- Bardovi-Harlig, Kathleen (1987). "Markedness and salience in second-language acquisition". In: *Language Learning* 37(3), 385–407.
- Baugh, Albert C. and Thomas Cable (2010). *A history of the English language*. 5th ed. Abingdon: Routledge.
- Behrens, Heike and Stefan Pfänder, eds. (2016). *Experience counts: Frequency effects in language*. Berlin: Mouton de Gruyter.
- Biber, Douglas (1993). "Representativeness in corpus design". In: *Literary and Linguistic Computing* 8(4), 243–257.
- (1999). *Longman grammar of spoken and written English*. 1st ed. Harlow: Longman.
- Biber, Douglas, Susan Conrad and Geoffrey Leech (2005). *Longman student grammar of spoken and written English*. 1st ed. 4th imprint. Harlow: Longman.
- Bishop, Hywel (2006). "Wales: Language situation". In: *Encyclopedia of language & linguistics*. Ed. by Edward K. Brown. 2nd ed. Oxford: Elsevier, 507–508.
- Blandford, Steve et al. (2010). *Screening the nation: Wales and landmark television*. Report for the BBC Trust/Audience Council Wales.
- Blumenthal-Dramé, Alice (2012). *Entrenchment in usage-based theories: What corpus data do and do not reveal about the mind*. Berlin and Boston: Mouton de Gruyter.
- Bortz, Jürgen (2005). *Statistik für Human- und Sozialwissenschaftler*. 6th ed. Heidelberg: Springer.
- Brake, Phylip (2004). *An introduction to Welsh*. München: Lincom.
- Brinton, Laurel J. and Elizabeth C. Traugott (2005). *Lexicalization and language change*. Cambridge: Cambridge University Press, 32–61.
- Britain, David (2007). "Grammatical variation in England". In: *Language in the British Isles*. Ed. by David Britain. Cambridge et al.: Cambridge University Press.
- British Library (2015). *BBC Voices*. Url: <http://sounds.bl.uk/Accents-and-dialects/BBC-Voices> (04 May 2015).
- Bybee, Joan (2006). "From usage to grammar: The mind's response to repetition". In: *Language* 82(4), 711–733.

- Caldwell-Harris, Catherine L., Jonathan Berant and Shimon Edelman (2012). "Measuring mental entrenchment of phrases with perceptual identification, familiarity ratings, and corpus frequency statistics". In: *Frequency effects in language representation*. Ed. by Dagmar Divjak and Stefan Th. Gries. Berlin and Boston: De Gruyter Mouton, 165–194.
- Cambridge University Press (2015). *Permissions requests*. Url: [http : / / www . cambridge . org / about - us / rights - permissions / permissions/permissions-requests](http://www.cambridge.org/about-us/rights-permissions/permissions/permissions-requests) (11 March 2015).
- Castelloe, John M. and Ralph G. O'Brien, eds. (2000). *Power and sample size determination for linear models*. Proceedings of the twenty-sixth Annual SAS Users Group International Conference. SAS Institute Inc.
- Cheshire, Jenny (1996). "Syntactic variation and the concept of prominence". In: *Speech past and present: Studies in English dialectology in memory of Ossi Ihalainen*. Ed. by Juhani Klemola, Merja Kytö and Matti Rissanen. Frankfurt am Main: Peter Lang, 1–17.
- Cheshire, Jenny, Viv Edwards and Pam Whittle (1995). "Urban British dialect grammar: The question of dialect levelling". In: *Verbale Kommunikation in der Stadt*. Ed. by Iwar Werlen. Tübingen: Narr, 67–110.
- Cheshire, Jenny et al. (2011). "Contact, the feature pool and the speech community: The emergence of Multicultural London English". In: *Journal of Sociolinguistics* 15(2), 151–196.
- "Cockney? So you're not from London" (2010). In: *Daily Mail*. 7 September 2010, 11.
- Collins, Beverley and Inger M. Mees (1999). "Cardiff: A real-time study of glottalisation". In: *Urban voices: Accent studies in the British Isles*. Ed. by Gerard J. Docherty and Paul Foulkes. London: Hodder Arnold, 185–202.
- Connolly, John H. (1990). "Port Talbot English". In: *English in Wales: Diversity, conflict and change*. Ed. by Nikolas Coupland. Clevedon et al.: Multilingual Matters, 121–129.
- Coupland, Nikolas (2001). "Stylization in radio talk". In: *Language in Society* 30(3), 345–375.
- Coupland, Nikolas and Alan R. Thomas (1990). "Introduction: Social and linguistic perspectives on English in Wales". In: *English in Wales: Diversity, conflict and change*. Ed. by Nikolas Coupland. Clevedon et al.: Multilingual Matters, 1–16.

- Crystal, David (2013). "Voices: A case study in the evolution of a linguistic climate at the BBC". In: *Analysing 21st century British English: Conceptual and methodological aspects of the "Voices" project*. Ed. by Clive Upton and Bethan L. Davies. London et al.: Routledge, 12–25.
- Davies, Janet (2014). *The Welsh language: A history*. Cardiff: University of Wales Press.
- Davies, John (1994). *A history of Wales*. London et al.: Penguin.
- Delort, Jean-Yves (2009). "Automatically characterizing salience using readers' feedback". In: *Journal of Digital Information* 10(1).
- Deumert, Ana (2003). "Markedness and salience in language contact and second-language acquisition: Evidence from a non-canonical contact language". In: *Language Sciences* 25, 561–613.
- Divjak, Dagmar and Catherine L. Caldwell-Harris (2015). "Frequency and entrenchment". In: *Handbook of cognitive linguistics*. Ed. by Ewa Dabrowska and Dagmar Divjak. Berlin and Boston: De Gruyter Mouton, 53–75.
- Dunning, Ted (1993). "Accurate methods for the statistics of surprise and coincidence". In: *Computational Linguistics* 19 (1), 61–74.
- Edwards, John (2009). *Language and identity: An introduction*. Cambridge et al.: Cambridge University Press.
- Edwards, John R. (1985). *Talk Tidy: The art of speaking Wenglish*. 1st ed. Cowbridge: D. Brown & Sons Ltd.
- (1986). *More Talk Tidy*. 1st ed. Cowbridge: D. Brown & Sons Ltd.
- Edwards, Viv (1993). "The grammar of southern British English". In: *Real English: The grammar of English dialects in the British Isles*. Ed. by Jim Milroy and Lesley Milroy. London: Longman, 214–238.
- Ellis, Nick C. (2012). "What can we count in language, and what counts in language acquisition, cognition, and use?" In: *Frequency effects in language learning and processing*. Ed. by Stefan Th. Gries and Dagmar Divjak. Vol. 1. Berlin: Mouton de Gruyter, 7–34.
- Elmentaler, Michael, Joachim Gessinger and Jan Wirrer (2010). "Qualitative und quantitative Verfahren in der Ethnodialektologie am Beispiel von Salienz". In: *Perceptual Dialectology: Neue Wege der Dialektologie*. Ed. by Christina Ada Anders, Markus Hundt and Alexander Lasch. Berlin and New York: Mouton de Gruyter, 111–150.

- Elmes, Simon (2013). "Voices: A unique BBC adventure". In: *Analysing 21st century British English: Conceptual and methodological aspects of the "Voices" project*. Ed. by Clive Upton and Bethan L. Davies. London et al.: Routledge, 1–11.
- Everts, Philip (2011). "Familiarity breeds consent: Issue salience and support for the use of military force". In: *Issue salience in international politics*. Ed. by Kai Oppermann and Henrike Viehrig. Oxon: Routledge.
- Filppula, Markku (1999). *The grammar of Irish English: Language in Hibernian style*. New York et al.: Routledge.
- (2012). "Irish English". In: *The Mouton world atlas of variation in English*. Ed. by Bernd Kortmann and Kerstin Lunkenheimer. Berlin and Boston: Mouton de Gruyter, 30–47.
- Filppula, Markku, Juhani Klemola and Heli Paulasto (2008). *English and Celtic in contact*. New York et al.: Routledge.
- Fischer, Kerstin (2010). "Quantitative methods in cognitive semantics: Introduction to the volume". In: *Quantitative methods in cognitive semantics: Corpus-driven approaches*. Ed. by Dylan Glynn and Kerstin Fischer. Berlin and New York: Mouton de Gruyter, 43–59.
- Franceschini, Fiorenzo, Maurizio Galetto and Marco Varetto (2004). "Qualitative ordinal scales: The concept of ordinal range". In: *Quality Engineering* 16(4), 515–524.
- Gallwey, April (2013). "The rewards of using archived oral histories in research: The case of the Millennium Memory Bank". In: *Oral History* 41(1), 37–50.
- Garrett, Peter (2010). *Attitudes to language*. Cambridge et al.: Cambridge University Press.
- Garrett, Peter, Nikolas Coupland and Angie Williams (1999). "Evaluating dialect in discourse: Teachers' and teenagers' responses to young English speakers in Wales". In: *Language in Society* 28(3), 321–354.
- Garrett, Peter, Angie Williams and Betsy Evans (2005). "Attitudinal data from New Zealand, Australia, the USA and UK about each other's Englishes: Recent changes or consequences of methodologies?" In: *Multilingua* 24, 211–235.
- Giora, Rachel (1998). "When is relevance? On the role of salience in utterance interpretation". In: *Revista Alicantina de Estudios Ingleses* 11(11), 85–94.
- Glynn, Dylan (2010). "Corpus-driven cognitive semantics: Introduction to the field". In: *Quantitative methods in cognitive semantics: Corpus-driven approaches*. Ed. by Dylan Glynn and Kerstin Fischer. Berlin and New York: Mouton de Gruyter, 1–41.

- Günther, Susanne (1997). "Direkte und indirekte Rede in Alltagsgesprächen: Zur Interaktion von Syntax und Prosodie in der Redewiedergabe". In: *Syntax des gesprochenen Deutsch*. Ed. by Peter Schlobinski. Opladen: Westdeutscher Verlag, 227–263.
- Gries, Stefan Th. (2009). *Statistics for linguistics with R: A practical introduction*. Berlin and New York: Mouton de Gruyter.
- (2010). "Useful statistics for corpus linguistics". In: *A mosaic of corpus linguistics: Selected approaches*. Ed. by Aquilino Sánchez and Moisés Almela. Frankfurt am Main: Peter Lang, 269–291.
- Hannahs, S. J. (2013). *The phonology of Welsh*. Oxford: Oxford University Press.
- Harrington, Jonathan (2006). "An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts". In: *Journal of Phonetics* 34, 439–457.
- Harrington, Jonathan, Sallyanne Palethorpe and Catherine I. Watson (2000). "Does the Queen speak the Queen's English?" In: *Nature* 408, 927–928.
- Haspelmath, Martin (2006). "Against markedness (and what to replace it with)". In: *Journal of Linguistics* 42(1), 25–70.
- Herbst, Thomas, Rita Stoll and Rudolf Westermayr (1991). *Terminologie der Sprachbeschreibung: Ein Lernwörterbuch für das Anglistikstudium*. Ismaning: Max Hueber Verlag.
- Hernández, Nuria (2006). *User's guide to FRED: Freiburg Corpus of English Dialects*. Url: <https://www.freidok.uni-freiburg.de/data/2489/> (5 May 2015).
- Hettler, Yvonne (2014). "Salienz, Bewertung und Realisierung regionaler Merkmale in Norddeutschland". In: *Linguistik Online* 66(4), 71–90.
- Hughes, Arthur, Peter Trudgill and Dominic Watt (2012). *English accents and dialects: An introduction to social and regional varieties of English in the British Isles*. 5th ed. Oxon and New York: Routledge.
- Hume, Elizabeth and Frederic Mailhot (2013). "The role of entropy and surprisal in phonologization and language change". In: *Origins of sound change: Approaches to phonologization*. Ed. by Alan C. L. Yu. Oxford: Oxford University Press, 29–47.
- Itti, Laurent and Pierre Baldi (2009). "Bayesian surprise attracts human attention". In: *Vision Research* 49(10), 1295–1306.
- Itti, Laurent and Christof Koch (2000). "A saliency-based search mechanism for overt and covert shifts of visual attention". In: *Vision Research* 40, 1489–1506.

- James, Allan (2011). "English(es) in post-devolution Wales: The sociolinguistic reconstruction of late modern Valleys Voice". In: *AAA – Arbeiten aus Anglistik und Amerikanistik* 36(1), 47–63.
- James, William (1890). *The principles of psychology II*. London: Macmillan.
- Jansen, Sandra (2014). "Salience effects in the North-West of England". In: *Linguistik Online* 66(4), 91–110.
- Johnstone, Barbara, Jennifer Andrus and Andrew E. Danielson (2006). "Mobility, indexicality, and the enregisterment of 'Pittsburghese'". In: *Journal of English Linguistics* 34(2), 77–104.
- Johnstone, Barbara and Scott F. Kiesling (2008). "Indexicality and experience: Exploring the meanings of /aw/-monophthongization in Pittsburgh". In: *Journal of Sociolinguistics* 12 (1), 5–33.
- Kerswill, Paul (1985). *A sociolinguistic study of rural immigrants in Bergen, Norway*. Cambridge University: Unpublished PhD thesis.
- (2014). "The objectification of 'Jafaican': The discursal embedding of Multicultural London English in the British media". In: *Mediatization and sociolinguistic change*. Ed. by Jannis Androutsopoulos. Berlin and Boston: De Gruyter.
- Kerswill, Paul and Ann Williams (2000). "'Salience' as an explanatory factor in language change: Evidence from dialect levelling in urban England". In: *Reading Working Papers in Linguistics* 4, 63–94.
- (2002). "Dialect recognition and speech community focusing in new and old towns in England: The effects of dialect levelling, demography and social networks". In: *Handbook of perceptual dialectology*. Ed. by Daniel Long and Dennis R. Preston. Vol. 2. Amsterdam: Benjamins, 178–207.
- Kerswill, Paul et al. (2007). *Linguistic Innovators: The English of Adolescents in London: Full research report*. ESRC End of Award Report. Swindon.
- Kimchi, Ruth and Mary A. Peterson (2008). "Figure-ground segmentation can occur without attention". In: *Psychological Science* 19(7), 660–668.
- King, Gareth (2015). *Modern Welsh: A comprehensive grammar*. 3rd ed. London and New York: Routledge.
- Kortmann, Bernd (2005). *English linguistics: Essentials*. Berlin: Cornelsen.
- (2010). "Areal variation in syntax". In: *Language and space: Theory and methods*. Ed. by Peter Auer and Jürgen Schmidt. Berlin and New York: Walter de Gruyter, 837–864.

- Kortmann, Bernd (2012). "Regional profile: The British Isles". In: *The Mouton world atlas of variation in English*. Ed. by Bernd Kortmann and Kerstin Lunkenheimer. Berlin and Boston: Mouton de Gruyter, 679–702.
- Kortmann, Bernd and Kerstin Lunkenheimer (2012). "Introduction". In: *The Mouton world atlas of variation in English*. Ed. by Bernd Kortmann and Kerstin Lunkenheimer. Berlin and Boston: Mouton de Gruyter, 1–11.
- (2013). *The electronic world atlas of varieties of English*. Url: <http://ewave-atlas.org/> (2 March 2015).
- Kortmann, Bernd and Christoph Wolk (2012). "Morphosyntactic variation in the anglophone world: A global perspective". In: *The Mouton world atlas of variation in English*. Ed. by Bernd Kortmann and Kerstin Lunkenheimer. Berlin and Boston: Mouton de Gruyter, 907–936.
- Kristiansen, Gitte (2006). "Towards a usage-based cognitive phonology". In: *International Journal of English Studies* 6(2), 107–140.
- Krug, Manfred (1998). "British English is developing a new discourse marker, innit? A study in lexicalisation based on social, regional and stylistic variation". In: *AAA – Arbeiten aus Anglistik und Amerikanistik* 23(2), 145–197.
- Krug, Manfred and Katrin Sell (2013). "Designing and conducting interviews and questionnaires". In: *Research methods in language variation and change*. Ed. by Manfred Krug and Julia Schlüter. Cambridge et al.: Cambridge University Press, 69–98.
- Kuperman, Victor and Julie A. Van Dyke (2013). "Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers". In: *Journal of Experimental Psychology: Human Perception and Performance* 39(3), 802–823.
- Labov, William (1966). *The social stratification of English in New York City*. Washington D.C.: Center for Applied Linguistics.
- (1972a). *Sociolinguistic patterns*. Washington D.C.: University of Pennsylvania Press.
- (1972b). "Some principles of linguistic methodology". In: *Language in Society* 1, 97–120.
- (2001). *Principles of linguistic change. Vol. 2: Social factors*. Oxford: Blackwell.
- (2010). *Principles of linguistic change. Vol. 3: Cognitive and cultural factors*. Chichester: Wiley–Blackwell.
- Labov, William et al. (2011). "Properties of the sociolinguistic monitor". In: *Journal of Sociolinguistics* 15(4), 431–463.

- Laferriere, Martha (1979). "Ethnicity in phonological variation and change". In: *Language* 55(3), 603–617.
- Langacker, Ronald W. (2013). *Essentials of Cognitive Grammar*. Oxford et al.: Oxford University Press.
- Lange, Claudia (2012). *The syntax of spoken Indian English*. Amsterdam: Benjamins.
- Leech, Geoffrey and Jan Svartvik (2013). *A communicative grammar of English*. 3rd ed. Oxon and New York: Routledge.
- Lenz, Alexandra Nicole (2003). *Struktur und Dynamik des Substandards: Eine Studie zum Westmitteldeutschen (Wittlich/Eifel)*. Stuttgart: Steiner.
- (2010). "Zum Salienzbeff und zum Nachweis salienter Merkmale". In: *Perceptual Dialectology: Neue Wege der Dialektologie*. Ed. by Christina Ada Anders, Markus Hundt and Alexander Lasch. Berlin and New York: Mouton de Gruyter, 89–110.
- Lorenz, Cornelia (2014). "Salienz unter Einheimischen und Zugezogenen: Ein empirischer Vergleich". In: *Linguistik Online* 66(4), 135–146.
- Lupyan, Gary (2010). "Attention: Cognitive influences". In: *Encyclopedia of attention*. Ed. by E. Bruce Goldstein. Thousand Oaks et al.: SAGE Publications, 71–74.
- Mac Mathúna, Liam (1990). "Contrastive aspects of English and Irish". In: *Homage to Ireland: Aspects of culture, literature and language*. Ed. by Birgit Bramsbäck. Uppsala: Almqvist & Wiksell, 81–98.
- Mahboob, Ahmar (2012). "Pakistani English". In: *The Mouton world atlas of variation in English*. Ed. by Bernd Kortmann and Kerstin Lunkenheimer. Berlin and Boston: Mouton de Gruyter, 531–539.
- Mair, Christian (2008). *English linguistics: An introduction*. 1st ed. Tübingen: Narr.
- McArthur, Tom, ed. (1998). *Concise Oxford Companion to the English language*. Oxford: Oxford University Press.
- Meriläinen, Lea and Heli Paulasto (2014). "Embedded inversion as an angloversal: Evidence from inner, outer, and expanding circle Englishes". In: *The Oxford Handbook of World Englishes*. Ed. by Markku Filppula, Juhani Klemola and Devyani Sharma. Oxford: Oxford University Press.
- Mesthrie, Rajend (2008). "Synopsis: Morphological and syntactic variation in Africa and South and Southeast Asia". In: *Varieties of English 4: Africa, South and Southeast Asia*. Ed. by Rajend Mesthrie. Berlin and New York: Mouton de Gruyter, 624–635.
- Meyer, Charles F. (2002). *English corpus linguistics: An introduction*. Cambridge: Cambridge University Press.

- Mihm, Arend (1985). "Prestige und Stigma des Substandards: Zur Bewertung des Ruhrdeutschen im Ruhrgebiet". In: *Sprache an Rhein und Ruhr: Dialektologische und soziolinguistische Studien zur sprachlichen Situation im Rhein-Ruhr-Gebiet und ihrer Geschichte*. Ed. by Arend Mihm. Stuttgart: Franz Steiner Verlag, 163–193.
- Miller, Jim (2008). "Scottish English: Morphology and syntax". In: *Varieties of English 1: The British Isles*. Ed. by Bernd Kortmann and Clive Upton. Berlin and New York: Mouton de Gruyter, 299–327.
- Montgomery, Chris and Jennifer Cramer (2016). "Developing methods in perceptual dialectology". In: *Cityscapes and perceptual dialectology: Global perspectives on non-linguists' knowledge of the dialect landscape*. Berlin and Boston: De Gruyter Mouton, 9–24.
- Oral History Society (2015). *Welcome*. Url: <http://www.ohs.org.uk/index.php> (5 May 2015).
- Oxford English Dictionary Online (2014). Oxford University Press. Url: <http://www.oed.com/> (10 September 2014).
- Parry, David, ed. (1999). *A grammar and glossary of the conservative Anglo-Welsh dialects of rural Wales*. Sheffield: The National Centre for English Cultural Tradition.
- Pattabhiraman, T. and Nick Cercone (1990). "Selection: Salience, relevance and the coupling between domain-level tasks and text planning". In: *Proceedings of the Fifth International Workshop on Natural Language Generation*, 79–86.
- Paulasto, Heli (2006). *Welsh English syntax: Contact and variation*. Joensuu: Joensuu University Press.
- Penhallurick, Robert (1996). "The grammar of Northern Welsh English: Progressive verb phrases". In: *Speech past and present: Studies in English dialectology in memory of Ossi Ihalainen*. Ed. by Juhani Klemola, Merja Kytö and Matti Rissanen. Frankfurt am Main: Peter Lang, 308–342.
- (2007). "English in Wales". In: *Language in the British Isles*. Ed. by David Britain. Cambridge et al.: Cambridge University Press, 152–170.
- (2008a). "Welsh English: Morphology and syntax". In: *Varieties of English 1: The British Isles*. Ed. by Bernd Kortmann and Clive Upton. Berlin and New York: Mouton de Gruyter, 360–372.
- (2008b). "Welsh English: Phonology". In: *Varieties of English 1: The British Isles*. Ed. by Bernd Kortmann and Clive Upton. Berlin and New York: Mouton de Gruyter, 105–121.

- (2012). “Welsh English”. In: *The Mouton world atlas of variation in English*. Ed. by Bernd Kortmann and Kerstin Lunkenheimer. Berlin and Boston: Mouton de Gruyter, 58–69.
- Perks, Rob (2001). “The Century Speaks: A public history partnership”. In: *Oral History* 29(2), 95–105.
- Perks, Rob and Jonnie Robinson (2005). “‘The Way We Speak’: Web-based representations of changing communities in England”. In: *Oral History* 33(2), 79–90.
- Pitkänen, Heli (2003). “Non-standard uses of the progressive form in Welsh English: An apparent time study”. In: *The Celtic Englishes III*. Ed. by Hildegard L. C. Tristram. Heidelberg: Winter, 111–128.
- Podhovnik, Edith (2008). *The phonology of Neath English: A socio-dialectological survey*. Unpublished PhD thesis. University of Wales, Swansea.
- Preston, Dennis R. (1986). “Five visions of America”. In: *Language in Society* 15(2), 221–240.
- (1989). *Perceptual dialectology: Nonlinguists’ views of areal linguistics*. Dordrecht-Providence: Foris.
- (1996). “Whaddayaknow?: The modes of folk linguistic awareness”. In: *Language Awareness* 5(1), 40–74.
- Preston, Dennis R. and Nancy A. Niedzielski (2000). *Folk linguistics*. Berlin and New York: Mouton de Gruyter.
- Purschke, Christoph (2014). “‘I remember it like it was interesting.’ Zur Theorie von Salienz und Pertinenz”. In: *Linguistik Online* 66(4), 31–50.
- Quirk, Randolph et al. (1985). *A comprehensive grammar of the English language*. London et al.: Longman.
- Rácz, Péter (2013). *Salience in sociolinguistics: A quantitative approach*. Berlin and Boston: Mouton de Gruyter.
- Rayson, Paul, Damon Berridge and Brian Francis (2004). “Extending the Cochran rule for the comparison of word frequencies between corpora”. In: *JADT 2004 : 7es Journées internationales d’Analyse statistique des Données Textuelles*.
- Research training group GRK DFG 1624 “Frequency effects in language” (2016). Url: <http://frequenz.uni-freiburg.de/abstract&language=en> (28 October 2016).
- Robinson, Jonnie, Jon Herring and Holly Gilbert (2009–2012). *British Library: Voices of the UK Project*. Unpublished manuscripts.

- Roller, Katja (2011). *The English of young bilinguals in South Wales: A study into phonetics and phonology*. Unpublished thesis (Zulassungsarbeit zum Ersten Staatsexamen). University of Bamberg.
- (2015). “Towards the ‘oral’ in oral history: Using historical narratives in linguistics”. In: *Oral History* 43(1), 73–84.
- Rosen, Anna (2012). “Channel Island English”. In: *The Mouton world atlas of variation in English*. Ed. by Bernd Kortmann and Kerstin Lunkenheimer. Berlin and Boston: Mouton de Gruyter, 98–105.
- (2014). *Grammatical variation and change in Jersey English*. Amsterdam et al.: Benjamins.
- Rosewarne, David (1984). “Estuary English”. In: *Times Educational Supplement*. 19 October 1984, 29.
- Rydén, Mats (1991). “The be/have variation with intransitives in its crucial phases”. In: *Historical English syntax*. Ed. by Dieter Kastovsky. Berlin: Mouton de Gruyter, 343–354.
- Samuel, Martin (2012). “Rhyming slang? Leave it out, John”. In: *Daily Mail*. 30 March 2012, 18.
- Schirmunski, Victor (1928, 1929). “Die schwäbischen Mundarten in Transkaukasien und Südukraine”. In: *Teuthonista* 5(1), 38–60, 157–171.
- (1930). “Sprachgeschichte und Siedelungsmundarten”. In: *Germanisch-Romanische Monatsschrift* 18, 113–122, 171–188.
- Schmid, Hans-Jörg (2000). *English abstract nouns as conceptual shells: From corpus to cognition*. Berlin and New York: Walter de Gruyter.
- (2007). “The notions of entrenchment and salience in cognitive linguistics”. In: *The Oxford handbook of cognitive linguistics*. Ed. by Dirk Geeraerts and Hubert Cuyckens. Oxford et al.: Oxford University Press.
- (2014). “Lexico-grammatical patterns, pragmatic associations and discourse frequency”. In: *Constructions collocations patterns*. Ed. by Thomas Herbst, Hans-Jörg Schmid and Susen Faulhaber. Berlin and Boston: Mouton de Gruyter, 239–294.
- Schütze, Carson T. (1996). *The empirical base of linguistics: Grammaticality judgements and linguistic methodology*. Chicago and London: The University of Chicago Press.
- Smith, Jennifer (2012). “Scottish English and varieties of Scots”. In: *The Mouton world atlas of variation in English*. Ed. by Bernd Kortmann and Kerstin Lunkenheimer. Berlin and Boston: Mouton de Gruyter, 21–29.

- Szelid, Veronika and Dirk Geeraerts (2008). "Usage-based dialectology: Emotion concepts in the Southern Csango dialect". In: *Annual Review of Cognitive Linguistics* 6, 23–49.
- Szmrecsanyi, Benedikt (2010). "The English genitive alternation in a cognitive sociolinguistics perspective". In: *Advances in cognitive sociolinguistics*. Ed. by Dirk Geeraerts, Gitte Kristiansen and Yves Peirsman. Mouton de Gruyter, 141–166.
- (2013). *Grammatical variation in British English dialects: A study in corpus-based dialectometry*. Cambridge: Cambridge University Press.
- Taylor, John R. (2012). *The mental corpus: How language is represented in the mind*. Oxford et al.: Oxford University Press.
- Taylor, Shelley E. and Susan T. Fiske (1978). "Salience, attention and attribution: Top of the head phenomena". In: *Advances in Experimental Social Psychology* 11, 249–288.
- Tench, Paul (1990). "The pronunciation of English in Abercrave". In: *English in Wales: Diversity, conflict and change*. Ed. by Nikolas Coupland. Clevedon et al.: Multilingual Matters, 130–141.
- The R Foundation (2015). *The R project for statistical computing*. Url: <https://www.r-project.org/> (31 August 2015).
- Thomas, Alan R. (1984). "Welsh English". In: *Language in the British Isles*. Ed. by Peter Trudgill. 1st ed. Cambridge et al.: Cambridge University Press, 178–194.
- (1997). "The Welshness of Welsh English: A survey paper". In: *The Celtic Englishes I*. Ed. by Hildegard L. C. Tristram. Heidelberg: Winter, 55–85.
- Timberlake, Alan (1977). "Reanalysis and actualization in syntactic change". In: *Mechanisms of syntactic change*. Ed. by Charles N. Li. University of Texas Press, 141–177.
- Tomasello, Michael (2005). *Constructing a language: A usage-based theory of language acquisition*. 1st paperback ed. Cambridge, Massachusetts and London: Harvard University Press.
- Tottie, Gunnel and Sebastian Hoffmann (2006). "Tag questions in British and American English". In: *Journal of English Linguistics* 34(4), 283–311.
- Trousdale, Graeme (2008). "A constructional approach to lexicalization processes in the history of English: Evidence from possessive constructions". In: *Word Structure* 1(2), 156–177.
- Trudgill, Peter (1986). *Dialects in contact*. 1st ed. Oxford and New York: Basil Blackwell.
- (1999). *The dialects of England*. 2nd ed. Oxford: Blackwell.

- Trudgill, Peter (2014). "Diffusion, drift, and the irrelevance of media influence". In: *Journal of Sociolinguistics* 18(2), 214–222.
- Trudgill, Peter and Jean Hannah (2002). *International English: A guide to the varieties of Standard English*. 4th ed. London et al.: Arnold.
- Upton, Clive and Bethan L. Davies (2013). "Preface". In: *Analysing 21st century British English: Conceptual and methodological aspects of the "Voices" project*. Ed. by Clive Upton and Bethan L. Davies. London et al.: Routledge, xii.
- Vaan, Michiel de (2008). *Etymological dictionary of Latin and the other Italic languages*. Leiden and Boston: Brill Academic Pub.
- Wells, John C. (1982). *Accents of English 1: An introduction*. Cambridge et al.: Cambridge University Press.
- (1986). *Accents of English 2: The British Isles*. Cambridge et al.: Cambridge University Press.
- (1997). "What is Estuary English?" In: *English Teaching Professional* 3, 46–47.
- Welsh Government (2012). *2011 Census: First results on the Welsh language*. Url: <http://wales.gov.uk/docs/statistics/2012/121211sb1182012en.pdf> (11 January 2015).
- Wenning, Carl J. (2014). *How many data points are enough?* In: *Student laboratory handbook*. Ed. by Carl J. Wenning. Url: <http://www2.phy.ilstu.edu/~wenning/slh/> (01 September 2015).
- Williams, Angie, Peter Garrett and Nikolas Coupland (1996). "Perceptual dialectology, folklinguistics, and regional stereotypes: Teachers' perceptions of variation in Welsh English". In: *Multilingua* 15(2), 171–199.
- (1999). "Dialect recognition". In: *Handbook of perceptual dialectology 1*. Ed. by Dennis R. Preston. Amsterdam: Benjamins, 369–383.
- Williams, Colin H. (1990). "The anglicisation of Wales". In: *English in Wales: Diversity, conflict and change*. Ed. by Nikolas Coupland. Clevedon: Multilingual Matters, 19–47.
- (2014). "The lightening veil: Language revitalization in Wales". In: *Review of Research in Education* 38, 242–272.
- Williams, Malcolm (2000). "The pragmatics of predicate fronting in Welsh English". In: *The Celtic Englishes II*. Ed. by Hildegard L. C. Tristram. Heidelberg: Winter, 210–230.

- (2003). “Information packaging in Rhondda speech: A second look at the research of Ceri George”. In: *The Celtic Englishes III*. Ed. by Hildegard L. C. Tristram. Heidelberg: Winter, 201–223.
- Wright, Peter (1981). *Cockney dialect and slang*. London: Batsford.
- Zhang, Tong, Mark Hasegawa-Johnson and Stephen E. Levinson (2006). “Extraction of pragmatic and semantic salience from spontaneous spoken English”. In: *Speech Communication* 48, 437–462.

Appendix

Appendix A: Salience questionnaire – informants from London

Questionnaire: English in the British Isles

Where do you think the speakers are from? (You can tick more than one box if you think that there are different possible locations where the speaker might be from.)

1. *The cat wants petted.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

2. *Who-all did you say was going to be there?*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

3. *There's nice to see you.* (meaning: *How nice to see you.*)

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

4. *He do be sick a lot.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

5. *Sort of a student he was.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

6. *I's going to town tomorrow.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

7. *It's an exhausting job. I'm happy to do it but.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

8. *I'm going to the cinema every week.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

9. *It is so nice today because a sun is shining.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

10. *It's just the way I do speak.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

11. *It is too difficult.* (meaning: *It is very difficult.*)

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

12. *I tell you what we might should do.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

13. *She's after selling the boat.* (meaning: *She has just sold the boat.*)

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

14. *Poor people starved due to the hunger.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

15. *Do you have tickets? No, sold already. (meaning: No, I sold them already.)*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

16. *I'm wondering is it Frank or not.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

17. *Thass rainen so much today!*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

18. *I'm here, amn't I?*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

19. *He gon build my house.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

20. *It was very cold last night, so I walk home very quickly.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

21. *They is something bad wrong with her.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

22. *I wanted that I should get leave.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

23. *The boys was interested, but Mary weren't.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

24. *She lives just across the hallway, isn't it?*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

25. *She got her a new car.*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

26. *This mustn't be true. (meaning: This can't be true.)*

England ☐ N ☐ S ☐ Ireland ☐ Scotland ☐ Wales ☐ Other _____ ☐ Don't know

Welsh English

Look at the following pairs of sentences – one sentence each was said by a Welsh English speaker. Which sentence do you think it is? (You can tick only **one** box per pair.)

- | | |
|--|--|
| 1. <i>When I came home last weekend, my mother cook me a delicious dinner.</i> | <input type="checkbox"/> Welsh English |
| <i>Douglas I think his name was.</i> | <input type="checkbox"/> Welsh English |
| 2. <i>I am playing the piano every Monday.</i> | <input type="checkbox"/> Welsh English |
| <i>I want that I should get a higher salary.</i> | <input type="checkbox"/> Welsh English |
| 3. <i>You like him, isn't it?</i> | <input type="checkbox"/> Welsh English |
| <i>He asked me had I been on holiday already.</i> | <input type="checkbox"/> Welsh English |
| 4. <i>I do meet her once a week.</i> | <input type="checkbox"/> Welsh English |
| <i>There's strange it was. (meaning: How strange it was.)</i> | <input type="checkbox"/> Welsh English |

Personal Information

Initials: _____

Sex: ☐ Male ☐ Female

Date of birth: _____

Ethnic self-identification: _____

Native language(s):
☐ English ☐ Other: _____

Language(s) used at home
 while growing up:
☐ English ☐ Other: _____

Native language(s) father:
☐ English ☐ Other: _____

Native language(s) mother:
☐ English ☐ Other: _____

Highest educational qualification:

- | | |
|---|--|
| <input type="checkbox"/> GCSE's/ O-Levels | <input type="checkbox"/> Bachelor's degree |
| <input type="checkbox"/> A-Levels | <input type="checkbox"/> Master's degree |
| <input type="checkbox"/> Apprenticeship | <input type="checkbox"/> PhD |
| <input type="checkbox"/> None | <input type="checkbox"/> Other: _____ |

Places lived at (at different ages).
 Please indicate the cities/ towns/
 villages:

0-10: _____

11-20: _____

21-30: _____

31-40: _____

41-50: _____

51-60: _____

61-70: _____

71-100: _____

Please indicate your opinion on the following statement using the scale:
 "I like hearing a range of dialects."

Strongly agree

Strongly disagree

☐-----☐-----☐-----☐-----☐-----☐-----☐

Do you have friends who speak a dialect of English other than your own?
 If yes, which dialects are these?

☐ Yes ☐ No

- | | |
|---|--|
| <input type="checkbox"/> Channel Islands English | <input type="checkbox"/> Irish English |
| <input type="checkbox"/> East Anglian English | <input type="checkbox"/> Manx English |
| <input type="checkbox"/> English Dialects in the North of England | <input type="checkbox"/> Orkney and Shetland English |
| <input type="checkbox"/> English dialects in the Southeast of England | <input type="checkbox"/> Scottish English |
| <input type="checkbox"/> English dialects in the Southwest of England | <input type="checkbox"/> Welsh English |
| <input type="checkbox"/> Other: _____ | |

What are the first five things about **Wales** that you can think of?

1. _____
2. _____
3. _____
4. _____
5. _____

What are the first five things about **Welsh English** that you can think of?

1. _____
2. _____
3. _____
4. _____
5. _____

Do you know any radio/ TV programmes from Wales?
 If yes, please name the programmes:

☐ Yes ☐ No

Do you know any radio/ TV presenters or comedians from Wales?
 If yes, please name the presenters/ comedians:

☐ Yes ☐ No

Do you have friends/ family members/ work colleagues from Wales?

☐ Yes ☐ No

Do you have friends/ family members/ work colleagues who speak Welsh English?
 If yes, how often do you speak with them?

☐ Yes ☐ No

☐ Every day ☐ Regularly ☐ Sporadically ☐ Never

Have you ever been to Wales?
 If yes, for how long in total?

☐ Yes ☐ No

☐ Up to a week ☐ Up to a month ☐ Up to 3 months ☐ Up to 6 months ☐ Longer than 6 months

Appendix B: Radio Wales Corpus

B.1 Markup²

Tag	Indicates	Description/examples
<u <i>speaker ID</i> >	Speaker identification	E.g. <u KR> (for a speaker called Katja Roller)
{<u <i>interviewer ID</i> > <i>text</i> }	Interviewer identification, interviewer utterance	E.g. {<u IntAM>: #Hello.} (for an interviewer called Anita Morgan)
#	Start of sentence	E.g. #Yeah, yeah.
(v ‘...’)	Non-verbal element, vocal	E.g. (v ‘laughs’) or (v ‘sniffs’)
(e ‘...’)	Non-verbal element, event (including non-human sounds)	E.g. (e ‘door shuts’) or (e ‘recorder beeps’)
(trunc) ... (/trunc)	Truncation	E.g. Marion and I have (trunc) fri (/trunc), been friends all our life.
(reg sic= ...) ... (/reg)	Regularisation of non-standard pronunciation	E.g. (reg sic=‘em) them (/reg) or (reg sic=‘t) it (/reg) The method facilitates searching for words in standard orthography whilst also providing information about non-standard pronunciation (cf. Hernández 2006: 35)
(overlap)	Overlap of two or more speakers	E.g. <u AMW>: #Oh, that one! (overlap) <?>: #Yeah.
(overlap <u <i>speaker ID</i> > # ...)	Interjection/overlap while someone else continues talking	E.g. <u AMW>: #Oh, I ’m knackered, (overlap <u CMD> #I ’m knackered) they don’t say...
(crosstalk)	Overlap of two or more speakers so that individual utterances are unintelligible	E.g. I wonder (crosstalk)
(unclear) ... (/unclear)	Unclear utterance	E.g. (unclear) word (/unclear) or (unclear) and you (/unclear)

²This markup is largely based on the markup of the Freiburg Corpus of English Dialects (FRED, cf. Hernández 2006: 35).

(gap '...')	Gap in recording	E.g. (gap 'tape interrupted') The tag is also used when parts of recorded conversations are unintelligible, for example, due to bad sound quality: (gap 'indistinct')
-------------	------------------	--

B.2 List of interviews in the Radio Wales Corpus

Interview	Date recording	Number words	Speaker initials	Year birth	Age	Sex
VOICES Bangor	28/01/2005	9,596	RT	1970	34	m
			CP	1967	37	f
			SS	1959	45	f
			CH	1941	63	f
VOICES Bethesda	2005	3,612	DB	1987 or 1988	17	f
			HC	1986 or 1987	18	m
			DG	1987 or 1988	17	m
			HGJ	1986 or 1987	18	m
			HP	1987 or 1988	17	f
			HTP	1986 or 1987	18	m
			CAR	1987 or 1988	17	f
VOICES Bon-y-maen	23/11/2004	6,704	JT	1987 or 1988	17	f
			LA	1961	43	f
			SB	1926	77	f
			EGR	1939	65	m
			JR	1943	61	f
			BS	1931	73	f
VOICES Builth Wells	26/11/2004	13,254	NLS	1986	18	f
			EMGP	1931	73	f
			JCH	1970	34	f
			RAH	1967	36	m
			JCE	1978	26	f
VOICES Flint	09/01/2005	18,791	JH	1989	15	m
			LD	1934 or 1935	70	m
			JDB	1930	74	m
			DLJ	1930 or 1931	74	m
VOICES Glynneath	2005	12,095	VR	1943 or 1944	61	m
			IG	1939 or 1940	65	m
			MG	1940 or 1941	64	f
			CW	1955 or 1956	49	f

			DW	1985 or 1986	19	m
			AT	1951 or 1952	53	f
VOICES Holyhead	15/11/2004	14,043	EGJ	1950	54	m
			JME	1949 or 1950	54	m
			AMG	1942	61	f
			JAW	1950 or 1951	53	f
VOICES Llanelli	23/03/2005	6,026	CL	1986	18	f
			L	1988	16	m
			RP	1969	36	f
			PR	1963	41	m
			ARR	1984	20	m
			GS	1985	19	f
			PS	1984	21	m
			IW	1968	36	m
VOICES Llangollen	28/01/2005	4,980	BC	1952	52	m
			DC	1959	45	m
			PAL	1951	53	f
			ELJ	1961	43	f
			KP	1986	18	f
VOICES Milford Haven	28/01/2005	9,704	WEJJ	1929	75	m
			BJ	1934	70	f
			ET	1935	69	f
			NCE	1954 or 1955	50	f
VOICES Newport	01/03/2005	5,097	RA	1958	46	f
			VC	NA	NA	f
			TJ	1966	39	f
			MS	1962	43	m
			KW	1973	31	f
VOICES Newtown	2005	3,074	GB	1954 or 1955	50	f
			GD	1953 or 1954	51	f
			DJP	1972 or 1973	32	f
			SW	NA	NA	f
			GW	1976 or 1977	28	m
VOICES Pontcanna	2005	8,467	DMA	1965	39	f
			CA	1983	21	f
			TJ	NA	NA	f
			JM	1968	36	f
			SM	1937	67	f
			RO	1986	18	f
VOICES Pontypridd	22/11/2004	10,982	MMG	1938	66	f
			LM	1952	52	f
			KW	1948	56	f

			AJB	1935	69	m
			EB	1938	66	f
			AL	1948	56	f
VOICES Rhos on Sea	28/06/2004	8,026	SW	1988	16	f
			AW	1963	41	f
			NW	1927	77	f
VOICES Rhosgadfan	06/10/2004	4,327	RJ	1980	24	m
			DR	1980	24	m
			MR	1980	24	m
VOICES Risca	14/02/2005	8,154	GP	1943	61	m
			DJ	1940	64	m
			GMDL	1953	51	f
			PG	1950	54	f
			VE	1943	60	f
VOICES Splott	18/01/2005	8,371	CP	1953	51	f
			HP	1984 or 1985	20	f
			DS	1963 or 1964	41	f
			JH	1986 or 1987	18	m
			SH	1938 or 1939	66	m
VOICES Talbot Green	24/01/2005	9,653	RS	1924	80	f
			SS	1926	78	m
			JH	1972	32	m
			PGE	1959 or 1960	45	m
			MW	1970	34	m
VOICES Tregaron	22/03/2005	7,653	CH	1959	45	f
			SH	1932	73	f
			HW	1952	52	m
			MD	1948	56	f
			NAE	1977	27	f
VOICES Treorchy	12/11/2004	11,680	BR	1923	81	f
			DH	1935	68	m
			GH	1938	65	f
			AB	1939	65	f
VOICES Wrexham	09/01/2005	6,788	CMD	1964 or 1965	40	f
			SGD	1961 or 1962	43	m
			CG	1960 or 1961	44	f
			AW	1961 or 1962	43	f
			AMW	1965 or 1966	39	f
MMB Brynamman	25/01/1999	4,586	CJM	1967	31	f
MMB Brynmawr I	26/01/1999	3,887	LJW	1931	67	m

MMB Brynmawr II	18/01/1999	4,167	FAD	1938	60	f
MMB Criccieth	29/04/1999	5,333	DCE	1964	34	f
MMB Gwaun-Cae- Gurwen	10/02/1999	4,704	JH	1970	29	f
MMB Llansamlet	04/11/1998	4,620	KM	1933	65	m
MMB Menai Bridge	27/04/1999	4,481	MAC	1964	35	m
MMB Neath	11/12/1998	3,893	JW	1938	60	m
MMB Neath area	12/01/1999	4,109	JAH	1967	31	f
MMB North Wales	10/11/1999	7,950	ELJ	1926	73	f
			ILJ	1921	78	m
			RR	1933	66	f
			IR	1932	67	m
MMB Pontardawe	16/12/1999	5,020	HE	1970	29	m
MMB Rhondda	15/11/1998	4,229	ED	1931	67	f
MMB St Harmon	18/11/1998	3,695	MED	1935	63	f
MMB Swansea	20/01/1999	4,420	DJ	1973	25	m
MMB Talybont	30/04/1999	3,836	GFH	1933	65	f
MMB Trefor	09/11/1999	4,121	RW	1929	70	m
			MJ	1971 or 1972	27	m
			GJ	NA	NA	m
MMB Treharris	18/11/1998	4,715	GW	1966	32	m
Total		268,843				

Appendix C: List of TV clips analysed³

Link to video	Show	Presenter	Length
https://www.youtube.com/watch?v=CDmkz1jGfOg	Rob Brydon's Annually Retentive	Rob Brydon	00:28:08
https://www.youtube.com/watch?v=ORTOYa-cvRg	Rob Brydon's Annually Retentive	Rob Brydon	00:27:49
http://www.youtube.com/watch?v=Onn_W7fgLeE	The Rob Brydon Show	Rob Brydon	00:29:10
http://www.youtube.com/watch?v=M82qejdUYD4&list=PLWYMemLlKrJ_dRsaj_Rd5iWnyF_fR_vHZ	The Rob Brydon Show	Rob Brydon	00:29:14
http://www.youtube.com/watch?v=mwcRpVixvb8	The Rob Brydon Show	Rob Brydon	00:29:48
http://www.youtube.com/watch?v=Cx0kQ9_Q0mY	The Rob Brydon Show	Rob Brydon	00:13:00
http://www.youtube.com/watch?v=8hlGl8ljFYM	The Rob Brydon Show	Rob Brydon	00:03:39
http://www.youtube.com/watch?v=ZxyGrHsKTmo	The Rob Brydon Show	Rob Brydon	00:04:36
http://www.youtube.com/watch?v=8x56QGeZrIY	The Rob Brydon Show	Rob Brydon	00:07:58
http://www.youtube.com/watch?v=ia1qF8kIGTY	The Rob Brydon Show	Rob Brydon	00:11:30
http://www.youtube.com/watch?v=45_qs9Ks16U	The Rob Brydon Show	Rob Brydon	00:09:19
No longer available online	Rob Brydon Live Standup	Rob Brydon	01:21:10

³ All links were last accessed in October 2015.

https://www.youtube.com/watch?v=RQrSXlMZT3I	The Paul O'Grady Show	Rob Brydon	00:05:07
http://www.youtube.com/watch?v=xu6DAgDtghE	Ask Rhod Gilbert	Rhod Gilbert	00:14:07
http://www.youtube.com/watch?v=56iWq4hGUqc	Never Mind the Buzzcocks	Rhod Gilbert	00:29:06
http://www.youtube.com/watch?v=uOeh2P3stP8	Live at Comedy Store	Rhod Gilbert	00:22:49
http://www.youtube.com/watch?v=pYtyetoQlxo	Rhod Gilbert's Work Experience	Rhod Gilbert	00:23:12
http://www.youtube.com/watch?v=RWYH0CV3Bmc	Rhod Gilbert's Work Experience	Rhod Gilbert	00:29:14
http://www.youtube.com/watch?v=PaKMTvNTDSA	Rhod Gilbert's Work Experience	Rhod Gilbert	00:29:14
http://www.youtube.com/watch?v=SNDoLJ8xdzk	Rhod Gilbert's Work Experience	Rhod Gilbert	00:29:16
http://www.youtube.com/watch?v=XofHS0_mkKs	Rhod Gilbert's Work Experience	Rhod Gilbert	00:29:14
http://www.youtube.com/watch?v=U10PaSy9t1M&list=PLYgCqOSKv5NerpZmqc\discretionary{-}{-}{-}{-}G8vZSGEkIK_7n60	Rhod Gilbert's Work Experience	Rhod Gilbert	00:14:59
https://www.youtube.com/watch?v=OMM_HjKlrOI&list=PLS-morIiTKJh26kJ9RKvj3P\discretionary{-}{-}{-}{-}1SeYRReF_G	Rhod Gilbert live comedy show	Rhod Gilbert	00:10:16
https://www.youtube.com/watch?v=rS6HjwPaaeg	Rhod Gilbert live comedy show	Rhod Gilbert	00:08:52

https://www.youtube.com/watch?v=rOGxHVVFP-M	Rhod Gilbert live comedy show	Rhod Gilbert	00:08:15
https://www.youtube.com/watch?v=-tcDXFBDtgE	Rhod Gilbert live comedy show	Rhod Gilbert	00:08:49
https://www.youtube.com/watch?v=ApgBbQQYOSU	Back to Llanbobl	Rhod Gilbert	00:07:53
https://www.youtube.com/watch?v=oVp3Q3bctTQ	Back to Llanbobl	Rhod Gilbert	00:09:35
https://www.youtube.com/watch?v=TQ4W7yB9Mow	Michael McIntyre's Christmas Comedy Roadshow	Rhod Gilbert	00:05:08
http://www.youtube.com/watch?v=gop40qXM5as	The One Show	Alex Jones	00:18:00
http://www.youtube.com/watch?v=_y02KVox1ho	The One Show	Alex Jones	00:15:00
http://www.youtube.com/watch?v=LvoztHyRXy8	The One Show	Alex Jones	00:13:00
http://www.youtube.com/watch?v=FNlslu5gHKg	The One Show	Alex Jones	00:13:22
https://www.youtube.com/watch?v=6M9xy0uu-Wc	The One Show	Alex Jones	00:12:00
https://www.youtube.com/watch?v=UNv4TjA2XN4	The One Show	Alex Jones	00:11:06
https://www.youtube.com/watch?v=HBtW7-LDXQk	The One Show	Alex Jones	00:08:00
http://www.youtube.com/watch?v=B_IM_aCoRQI	The One Show	Alex Jones	00:09:44

http://www.youtube.com/watch?v=IFCxE0ORdBI	The One Show	Alex Jones	00:20:00
http://www.youtube.com/watch?v=B8L6DFCo7xw	The One Show	Alex Jones	00:16:18
http://www.youtube.com/watch?v=gTrrvAcoM_M	The One Show	Alex Jones	00:14:53
http://www.youtube.com/watch?v=LAAtGNuKPZs8	The One Show	Alex Jones	00:14:41
http://www.youtube.com/watch?v=7Ya3R_D4ldU	The One Show	Alex Jones	00:13:00
http://www.youtube.com/watch?v=scs3BQSOzN4	The One Show	Alex Jones	00:13:35
http://www.youtube.com/watch?v=21JRDPpBI6U	The One Show	Alex Jones	00:13:25
http://www.youtube.com/watch?v=tOJmzI4iLyA	The One Show	Alex Jones	00:13:24
http://www.youtube.com/watch?v=fwPbpmrnOR0	The One Show	Alex Jones	00:10:23
http://www.youtube.com/watch?v=n4G-ME63RGs	The One Show	Alex Jones	00:09:52
https://www.youtube.com/watch?v=DA1PQONVJCs	The One Show	Alex Jones	00:11:56
https://www.youtube.com/watch?v=e5VgdZvfJks	The One Show	Alex Jones	00:12:24
https://www.youtube.com/watch?v=LcXXqbMoU-g	The One Show	Alex Jones	00:09:43
https://www.youtube.com/watch?v=Jbeeh1EAmPs	The One Show	Alex Jones	00:04:07

Not available online	Gavin & Stacey, Season 1 (6 episodes)		02:45:18
Not available online	Gavin & Stacey, Season 2 (7 episodes)		03:11:32
	Total		19:57:36

Appendix D: Questionnaire – perception of deviations from Standard English

Questionnaire: Deviations from Standard English

Some of the sentences below contain **deviations from Standard English**. Please circle all deviations that you find. E.g.:

I's going to town tomorrow.

1. *They is something bad wrong with her.*
2. *It's just the way I do speak.*
3. *Can you tell me what the difference is?*
4. *I'm here, amn't I?*
5. *The boys was interested, but Mary weren't.*
6. *The man who lives there is very nice.*
7. *It was very cold last night, so I walk home very quickly.*
8. *I don't know anything about snakes.*
9. *She's after selling the boat.*
10. *That one is for mummy, this one is for daddy.*
11. *I'm wondering is it Frank or not.*
12. *Sort of a student he was.*
13. *She lives just across the hallway, isn't it?*
14. *So you worked in the fields? - Oh yes, all the time.*
15. *Aunty Betty and the gang are here.*
16. *Give me five minutes to get the thing for you.*
17. *Let me finish this letter first.*
18. *The cat wants petted.*
19. *I did country dancing during the war.*
20. *I thought I might as well go on my own.*
21. *It's a really hard job, you know.*
22. *I'm going to the cinema every week.*
23. *It's an exhausting job. I'm happy to do it but.*
24. *I've been studying English for many years.*
25. *Do you like playing ping pong?*

26. *There's nice to see you.*

27. *It is so nice today because a sun is shining.*

28. *She got her a new car.*

29. *It is compulsory to attend the workshop.*

30. *That's much easier to follow.*

31. *I wanted that I should get leave.*

32. *They were married for sixty years.*

Where do you think the person who says the following sentence comes from?

I know a few people up there innit → The person comes from _____

Personal Information

Sex: ☐ Male ☐ Female

Year of birth: _____

Ethnic self-identification: _____

Native language(s):

☐ English ☐ Other: _____

Language(s) used at home
while growing up:

☐ English ☐ Other: _____

Native language(s) father:

☐ English ☐ Other: _____

Native language(s) mother:

☐ English ☐ Other: _____

Highest educational qualification:

☐ GCSE's/ O-Levels

☐ A-Levels

☐ Apprenticeship

☐ None

☐ Bachelor's degree

☐ Master's degree

☐ PhD

☐ Other: _____

Places lived at (at different ages).
Please indicate the cities/ towns/
villages:

0-10: _____

11-20: _____

21-30: _____

31-40: _____

41-50: _____

51-60: _____

61-70: _____

71-100: _____

Please indicate your opinion on the following statement using the scale:

"I like hearing a range of dialects."

Strongly agree

Strongly disagree

□-----□-----□-----□-----□-----□-----□

German summary

Salienz in der Grammatik des walisischen Englisch: Ein gebrauchsbasierter Ansatz

Diese Arbeit untersucht, inwieweit die Salienz dialektgrammatischer Konstruktionen durch deren Häufigkeit im Sprachgebrauch bedingt ist. Salienz wird unter einem soziolinguistischen Blickpunkt analysiert und bezieht sich in der vorliegenden Studie auf den Grad, zu dem ein sprachliches Merkmal als typisch oder charakteristisch für eine bestimmte regionale Varietät – hier: das walisische Englisch – wahrgenommen wird.

Befunde aus der Psychologie, der Kognitionswissenschaft und der gebrauchsbasierten Linguistik deuten darauf hin, dass die Wahrnehmung von sprachlichen Elementen durch deren Auftretenshäufigkeiten im Sprachgebrauch beeinflusst wird (vgl. z.B. Bybee 2006; Ellis 2012). Allerdings ist bislang wenig über die Rolle von Frequenz im Kontext der soziolinguistischen Salienz dialektgrammatischer Phänomene bekannt. Die vorliegende Studie untersucht mögliche Einflüsse von Tokenfrequenzen (im eigenen Dialekt und im Vergleich mit anderen) auf die Wahrnehmung morphosyntaktischer Konstruktionen des walisischen Englisch. In den Blick genommen werden sowohl intralektale Salienz (*intralectal salience*) – die Auffälligkeit von Phänomenen des walisischen Englisch für Waliser – als auch interlektale Salienz (*interlectal salience*) – die Auffälligkeit solcher Phänomene für Sprecher aus anderen Dialektregionen (vgl. Auer 2014). Zudem werden weitere mögliche Determinanten soziolinguistischer Salienz, etwa die geographische Ausbreitung eines Merkmals oder sozial-attitudinale Faktoren, untersucht und mit Frequenz in Beziehung gesetzt.

Die Salienz verschiedener grammatischer Konstruktionen des walisischen Englisch wird mithilfe einer Fragebogenstudie mit 150 Informanten aus Wales (intralektale Sali-

enz) und 150 Informanten aus London (interlektale Salienz) ermittelt. Zur Bestimmung der Gebrauchsfrequenzen der Konstruktionen werden Korpusanalysen herangezogen. Neben dem eigens angelegten *Radio Wales Corpus* mit gesprochen sprachlichen Daten aus Wales wird das *Linguistic Innovators Corpus* mit Sprechern aus London verwendet (vgl. Kerswill u. a. 2007). Weitere mögliche Determinanten von Salienz werden beispielsweise mithilfe des *Electronic World Atlas of Varieties of English* (eWAVE; geographische Ausbreitung) und Fragebogendaten (sozial-attitudinale Faktoren) analysiert.

Die Ergebnisse der Studien deuten darauf hin, dass Frequenz mit Salienz auf mehreren Ebenen interagiert. In Übereinstimmung mit Rácz (2013) haben interlektal saliente Merkmale eine geringe Auftretenswahrscheinlichkeit im Dialekt der Außenstehenden (Londoner Englisch) im Vergleich zum Zieldialekt (walisisches Englisch). So kommt beispielsweise die saliente „focus fronting“-Konstruktion mit deutlich höheren Tokenfrequenzen im Radio Wales Corpus als im Linguistic Innovators Corpus vor. Zudem korreliert Salienz mit Tokenfrequenzen im walisischen Englisch. Salientere Merkmale haben generell – aber nicht immer, wie das Beispiel der vergleichsweise niedrigfrequenten aber salienten „invariant tag question *isn't it*“ zeigt – eine höhere Auftretenswahrscheinlichkeit im Radio Wales Corpus als weniger saliente Phänomene. Die positive Korrelation kann sowohl für intra- als auch interlektale Salienz festgestellt werden, ist aber hinsichtlich intralektaler Wahrnehmungen deutlicher ausgeprägt. Neben Frequenz scheinen weitere Faktoren, wie etwa sozial-attitudinale Dispositionen der Sprecher, einen Einfluss darauf zu haben, wie ein Merkmal wahrgenommen wird. So wurden grammatische Phänomene des walisischen Englisch eher von Sprechern (Walisern und Londonern) erkannt, die eine positive Einstellung gegenüber dialektaler Vielfalt aufwiesen und/oder die Freunde oder Bekannte aus Wales hatten (Londoner). Insgesamt deuten die Ergebnisse auf die Komplexität soziolinguistischer Salienz hin. Während der frequenz-basierte Ansatz Wahrnehmungen von *durchschnittlichen Sprechern* zu einem gewissen Grad vorhersagen kann, werden *individuelle Wahrnehmungen* beispielsweise auch durch persönliche Erlebnisse und Einstellungen geprägt.

Die systematische Auslotung von Chancen und Grenzen eines frequenz-basierten Erklärungsansatzes für Salienz im Bereich der Dialektgrammatik liefert neue Erkenntnisse für die gebrauchsbasierte Linguistik und die perzeptuelle Dialektologie. Mit der

Analyse grammatischer Phänomene der bislang wenig erforschten Varietät des walisischen Englisch leistet die Arbeit zudem einen Beitrag zur Varietätenlinguistik.

Welsh summary

Y berthynas rhwng mynychder ac amlygrwydd mewn morffogystrawen: Achos Saesneg Cymreig

Mae'r gwaith hwn yn ymchwilio i ba raddau y pennir amlygrwydd nodweddion gramadegol rhanbarthol gan fynychder eu defnydd mewn iaith. Dadansoddir amlygrwydd o safbwynt cymdeithasol-ieithyddol, ac yn y gwaith hwn cyfeiria at y graddau y gwelir nodwedd ieithyddol fel un sy'n nodweddiadol o amrywiaeth ranbarthol benodol – yma: Saesneg Cymreig.

Awgryma ddarganfyddiadau seicoleg, gwyddoniaeth wybyddol ac ieithyddiaeth sy'n seiliedig ar ddefnydd fod canfyddiadau o iaith yn cael eu dylanwadu gan debygolrwydd digwyddiadau mewn defnydd iaith (cf. e.e. Bybee 2006; Ellis 2012). Hyd yma, fodd bynnag, nid oes llawer yn hysbys am rôl mynychder yng nghyd-destun mynegeio cymdeithasol nodweddion gramadegol rhanbarthol (cf. Rącz 2013). Gan ddefnyddio Saesneg Cymreig fel enghraifft, archwilia'r prosiect hwn effeithiau posib (absoliwt a chymharol) mynychder nodwedd ar amlygrwydd nodweddion morffogystrawen mewnlectaidd (o fewn ardal dafodiaith benodol – *intralectal*) a rhynglectaidd (y tu allan i ardal dafodiaith benodol – *interlectal*) (cf. Auer 2014). Ar ben hynny, caiff penderfynyddion posib pellach amlygrwydd cymdeithasol-ieithyddol, megis lledaeniad daearyddol nodwedd yn Ynysoedd Prydain neu ffactorau cymdeithasol-agweddol, eu dadansoddi a'u cysylltu â mynychder.

Profir amlygrwydd nifer o nodweddion gramadegol Saesneg Cymreig drwy arolwg ar sail holiadur gyda mwy na 150 o hysbyswyr o Gymru (amlygrwydd mewnlectaidd) a 150 o hysbyswyr o Lundain (amlygrwydd rhynglectaidd). Pennir mynychder defnydd o nodweddion drwy ddadansoddiadau corpws. Ar wahân i'r corpws y casglais fy

hun, Corpws Radio Wales, sy'n cynnwys data cyfweiliadau llafar o Gymru, defnyddiaf y Linguistic Innovators Corpus gyda siaradwyr o Lundain (cf. Kerswill et al. 2007). Dadansoddir penderfynyddion posib eraill amlygrwydd, er enghraifft, defnyddio Atlas Electronig y Byd o Amrywiaethau o Saesneg (eWAVE; lledaeniad daearyddol) a data o holiaduron (ffactorau cymdeithasol-agweddol).

Awgryma canlyniadau'r astudiaethau fod mynychder yn rhyngweithio ag amlygrwydd ar sawl lefel. Ar hyd llinellau Rácz (2013), mae gan nodweddion rhynglectaidd amlwg debygolrwydd isel o ddigwydd mewn tafodiaith pobl 'o'r tu allan' (Saesneg Llundain) o'i gymharu â'r dafodiaith darged (Saesneg Cymreig). Er enghraifft, mae'r gystrawen amlycaf "ffocws o flaen" yn nodwedd sy'n ymddangos llawer amlach yn y Corpws Radio Wales nag yn nata Linguistic Innovators. Ar ben hynny, mae'r amlygrwydd yn cyfateb â mynychder nodweddion absoliwt mewn Saesneg Cymreig. Mae'r nodweddion amlycaf yn gyffredinol – ond nid bob amser, fel y dengys enghraifft cymharol brin, ond amlwg, y "cwestiwn tag digyfnewid *isn't it*" – yn fwy mynych yn y Corpws Radio Wales na ffenomenau llai amlwg. Ceir cydberthyniad cadarnhaol o ran amlygrwydd mewnllectaidd a rhynglectaidd; fodd bynnag, mae'n gryfach o ran canfyddiadau mewnllectaidd. Ar wahân i amllder, cafodd fy astudiaeth fod ffactorau megis thueddiadau cymdeithasol-agweddol siaradwyr hefyd yn dylanwadu ar ganfyddiad o'r nodwedd. Adnabuwyd nodweddion gramadegol Saesneg Cymreig, er enghraifft, yn amlach gan siaradwyr (Cymry ac yn Llundain) ag agweddau cadarnhaol at amrywiaeth dafodieithol a/neu â ffrindiau neu gymdeithion o Gymru (pobl o Lundain). Ar y cyfan, dengys y canlyniadau gymhlethdod amlygrwydd cymdeithasol-ieithyddol. Er y gall y dull sy'n seiliedig ar amllder ragweld canfyddiadau defnyddwyr iaith cyffredin i ryw raddau, caiff canfyddiadau unigol hefyd, er enghraifft, eu bathu gan agweddau a phrofiadau personol.

Mae ymchwiliad systematig i botensial a chyfyngiadau ymagwedd esboniadol ar sail amllder at fynychder ym maes gramadeg tafodiaith yn darparu craffter newydd ar gyfer ieithyddiaeth ar sail defnydd a thafodieitheg ganfyddiadol. Mae'r gwaith yn cyfrannu ymhellach at ieithyddiaeth gymdeithasol amrywiadol drwy ddadansoddi nodweddion gramadegol yr amrywiaeth Saesneg Cymreig sydd heb ei hastudio'n ddigonol hyd yma.

This book explores salience in dialect grammar, i.e. the degree to which features are perceived as characteristic of Welsh English. The author approaches salience from a usage-based perspective, testing the extent to which a feature's salience can be predicted by its frequency in language use. So far, not much is known about the impact of frequency on the conspicuousness of regional grammatical features. The present study provides a systematic analysis of the relation between salience and different forms of token frequency (e.g. frequency in the target dialect, frequency differences between the target dialect and the dialect of outsiders). In each of these analyses, salience for insiders (people from Wales) is contrasted with salience for outsiders (people from London). Assuming that frequency interacts with other factors, the study also explores how salience in dialect grammar is related to a feature's areal spread, its linguistic structure and social-attitudinal characteristics of the listeners.

Six Welsh English features are investigated on the basis of three sets of data: questionnaires (salience, structural and social factors), corpora (frequency) and eWAVE (*Electronic World Atlas of Varieties of English*; areal spread). In addition to making use of existing corpora, the author presents a new, self-compiled corpus of Welsh English, the *Radio Wales Corpus*.

The systematic account of benefits and limitations of a frequency-based approach to salience in dialect grammar provides new insights for usage-based theories of language and (perceptual) dialectology. Overall, this study helps consolidate the emerging field of *cognitive/usage-based dialectology*.

Katja Roller majored in English (teacher qualification) at the Universities of Bamberg (Germany) and Swansea (Wales). She obtained her degree in 2012. From 2012 to 2015 she was a member of the Freiburg-based doctoral research training group "Frequency Effects in Language" (DFG GRK 1624). She defended her PhD thesis at Freiburg University in May 2016. This book is a revised version of her dissertation.

ISBN 978-3-928969-67-3



9 783928 969673

UNI
FREIBURG