Leibniz Universität Hannover
Philosophische Fakultät
Englisches Seminar
Sommersemester 2014
Masterarbeit
Erstprüferin: Prof. Dr. Ulrike Altendorf
Zweitprüfer: Jun.-Prof. Dr. Daniel Schnitzlein

Applying multivariate estimation models to investigate factors promoting the use of the passive voice in the writing of EFL learners

Tobias Gärtner
Clemensstraße 10
30169 Hannover
tobias_gaertner@htp-tel.de
Mat.Nr.:2724160
Master of Education Englisch / Ev. Theologie
10. Semester

Gewidmet meinen Eltern in Dankbarkeit

Contents

List of Figures

List of Tables

Listings

List of Equations

1 Introduction

The passive has been approached from different theoretical perspectives. Also, it has been analysed as a part of the research on complexity and has been counted and statistically tested. There seems little reason to deal with it again.

Yet, this master's thesis shows that up to now only a few aspects of the passive have been taken into account. The passive has either been manually counted or basic search algorithms have been used to let a computer do the work. In contrast to that this paper introduces new tools that do not only count the number of passives per text but are also able to distinguish more than 45 different tenses, aspects, voices and modes. The following chapters present standard statistical procedures taken from psychology and economics which have rarely been used in linguistics. With theses tools to the dependency of the passive voice on multiple variables is illustrated. Yet, since all these variables are connected with each other, their mutual influence is analysed with multivariate and not just univariate methods.

The aim is nonetheless not only to show new ways of linguistic analysis but primarily to show that the characteristics and dependencies of the passive are complex. Despite the linguistic orientation of this master's thesis the description and discussion of the newly applied methods require half of the pages. Furthermore, the above mentioned complexity of the passive constructions as well as the methods applied in this paper restrict the number of discussed passive constructions. The remaining figures, graphs, tables and source codes can be found in the Appendix (starting from page 70).

The overall goal of this paper is to investigate the influence of administrative, social and linguistic variables on the frequency of different passive constructions in the essay writing of undergraduate university students of English. All of these students are native speakers of 17 different languages and non-native speakers of English. In order to do that, a linguistic parser, data-base tools, Perl and Windows Shell scripts and R are used. The Figure 1 on the following page gives an outline of the procedures which are described in Chapter 3 and applied in Chapter 4.

Chapter 2 will illustrate the deviances between the above sketched procedures and the state of the art literature which is why the aim of this paper is reformulated in the beginning of Chapter 3. Due to the lack of referential literature, statistical choices are made on the basis of statistical tests and parameter values, and not on the basis of theoretical considerations.

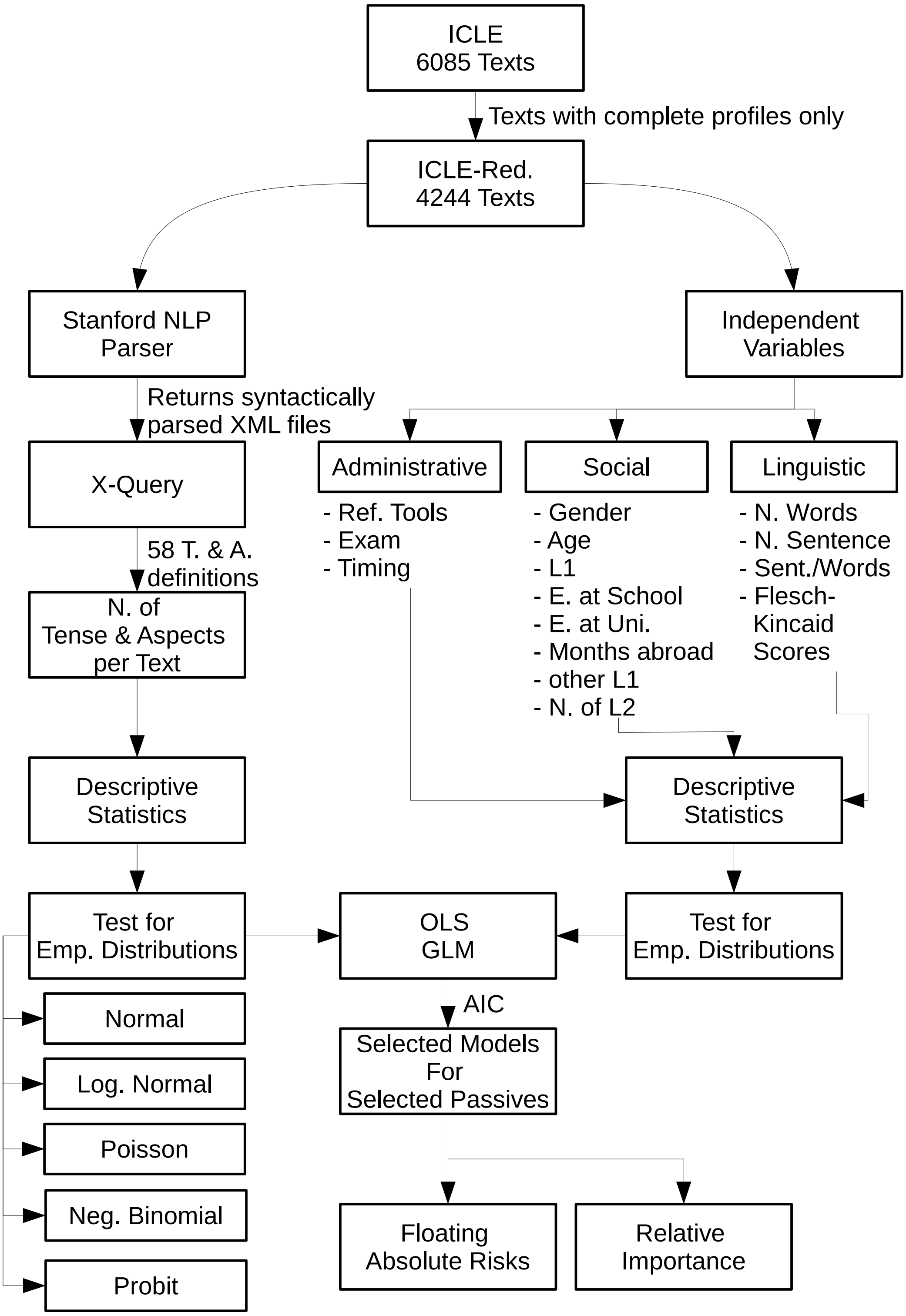


Figure 1: Roadmap for the Analysis of the Passive Voice.

2 State of the Art

2.1 Linguistic Literature Review

In the recent past the passive has been of central interest for several researchers. Although this thesis is data driven and thereby quantitative, the results of qualitative studies shall not be omitted.

Puckica (2009) differentiates between central and marginal passives. The central passive can also be split up into two categories. The passive as it is taught in many schools and which is used in this thesis is called " simple central passive " (Puckica 2009, 221) and consists of a form of to 'be' or 'get' and a past participle. Puckica defines certain participle constructions as periphrastic central passives. These phrases do not contain a finite yet only a non-finite verb form, i.e. the past participle, and denote a passive meaning (Puckica 2009, 222). Like the periphrastic central passive the marginal passive is expressed by present participles and adjectives based on verbs (Schmid 2011, 175). All of the following sentences are considered to express the passive voice according to Puckica (2009, 216,221,223):

(1) The employee was fired (by the manager).
(2) The employee fired by the manager is [...] .
(3) Again, this is a serious defect and [it] needs checking [...].
(4) Those tenements shall be recoverable [...].

It needs to be noted that Puckica focuses merely on the semantic function and not on the syntactical and morphological form of the passive.

James Blevins (2003) states that many constructions are mislabelled as being passive. According to him many phrases labelled as being passive are subjectless and thereby form an " impersonal" or " autonomous" voice (Blevins 2003, 474). This implies that the object of an active voiced sentence does not necessarily become the subject when passivised but the function of the subject can be omitted. Like Puckica Blevins follows the Head Driven Phrase Structure Grammar Framework (HPSG), which is a purely qualitative approach. It investigates linguistic phenomena in a context-free environment. Not the form is crucial but the function, i.e. the position of the subject can be occupied in terms of the word order but at the same time empty in terms of the function.

Marianne Hundt (2004) analyses the development which the passive in general and the progressive passive in particular underwent during the last 400 years, i.e. the entire modern English period. She uses A Representative Corpus of Historical English Registers (ARCHER) for her quantitative study. According to Hundt's findings the frequency and popularity of the progressive, the passive and the progressive passive are rising since the early 19^{th} century (Hundt 2004, 104f.). Additionally, the progressive passive seems to be two times more frequent in British than in American English (Hundt 2004, 111)[1].

In general the progressive passive has received much attention. Lieselotte Anderwald (2014) shows that, corresponding to Hundt's (2004) results, the number of grammar books dedicated to the progressive passive also rose from 1800 onwards. Anderwald also finds that especially British grammarians were more open to the new forms of passive constructions whereas American grammarians took longer to consider it appropriate. Anthony Warner (1995) claims that the development of the progressive passive is due to the loss of 'thou' and the resulting change of 'have' and 'be'. He supports his claim using the HPSG approach and therefore has no quantitative results.

According to Langacker & Munro (1975) the discussion on the syntactical structures of the passive takes place in two directions. Under the assumption that there is something like an underlying grammatical structure, which excludes semantic approaches such as HPSG, the passive can, according to Chomsky, either be a transformation of an active sentence, or, as criticised by Lakoff and Langacker as a subjectless construction (Langacker & Munro 1975, 792). As a result of the discussion there is the question of how to treat the verb in a passive sentence. Langacker & Munro state that 'be' can either be regarded as a grammatically necessary auxiliary in reference to Chomsky, or, referring to Lakoff, it can be the main verb of the sentence with the participle as a complement (Langacker & Munro 1975, 792).

One of the first quantitative studies on the frequency of the passive is the study by Weiner & Labov (1983). Following the socio-linguistic approach as put forth by Labov (1972), Weiner & Labov analyse the active/passive ratio in spoken English in Philadelphia. For their statistical analysis they use the second version of the Variable Rule (VARBRUL) algorithm (Weiner & Labov 1983, 39) and a log-likelihood based χ^2 test. The VARBRUL or GoldVarb algorithm is an early version of what is later called logistic regression (Eddington 2010). It predicts whether, under the influence of a socio-linguistic variable, a grammatical rule holds true (weights > 0.5) or not (weights < 0.5). The χ^2 test evaluates if the VARBRUL weights of the sub-samples are significant in contrast to the entire sample. Yet, it needs to be mentioned that this approach is not multivariate and therefore easily falls prey to the Simpson Paradox (Simpson 1951)[2]. Despite these statistical restrictions Weiner & Labov show that neither social class, age or gender have any influence on the frequency of the passive in spoken English. Only genre, i.e. the contrast between careful and casual speech, seems to have an impact, with speakers favouring

[1] These findings have consequences for the analysis of NNS writing. See Chapter 3.2 for a further discussion.

[2] The Simpson Paradox states that the contrast between sub-samples, separated on the basis of the parameter values of a variable, can be biased by a second variable and that the composition of the sub-samples can bias the overall results. It is therefore necessary to take a closer look at the single results.

the passive in careful speech. Moreover, Weiner & Labov argue that pragmatic and referential concerns influence the preference of the passive over the active. If the object of the preceding sentence can be turned into the subject of the passivised sentence, the passive is significantly more often chosen than in sentences where this is not possible (Weiner & Labov 1983, 46f.). Although these results might not withstand a deeper statistical analysis with other tools, they nonetheless support the idea of genre dependency of the passive, which was later reformulated by Douglas Biber (1992).

Douglas Biber (1992) investigates the complexity of spoken and written discourse. He uses 33 linguistic markers of complexity, grouped into the following six dimensions: "Reduced Structure and Specificity, Structural Elaboration of Reference, 'Framing' Structural Elaboration, Integrated Structure, and Passive Constructions" (Biber 1992, 133). According to him the main variable, which is supposed to influence discourse complexity, is the text genre. Therefore, 23 text genres, as they appear in the Lancaster Oslo Bergen (LOB) and the London Lund Corpus (LLC), have been evaluated. Combined both corpora consist of 481 text with a cumulated word count of approximately 960,000. Since there is no explicit discussion and because of the date of publication it can be assumed that the 33 linguistic features have been counted by hand. Biber states that these features only scratch the "surface" (1992, 141) of complexity, since cohesion or rhetorical devices are not part of the analysis. In terms of statistical evaluation, Biber uses a Confirmatory Factor Analysis (CFA). The CFA assumes latent underlying structures, in this case the above mentioned six dimensions. Biber formulates statistic models that contain increasing numbers of complexity dimensions. He tests if the addition of a further dimension significantly improves the goodness of fit (1992, 142ff.). The CFA shows that the use of the passive is a strong indicator for complexity and that it has a high positive correlation with reduced structures, specificity and integrated structures. This means that speakers, who frequently use complex constructions in their language performance, also use passive constructions (Biber 1992, 147ff.). Consequently, the passive is mainly used in text genres that are already complex due to their content, i.e. official documents and academic prose. To include his findings into this paper, the text genre is also used as a independent variable in the estimation models (see Chapter 3.7). If Biber's theory holds true, then the text genre will have a significant influence on the frequency of the passive.

One of the earliest investigations of the use of the passive by NNS is Sylviane Granger's "The *be + Past Participle* Construction in Spoken English" from 1983. Because it is more than 30 years old most of the then revolutionary approaches and methods have been altered. Thus, only brief insights into her research will be given. The corpus basis for her research covers 32 texts with about 160,000 words collected in the 1960s and the 1970s. Nine years before Biber Granger also finds a genre dependency of the passive voice. According to her results, the passive appears most frequently in careful and complex speech such as discussions and orations (Granger 1983, 270). This impression is supported by a χ^2 test evaluating the homogeneity of the distribution of the passive within the different genres. The text genre or style are the only variables analysed.

In 1997 Sylviane Granger analysed passive voice again. This time LOCNESS and parts of ICLE are used. The passive is used to discuss the chances and limitations of an automatic retrieval of grammatical constructions (Granger 1997, 365). According to Granger the limitations arise from the performance of the used tagger. I have neither found the programme on the internet nor any reference to it published in the 21^{st} century and therefore could not retry her experiment. Concerning the passive voice, Granger shows that NNS apply significantly fewer passive constructions than NS.

Recently Granger published another paper on the passive (2013). In this paper her search algorithm has limitations, since it only searches for a form of 'to be' directly followed by a past participle. This results in the fact that passive constructions with additional words between the form of 'to be' and the participle are excluded (Granger 2013, 5f.). Nevertheless, she finds a three dimensional dependency of the passive. The first dimension refers to Kachru's circle model (1992). According to Granger (2013, 6f.) outer and expanding circle varieties underuse the passive compared to inner circle varieties[3]. In a second dimension the passive does not seem to be proficiency dependent. Judging from the figures presented in her paper, the difference between the single proficiency levels does, at least for Japanese learners, not seem to be significant. However, according to the previously discussed literature and to the formulation of Granger's hypothesis, it can be assumed that these findings are likely due to chance and perhaps cannot be generalised. The third dimension is exterior influence. Sylviane Granger shows that the formulation of the task can already influence the frequency of specific language features (2013, 7).

Although there are many more papers and books on various aspects of the passive, only a selection can be discussed. Hence, the results of only one further research paper will be addressed. For a full analysis of corpus based research see Granger (2013).

One of the few papers that use similar methods as those that are described in the following chapters is Eli Hinkel's *Tense, aspect and the passive voice in L1 and L2 academic texts* (2004). In total there are 746 texts with 226,054 tokens written by 115 NS and 631 NNS from China, Japan, Korea, Indonesia, Vietnam and several Arabic-speaking countries form the basis for her contrastive quantitative study. The NNS texts are taken from entrance tests for four US universities. For further details on the composition of the NNS group see Hinkel (2004, 11-13). In terms of word count and text genre Hinkel's corpus is similar to ICLE (Hinkel 2004, 12f./Chapter 3.5). Tenses, aspects and voices are counted manually and a ratio in relation to the word count is computed (Hinkel 2004, 13). In order to test differences between the NS and NNS group, the Mann-Whitney U

[3] This claim is not statistically supported.

test is chosen (Hinkel 2004, 14/Sheskin 2000). Hinkel calls this test "conservative" (2004, 14) as it reduces the metric scaled counts to ordinal scaled data. Her findings support the idea that speakers of some Asian languages such as Chinese, Japanese, Indonesian, and Vietnamese have problems acquiring complex English tenses and aspects (Hinkel 2004, 14). The progressive and perfect aspect as well as the predicative 'would' each have median rates of zero for all NNS varieties. While the simple past is significantly overused by all NNS varieties, the remaining non-simple forms, including the passive, are significantly underused. For NS the passive is the third most frequent construction right after the present and past simple whereas for Chinese and Indonesian students it is on position four. For Arabic students the passive voice is on the second position. In general the median frequencies of the different constructions greatly vary between the single L1 groups. For Hinkel the observed deviances are largely due to inappropriate textbooks and native language specific transfer (2004, 8, 17).

2.2 The Analysis of Count Data

As it has been shown in the last chapter, the majority of linguistic literature focusses on a precise linguistic description of the forms, functions and implications of passive voice. These papers and books which use quantitative methods usually do not go beyond comparing percentages, χ^2 tests or early versions of logistic regressions. Yet, my approach requires statistical tools which allow for a simultaneous analysis of several variables. To find the appropriate tool either all available tools can be used and then tested for the best fit or appropriate literature can be consulted. Unfortunately, there is only a small amount of literature available on the subject. One of the leading statisticians in linguistics, Stefan Gries, also just briefly discusses multivariate approaches without going into detail (2003) or just mentions them as a further possibility (2013).

In order to find an appropriate model type the characteristics of the independent variable have to be declared. The number of passives per text can only be zero or above. Additionally, it can only be an integer. Because of these characteristics the passive and many other linguistic phenomena follow the definition of count data (Lancaster 2004, Cameron & Trivedi 2005, Winkelmann 2008). The corresponding regression models are for example Poisson, Negative Binomial, Zero-Inflated and Hurdle Regression models (Cameron & Trivedi 2005, Chapter 20). To prevent this paper from focussing too much on statistics only the first two model types shall be examined. The underlying mathematical formulae derive from McCullagh & Nelder (1972). Further information can be found in Lancaster (2004),Cameron & Trivedi (2005) and Winkelmann (2008), the respective *R* packages by Venables & Ripley (2002), Rigby & Stasinopoulos (2005), Zeileis, Kleiber & Jackman (2008), Jackman (2012) and Chapter 3.4.

3 Methodology

Following the literature review in the previous chapter the goal of this paper needs to be reformulated. In order to allow for a deeper analysis of previous research, different methods will be used in this master's thesis. First of all search algorithms are formulated to avoid manual searches which are prone to errors and to find more different grammatical constructions. Second, statistical procedures have to be applied that fit the data. Passive constructions per text are countable metric data. This type of data offers more statistical possibilities which should be used. Third, as language is not mono-causal and since Simpson has illustrated that different variables influence each other, multi-variate tools, which take both considerations into account, need to be applied to achieve the best fitting model.

Consequently this master's thesis will formulate search algorithms capable of automatically returning most of the passive constructions[4]. They will be described on the basis of the standard techniques of econometrics. Multi-variate regression models will be formulated to appropriately take account of the linguistic and social information given and acquired. Even though the following chapter on the methodology applied in this thesis is rather extended, the paper nevertheless aims for a corpus-linguistic focus, which makes it comparable to what has been done before. It will be shown that the passive is heavily dependent on the L1, whereas other variables have little to no influence at all. At the same time relatively little can be said about some other variables as their distribution is too narrow for a deeper analysis.

The following sub-sections will briefly introduce the methods and variables used. They are not sorted by academic disciplines but by the order they are applied[5]. Key concepts are picked up to be included into a glossary which can be found at the end of this thesis.

3.1 The Use of Annotations

This paper combines several academic disciplines. Each discipline has its own way to apply annotations, which will be outlined in this chapter.

There are four types of non-text items in this master's thesis. For each item type there is a corresponding list which can be found right after the Table of Contents. Tables and figures have their caption underneath them, whereas source codes have the caption above them. The font of the source codes resembles the font of their respective language editor, i.e. HTML looks different from Perl and from R. Mathematical formulae have no caption but are named and explained in the paragraphs above them. In order to find them more easily, there is a list of mathematical functions which includes the names of the concepts

[4] Chapter 3.3 will discuss these shortcomings.

[5] Confer Figure 1 on page 2.

that the respective formula expresses.

Abbreviations and Greek letters for mathematical concepts are explained when used for the first time. A list of them can be found in the glossary.

There are no further explanations for basic mathematical operators like plus (+), minus (−), greater-equal (≥) or absolute value (| |).

Quotes are put into double inverted commas (" ") whereas linguistic examples are put into single inverted commas (' ').

3.2 From CA to CIA to CONNSIA

The theoretical framework of this paper is based on Sylviane Granger's (1996) revolutionary article *From CA to CIA and back*. In this article she defines a new branch of corpus linguistics. Up to that point corpus linguists focussed on comparing different languages and translations with each other.

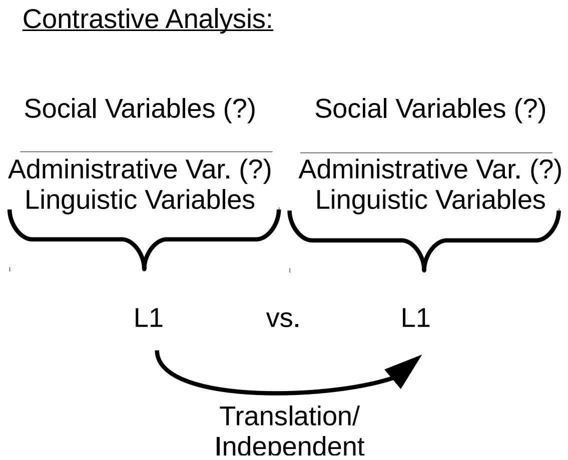


Figure 2: Schematic Description of the Contrastive Analysis.

Being called Contrastive Analysis this method works on two levels. On a first level texts with the same content but in different languages are compared with each other. Granger (1996, 39) uses the example of a British and a French newspaper reporting on the same topic, assuming that both articles are written by native speakers of the respective language. On a second level CA compares the translation of a text into another language with the original text. Again it is assumed that the writers of both texts are native speakers of their respective language. Problems arise in trying to generalise the results of the CA (Granger 1996, 42), because the corpora the CA is based on are comparatively small and contain few additional information on the authors. Thus, a reliable statistical analysis becomes difficult to impossible (see Figure 2). Moreover, the possible genres for analyses are limited since there is only a small number of texts that is professionally translated. As this paper tries to investigate the performance of English by non-native speakers the CA cannot be applied.

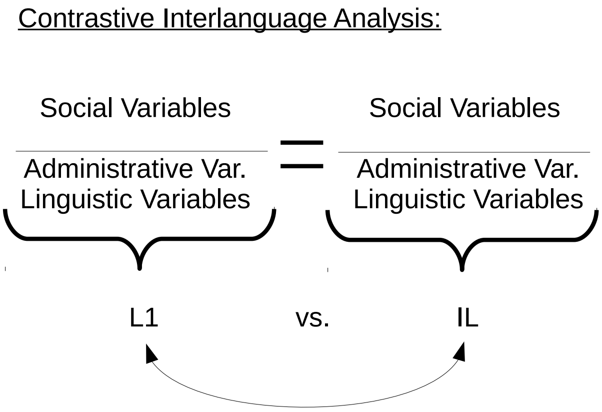


Figure 3: Schematic Description of the Contrastive Interlanguage Analysis.

In 1969 Larry Selinker developed the concept of the so called Interlanguage (IL). This interlanguage is a ”separate linguistic system” (Selinker 1972, 214) of language learners who almost reach a native-speaker language performance. This linguistic system is independent from the L1 as well as from the Target Language. According to Selinker, it is necessary to be able to predict the interlanguage in order to enhance the learning process for learners of English as a Foreign Language (EFL) (1972, 214). To predict the IL performance other methods and corpora are required than the CA provides. As a reaction to these shortcomings Granger advanced the CA by including the IL and thereby allowed for the comparison of an L1 with one or several ILs, or different ILs with each other (1996, 43-45). In order to do so the distinct sub-corpora have to be comparable in their characteristics. Yet, Granger states that only comparing different ILs is not enough for a sufficient linguistic analysis and therefore it is necessary to have ”access to a comparable corpus of NL English” (Granger 1996, 45). For further analyses not only the genre but also other variables need to be similar to assure a statistical comparability (see Figure 3). Sylviane Granger mentions that the University of Louvain is assembling such a corpus. By now this corpus is available under the name LOCNESS. Nevertheless, LOCNESS cannot fulfil the needs of the Contrastive Interlanguage Analysis (CIA). It is not comparable due to the lack of information on the administrative circumstances and the social characteristics of the writers. Hence, the current framework, i.e. the CIA, has to be advanced again to allow for an investigation of ILs without the need for an NS corpus.

The lack of an adequate NS corpus is not the only shortcoming of the CIA. Under the assumption that only the L1 influences the frequency of the passive, or any other construction, three groups may be given. The first group may consist of American students speaking flawless General American English (GA). The same number of British students speaking flawless Received Pronunciation (RP) may be the second group. The last group could be made up of Bulgarian students. A significance test such as a χ^2 test reveals that the difference between the GA and the RP group is significant. This results in a dilemma, as the Bulgarian students cannot follow both target varieties at the same time. Additionally, if by chance half of the Bulgarian students speak perfect RP and the other half perfect GA, neither of both, despite their perfect English as a Foreign Language competence, is regarded perfect under the given circumstances. Mixing both NS groups does not help either, as the statistically and linguistically significant differences are thereby hidden. Therefore, in order to compare the different NNS groups with NS groups, a corpus has to be found or created that contains various NS and NNS groups at the same time, which have the same characteristics, and where the single NNS declare which target norm they would like to follow. In an ideal case this declaration is also tested.

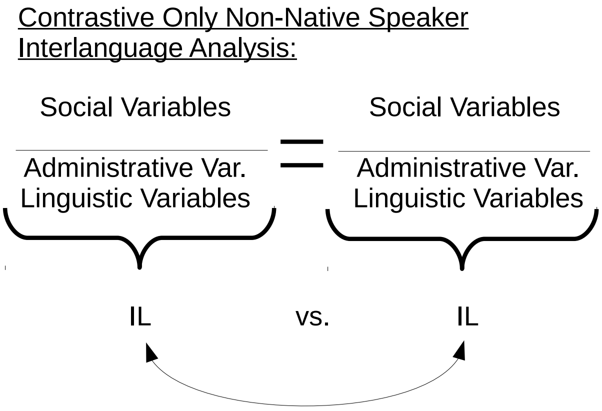


Figure 4: Schematic Description of the Interlanguage Analysis.

As this thesis investigates the usage of the passive under the interlanguage assumption and since a corpus as defined above does not exist, I reject the theory of Contrastive Interlanguage Analysis and instead suggest the Contrastive Only Non-Native Speaker Interlanguage Analysis (CONNSIA/see Figure 4). CONNSIA assumes that the statistical procedures presented in the following chapters as well as the theoretical assumptions stated above justify working without an NS reference corpus. It only relies on NNS language performance.

3.3 Tenses, Aspects, Voices and Modes

The English language knows three tenses, four aspects, two voices and modes (Huddleston & Pullum 2007). This results in twelve active indicative, twelve passive indicative, four active conditional and four passive conditional forms. Additionally, there are two future simple forms, namely 'will future' and 'going-to future', both available in active as well as in passive voice. Also, some more forms need to be added due to restrictions of the Stanford NLP parser used in this paper. So, in total there are 56 distinct forms. To define and describe each of these forms within this chapter would go beyond the constraints of this paper. The X-Query definitions for all tenses can be found in the corresponding chapter in the Appendix (from Page 72 onwards).

The following Table 1 (page 13) uses the sentence fragment 'they write a letter' in all tenses, aspects, modes and voices to give a grammatical foundation for the upcoming analyses. At this point it needs to be mentioned that the conditional mode is included in Table 1 as well as in the following tables. Yet, it will be excluded from in-depth analyses for two reasons. The parser does not distinguish between modal auxiliaries, i.e. 'can', 'must', 'should' etc., i.e. all receive the same label despite their different function. Also, none of the literature presented in Chapter 2 discusses the conditional passive and therefore I will do without it. Additionally, I cannot assure to satisfyingly define all conditional forms[6] and thereby also cannot programme a reliable search algorithm to include the conditional mode. Therefore, this paper will exclude the conditional from the detailed analysis. For a more detailed discussion of these grammatical categories confer Quirk, Greenbaum, Leech & Svartvik (1985), Pullum & Huddleston (2002) and Huddleston & Pullum (2007).

3.4 Statistics

Since the tools in this chapter are taken from econometrics, a more detailed presentation of their function for linguistic purposes is required. The statistics chapter is divided into three parts. The first part introduces the basic descriptive statistics, i.e. those basic formulae and values that are required for the following more advanced procedures. Some estimation models require their dependent and independent variables to follow certain distributions and therefore the second part introduces a number of empirical distributions. The last part combines results of the previous two sub-sections and shows how estimation models are formulated in order to give the best forecast of the frequency of tenses and aspects.

[6] Especially since the parser does not recognise the subjunctive because it has the same form as the infinitive.

Active

	Simple	Perfect	Progressive	Perfect Progressive
Present	write	has written	is writing	has been writing
Past	wrote	had written	were writing	had been writing
Future	will write	will have written	will be writing	will have been writing

Passive

	Simple	Perfect	Progressive	Perfect Progressive
Present	is written	has been written	is being written	has been being written
Past	was written	had been written	was being written	had been being written
Future	will be written	will have been written	will be being written	will have been being written

Conditional Active

	Simple	Perfect	Progressive	Perfect Progressive
Present	should write	should have written	should be writing	should have been writing

Conditional Passive

	Simple	Perfect	Progressive	Perfect Progressive
Present	should be written	should have been written	should be being written	should have been being written

Table 1: Tense, Aspect, Voice and Mode examples using the abbreviated phrases ”they write a letter”, ”a letter is written”, ”they should write a letter” and ”a letter should be written”.

3.4.1 Descriptive Statistics

Two measurements for the central tendency will be used: the median and the mean. The median, as the name already suggests, is the value in the middle of the distribution. For that purpose the raw values x_i are sorted by their value $x_{(i)}$ followed by a definition of the 0.5 quantile (see Equation 1).

$$x_{0.5} = \begin{cases} x_{\left(\frac{n+1}{2}\right)} & \text{for odd } n \\ \frac{1}{2} \cdot \left(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}\right) & \text{for even } n \end{cases} \tag{1}$$

In order to calculate the mean (μ) the raw values do not need to be sorted. For the mean all values are added up and are then divided by the number of observations (see 2).

$$\mu = \frac{\sum_{i=1}^{N} x_i}{n} \tag{2}$$

Both values have advantages and disadvantages. In contrast to the mean, the median is robust towards outliers, i.e. extreme values on either side do not influence the median. On the other hand the median can lead to wrong impressions with non-linear distributions, such as x^3 between 0 and 100, since in this case the mean (252,500) is more than twice as high as the median (125,000).

The mean is the first moment of a distribution. It only states the central tendency but neither the width, nor the symmetry, nor the relation of tails to shoulder of the distribution it describes. Therefore, three more moments are required.

The second central moments are variance (σ^2/ Equation 3) and standard deviation (σ/ Equation 4). Variance is the average squared distance from the mean, while the standard deviation is the average difference expressed in the same unit as the observations. The advantage of the standard deviation over the variance is that the standard deviation has the same unit as the raw values of the distribution.

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu)^2 \tag{3}$$

$$\sigma = \sqrt{\sigma^2} \tag{4}$$

With the first two moments the highest concentration and the width of the distribution can be described. What remains unknown is whether the distribution is symmetric or otherwise, if one tail is heavier than the other. A symmetric distribution such as a normal distribution has a skewness (γ) of 0. A positive skewness signifies a heavier right than left tail and for a negative skewness vice versa. The higher $|\gamma|$ the heavier the imbalance (see Equation 5).

$$\gamma = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^3 \tag{5}$$

The last moment is the kurtosis (κ/ Equation 6). The kurtosis expresses the relation between the shoulders and the tails of a distribution. The reference point is a normal distribution, where $\kappa = 3$. The heavier the shoulders in comparison to the tails the higher the kurtosis. A negative kurtosis stands for an imbalance with heavier tails than shoulders.

$$\kappa = \frac{1}{n}\sum_{i=1}^{n}(\frac{x_i - \mu}{\sigma})^4 \tag{6}$$

With the help Pearson's correlation coefficient r (see Equation 7) the possible connection between two metric variables can be shown. Let X and Y be two metric variables. Pearson's correlation is a measure for the linear relationship between X and Y and is defined by Equation 7. It attains values between -1 and 1. This means, if r = 1 then Y raises by one unit on average if X raises by one. If r=0 then they do not have a linear relationship.

$$r_{xy} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\left(\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2\right)^{\frac{1}{2}}\left(n\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2\right)^{\frac{1}{2}}} \tag{7}$$

Yet, the correlation coefficient only shows connectedness and not reason. Thus, if the coefficient between the number of storks and babies is 0.80, this only means that in regions with many storks many babies can be found and not that storks bring babies.

3.4.2 Empirical Distributions

There are several ways to find the best model to predict a phenomenon. The easiest way is to find it in literature. But, as it has already been mentioned in the introduction, the statistical procedures and techniques used in this paper are rather new to linguistics. There is next to no literature giving hints on model selection[7]. The second way is to set up all kinds of models and to test them for best fit, using something like the Akaike Information Criterion (AIC)[8]. Yet, this approach has one great disadvantage: there are too

[7] Gries (2013) is one of the latest and most advanced statistics book for linguists, yet his explanations end with linear models.

[8] The AIC evaluates the fit and complexity of a model on the basis of the log-likelihood. It is a relative measure to compare different model types with each other. Also see Equation 23.

many different types of models to test them all and most of the models have requirements that should not be violated. Therefore, it is most reasonable to test the dependent variables, in this case the passive, for different empirical distributions and to select the appropriate model in a second step.

An example will be used to illustrate the process of testing a variable for various empirical distributions. The example be the body weight of becoming mother's in pounds at their last menstrual period. The body weight be the variable to be tested[9] and the Gaussian normal distribution be the distribution to be tested for.

In a first step the histogram shall be plotted. For the histogram the range of the variable is divided into categories and the number of values fitting into these categories counted. The hight of these categories can either be the frequency, i.e. how often the value occurs in absolute numbers, or the density, i.e. the relation between the number of absolute values in category m and the total n. There is no hard and fast rule how many categories should be used. Nonetheless, Bortz (2010) suggests referring to Sturges (1926) to use $1 + 3.32 \cdot lg(n)$[10] categories[11]. The histogram for the above mentioned data using Sturges' formula to calculate the number of categories can be found on the following page.

This histogram can now be compared to model distributions with the same characteristics, i.e. same parameters. Some of these parameters can be directly calculated, such as μ and σ[12]. Others have to be estimated with the help of the maximum likelihood estimator (MLE). The probability density function (PDF) of a normal distribution has two parameters μ and σ, which are already known from the previous chapter. In order to come to the curve shown in Figure 5 the two parameters have to be inserted into the respective PDF:

$$p(x; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{8}$$

In the case of the body weight of pregnant women the range is 80 to 250 pounds. Therefore, it makes sense to also restrict x to these values. The thereby created curve can be seen in Figure 5 (page 17). The closer the curve fits the histogram the better the fit. Yet, the histogram only gives a first impression of whether or not the distribution actually fits. Additionally, comparing histograms becomes with more than 50 passive constructions and a dozen empirical distributions unnecessarily time consuming.

As an alternative to visual comparisons, significance tests can be used to evaluate the

[9] The example data is taken from the *MASS* package by Venables & Ripley (2002). The authors have collected the data at the Baystate Medical Center, Springfield, Mass in 1986.

[10] Lg stands for the logarithm to the base e whereas for all log logarithms the base will be stated.

[11] R's *hist()* function also allows to set the number of categories according to Sturges' formula.

[12] To be precise μ and σ are also maximum likelihood estimates. Nonetheless, their calculation can be achieved by less complex means.

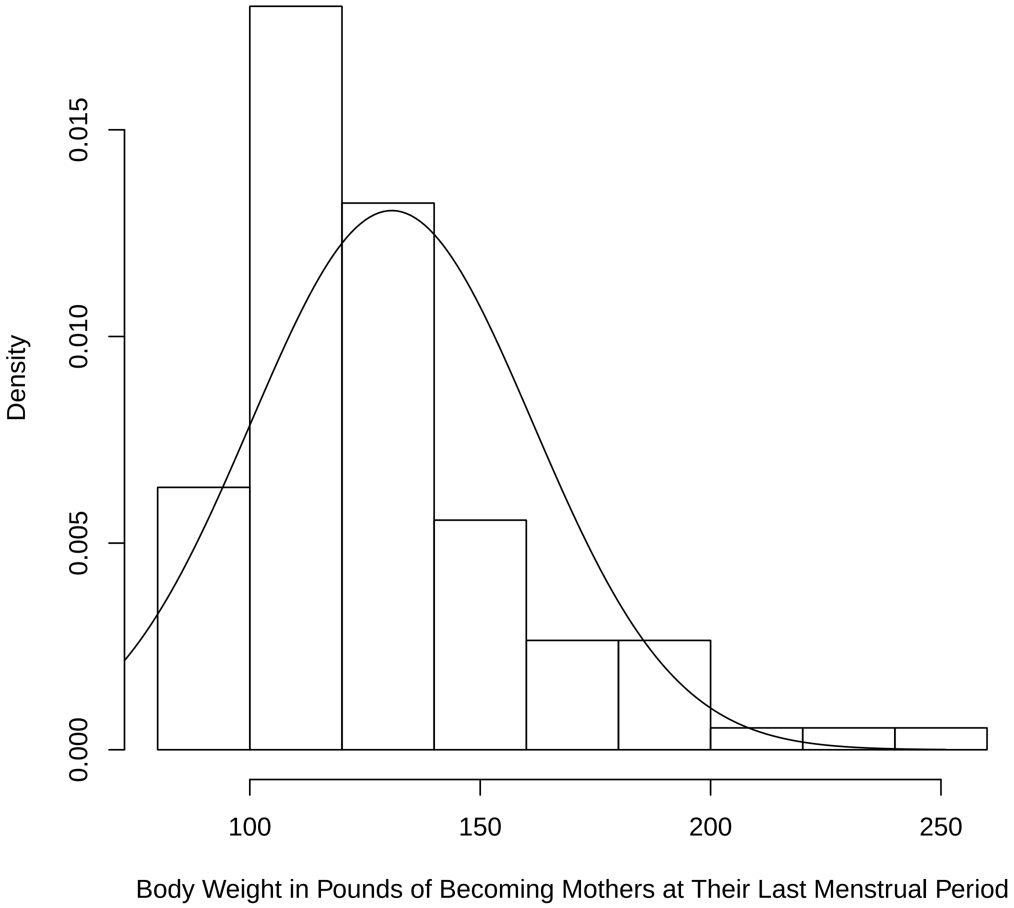


Figure 5: Histogram of Body Weight in Pounds of Becoming Mothers at Their Last Menstrual Period with Indicated Normal Distribution.

fit. In order to come to reliable conclusions the similarity can be tested. There are various significance tests to test the hypothesis H_0 : *sample distribution* $=$ *model distribution*. The significance test used in this paper is the Kolmogorov-Smirnov test (Sheskin 2000). It calculates the maximal distance between the cumulative density functions[13] of the sample and the model distribution using the following formula:

$$d_{max} = \left|S\left(X_i - F_0\left(X_i\right)\right)\right| \tag{9}$$

where $S(X_i)$ = cumulative frequency/n
and X_i the value of X at the position i
and F_0 the model distribution
The critical d-value for $p > 0.5$ is $1.22/\sqrt{n}$.

If the d-values are above the critical threshold the null hypothesis needs to be rejected,

[13] The cumulative density function (CDF) adds up the probability of each x_i from x_0 to x_n.

i.e. sample and model distribution significantly differ from each other. Certainly, the Komolgorov-Smirnov test is not the best test for all model distributions, yet, specialised significance tests such as the Shapiro-Wilk test (Shapiro & Wilk 1965) are restricted to certain distributions and therefore comparability cannot always be assured (cf. Formula 10). Additionally, the Shapiro-Wilk test is restricted to 5,000 items which means that the procedures shown in this paper cannot be repeated using larger corpora such as the British National Corpus.

The Shapiro-Wilk Test:

$$w = \left[\sum_{i=1}^{\lfloor 2/n \rfloor} a_{n-i+1} \left(y_{n-i+1} - y_i \right) \right] / \sum_{i=1}^{n} (y_i - \overline{y})^2 \tag{10}$$

where $\lfloor 2/n \rfloor$ = greatest integer of n/2

a_{n-i+1} = coefficients tabulated in Shapiro & Wilk 1965.

To illustrate the different model distributions, the following two plots (page 19) show all tested distributions. The formulae for the additional probability density, respectively probability mass functions,the Cauchy, Exponential, Poisson and Negative Binomial distribution, can be found in the Appendix starting from Page 85. The following parameters have been used: $\mu = 5, \sigma = 1, \lambda = 5, p = 0.5$ and $r = 5$.

The above mentioned example of the body weight of pregnant women is a clear cut case. Neither the Kolmogorov-Smirnov nor the Shapiro-Wilk test support the normality theory. In both cases the p-value is below the smallest value that can be calculated by R (p-value < 2.2e-16). In this case the next distribution is fitted and tested until an appropriate function is found. Unfortunately, it is not always possible to find a distribution that fits according to a significance test. In that case only a choice based on 'better than' can be made.

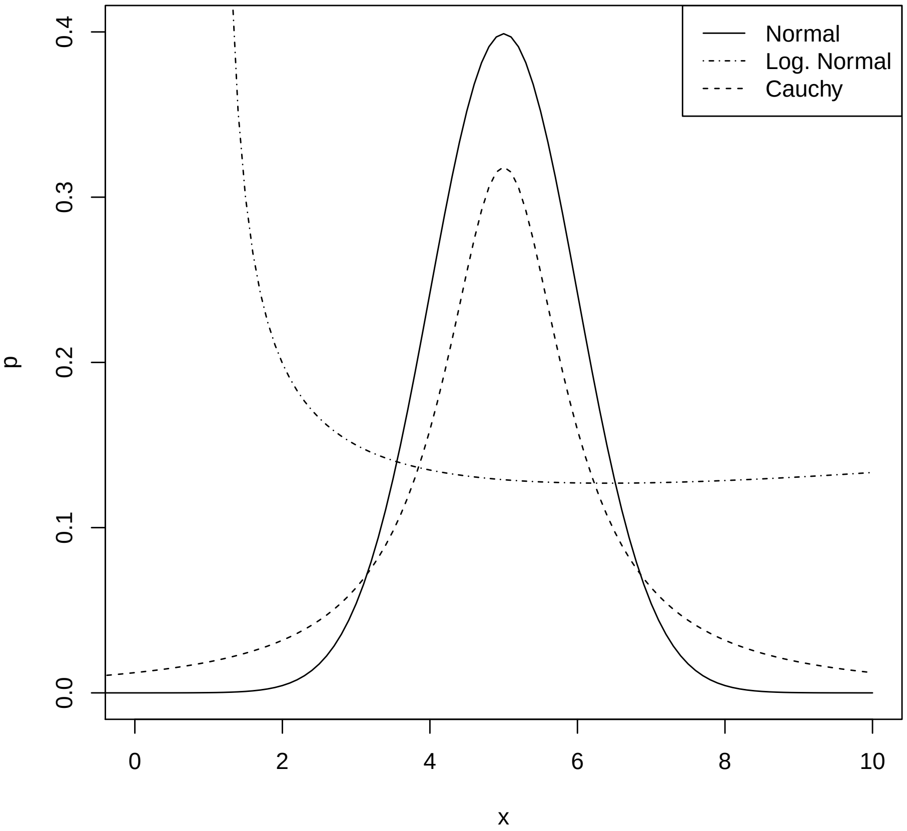


Figure 6: Normal, Log. Normal and Cauchy Example Distributions

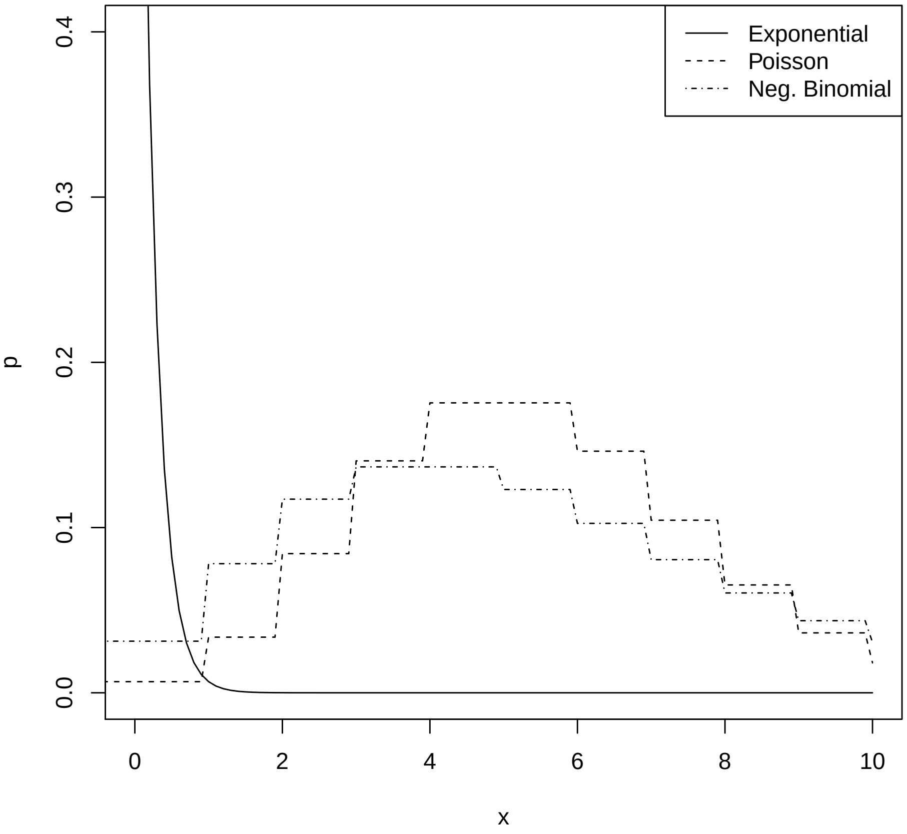


Figure 7: Exponential, Poisson and Neg. Binomial Example Distributions

3.4.3 Inferential Statistics

As mentioned in the introduction this master's thesis tries to find a way to predict the number of passive constructions per text. Predicting these constructions is not possible with the methods and procedures presented in the previous chapter.

The following example will illustrate the methodological restrictions of descriptive statistics. Let a χ^2 test show that group A uses significantly more passives per text than group B. The selection criterion for group A and B be characteristic c. This characteristic c be a metric variable. A significance test has revealed that on the basis of criterion c group A and B significantly differ. Depending on whether another group D fits to either group A or B, with the help of a χ^2 test one could say whether D is different from one of both groups. It still cannot be estimated, how many passives either of the three groups uses. Additionally, it remains unclear if another characteristic than c might have the most impact on the frequency of passives. Therefore, a method needs to be applied that estimates the number of passives per text without using predefined groups and that is at the same time able to integrate more than one variable.

Because the methods presented in this chapter are rather uncommon in linguistics, I will proceed by starting with an explanation of raw values and will develop the other procedures step by step. To illustrate the methods, the first 100 values of the text length (y), number of sentences (x_1), age (x_2) and months abroad (x_3) variables from the ICLE sample will be used. The text length functions as the dependent, or predicted variable and the other variables thereby function as independent, or predictor variables. The number of words per text is a discrete[14] count variable and thereby not the best variable for a linear model, yet it will be treated as a quasi-continuous variable for the sake of illustration (Wooldridge 2002, Tang, He & Tu 2012).

The basic idea behind a linear regression is to predict a dependent variable (y) on the basis of a number of independent variables (x_n) using a straight line. With one dependent and one independent variable the prediction becomes two dimensional and thereby optically intelligible. By plotting the independent variable on the y-axis and the dependent on the x-axis a scatter plot, as used in the descriptive statistics, appears. Pearson's correlation coefficient r (see Equation 7) can then be applied to determine how close the connection between the two variables is. In order to predict y, a straight line needs to be found that is closest to all points in the scatter plot. Straight lines can be formulated with a linear equation:

$$y = \beta_0 + \beta_1 x_1 \tag{11}$$

[14] Discrete variables have integer, i.e. whole numbered values.

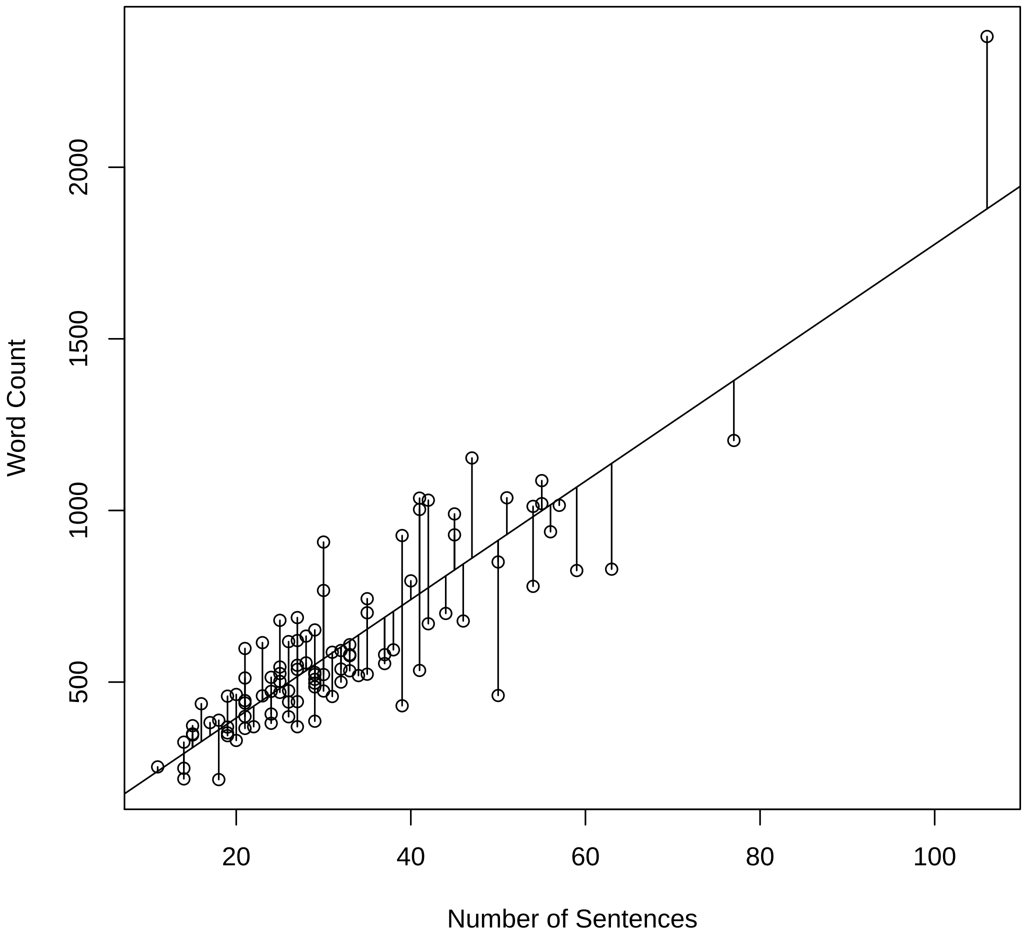


Figure 8: Scatterplot of the first 100 ICLE Text's Word Count and Number of Sentences with Indicated Model Function and Residuals.

Since only in a few rare cases a perfect prediction of y is possible, the model above has to be altered insofar, as it now includes a normal distributed noise term (ϵ) and y is fitted as $\hat{y}$:

$$\widehat{y} = \beta_0 + \beta_1 \cdot x \tag{12}$$

This estimation is based on the following model:

$$y = \beta_0 + \beta_1 \cdot x + \varepsilon \tag{13}$$

Picking up the words per text example from above, a scatter plot with indicated regression line, i.e. the straight line formulated in Equation 12, and residuals, i.e. the distance between the predicted and the actual line, is shown in the Figure 8 on the previous page.

In order to get a regression line the β_0 and the β_1 coefficients therefore need to be calculated.

$$\beta_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{14}$$

One might have noticed that this formula simply is the covariance of x and y divided by the variance of x (Wooldridge 2002, Bortz 2010). Calculating the intercept is easy once the β coefficient is known:

$$\beta_0 = \overline{y} - \beta_1 \cdot \overline{x}_1 \tag{15}$$

Since most models consist of more than one predictor variable another approach has to be chosen. Up to two predictor variables can be calculated with the following formulae:

$$\beta_1 = \frac{r_{y1} - r_{y2} \cdot r_{12}}{\sqrt{1 - r_{12}^2}} \cdot \frac{\sigma_y}{\sigma_1\sqrt{1 - r_{12}^2}} \tag{16}$$

$$\beta_2 = \frac{r_{y2} - r_{y1} \cdot r_{21}}{\sqrt{1 - r_{21}^1}} \cdot \frac{\sigma_y}{\sigma_2\sqrt{1 - r_{21}^2}} \tag{17}$$

Further variables require matrix algebra. First of all the values of the dependent variable are put into a vector. The same is done with the independent variables, i.e. the β coefficients and the ϵ terms, in an n times $k + 1$ sized matrix where n is the number of observations and k the number of variables.

$$y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}, x = \begin{pmatrix} 1 & x_1 1 & x_1 2 & x_1 3 & \dots & x_1 k \\ 1 & x_2 1 & x_2 2 & x_2 3 & \dots & x_2 k \\ & & & \vdots & & \\ 1 & x_n 1 & x_n 2 & x_n 3 & \dots & x_n k \end{pmatrix}, \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

With matrix algebra the linear Equation 12 can be reduced to $y = \beta x + \epsilon$ where $x\beta$ is the matrix vector product. The β coefficients are estimated by minimising the following linear Equations (Wooldridge 2002, 71f.):

$$\begin{gathered}
\textstyle\sum_{i=1}^{n}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_1 - \dots - \widehat{\beta}_k x_i k) = 0 \\
\textstyle\sum_{i=1}^{n} x_i 1(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_1 - \dots - \widehat{\beta}_k x_i k) = 0 \\
\textstyle\sum_{i=1}^{n} x_i 2(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_1 - \dots - \widehat{\beta}_k x_i k) = 0 \\
\vdots \\
\textstyle\sum_{i=1}^{n} x_i k(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_1 - \dots - \widehat{\beta}_k x_i k) = 0
\end{gathered}$$

With these calculations, a linear equation can be formulated with which the y values can be predicted. In a next step the model has to be evaluated, i.e. it needs to be estimated how far apart the predicted and the actual y values are and which of the independent variables is relevant for the prediction. The distance between the predicted $\hat{y}$ and the actual y values, i.e. the connection between the points and the regression line in Figure 8 are called residuals (e). The smaller e the better the model. The following five Equations 18 to 22 calculate: the β coefficients[15], the residuals e, the variance of the residuals σ_e^2, the variance of the β coefficients σ_β^2 or, if the square root is drawn, the standard error of the β coefficients, and the t statistics for which a p-value like for Pearson's χ^2 test can be given (Wooldridge 2002, Bortz 2010, Darlington 1968):

$$\beta = x'x^{-1}x'y \tag{18}$$

$$\varepsilon = y - \beta x \tag{19}$$

$$\sigma_e^2 = \frac{e'e}{n-k-1} \tag{20}$$

$$\sigma_\beta^2 = \sigma_e^2(x'x)^{-1} \tag{21}$$

$$\text{t-ratio} = \frac{\beta}{\sqrt{\sigma_\beta^2}} \tag{22}$$

The higher the t values the better, since a high t value results in a low p value which means that the respective variable has a significant influence. Insignificant independent variables can be excluded from the model unless needed for theoretical reasons, i.e. with regards to its content.

To evaluate the fit of the entire model two values can be used. The R^2 and the adjusted R^2 values express the amount of variance of y that can be explained on the basis of the linear model[16]. R^2 can be calculated by subtracting the quotient of the variance of the residuals and the variance of the independent variable from one. Multiplied by 100 it can

[15] x' stands for the transposed matrix of x.

[16] For non-linear models various pseudo R^2 values can be calculated. To maintain the comparability between linear and non-linear models theses pseudo R^2 will not be used. Instead another value is used which is calculated for all types of models on the basis of the same formula.

be read as percentage (Bortz 2010).

The second value to judge a regression model is the Akaike Information Criterion (AIC). For models with normal distributed residuals the AIC (Akaike 1974) can be calculated with the following formula (Burnham, Anderson & Diggle 2002, 63):

$$\mathrm{AIC} = n \cdot \log(\frac{\sum_{i=1}^{n} e_i^2}{n}) + 2 \cdot k \tag{23}$$

Yet, as the residuals of non-linear regression models are not necessarily normally distributed this formula only holds for linear models, which have qua definition normal distributed, independent, homoscedastic residuals (Bortz 2010, Wooldridge 2002, Burnham et al. 2002). For models without normally distributed residuals the logarithm of the likelihood is used to compute the AIC (Burnham et al. 2002, 62):

$$AIC = -2log(\mathcal{L}(\hat{\theta}|y)) + 2 \cdot k \tag{24}$$

The likelihood function $\mathcal{L}$ is dependent on the model type and therefore not specified at this point. For further information refer to McCullagh & Nelder (1972), Akaike (1974), Burnham et al. (2002) and Dobson (2002).

The AIC works on a basis of relative values (Burnham, Anderson & Huyvaert 2010), i.e. different models are compared by selecting the model with the lowest value as ”be[ing] ’closest’ to the unknown reality that generated the data” (Burnham et al. 2002, 62). Adding more independent variables to improve the model is being punished by the later part of the formula. In order to make the AIC more concrete the adjusted R^2 value for the linear regression will also be stated in the tables, so that a comparison of the different models can be started from the linear regression model. There are more sophisticated methods to compare different estimation models (Burnham 2010), but as the focus will lie on the characteristics of the best model and not on the process of finding this best model, only the AIC will be used.

As mentioned above, linear models are meant for normally distributed error terms. It may be that the number of passives per text fulfils these requirements by following a quasi continuous distribution, but chances are that it will not. Therefore, alternative models need to be formulated that allow for the analysis of count data.

The tool required for dependent variables following non-normal distributions is called generalised linear model (GLM) and has been developed by McCullagh & Nelder (1972). Explaining the different types of GLMs in a mathematical sense would go beyond the constraints of this linguistic master’s thesis. For the sake of completeness the procedure of a Poisson regression is outlined (Winkelmann 2000, 77). The following formula shows the logarithmic likelihood function of a Poisson regression:

$$\ell(\beta; y, x) = ln \prod_{i=1}^{n} f(y_i|x_i; \beta) \tag{25}$$

$$= \sum_{i=1}^{n} -e^{x_i'\beta} + y_i x_i'\beta - ln(y!) \tag{26}$$

To maximise this function to β, the partial derivative has to be solved with respect to β. Up to now the procedure is similar to that for a linear regression as explained above:

$$\frac{\partial \ell(\beta; y, x)}{\partial \beta} = \sum_{i=1}^{n} [y_i - e^{x_i'\beta}]x_i = 0 \tag{27}$$

Equation 27 can be solved using numerical techniques such as the Newton-Raphson method. Alternative methods for finding the respective maximum, such as proposed by McCullagh & Nelder (1989), can be used.

It has to be mentioned that the coefficients have to be interpreted differently for GLMs than for linear models. While for the linear model

$$\hat{y} = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon, \tag{28}$$

for the Poisson and negative binomial distributions

$$\hat{y} = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon) \tag{29}$$

or

$$lg(\hat{y}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \epsilon. \tag{30}$$

The Poisson and the negative binomial regression only differ in the additional dispersion parameter of the negative binomial regression, which is of no importance at this point (Dobson 2002).

Since we do not only want to know which variables influence the frequency of the passive voice in NNS writing but also to which degree, the relative importance of the single independent variables needs to be estimated. Blinder and Oaxaca have developed a decomposition for linear models that estimates how strongly a single variable influences the discrepancy between two groups (Jann 2008). Bauer & Sinning (2008) advanced this procedure to make it available for non-linear regression models. Grömping (2006) introduces Lindeman's, Merenda's and Gold's R^2 decomposition that returns the partial R^2 value of each independent variable. Since the Blinder-Oaxaca decomposition is only available for STATA whereas Grömping (2006) adopted Lindeman's, Merenda's and Gold's R^2 decomposition for R, the later will be used. The single partial R^2 values returned by the R function add up to the total R^2 and thereby allow a judgement about the relative importance of an independent variable.

The last part of this methodology section hints on how to read the parameter values. In the case of a linear regression the $\hat{y}$ values can directly be read as estimated frequency of a certain construction per text under given circumstances. These circumstances will be explained further down. In case of Poisson and negative binomial regressions there are two possibilities to interpret the coefficients. Either the model is interpreted as $lg(\hat{y}) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$ or as $\hat{y} = exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)$ (see eg. Winkelmann 2000).

The following example shall illustrate this procedure. If $\hat{y} = 5$ for the coefficient combination of a certain text, if the model allows a perfect prediction for the present perfect passive and if a linear regression is used, then there are five present perfect passives in that text. If instead of the linear regression model a Poisson regression model is used and the combination of the β coefficients add up to 5, there are 148.413 present perfect passives in that text. Therefore, the β coefficients are comparatively small for Poisson and negative binomial regressions.

Some of the independent variables cannot rise or drop since they are categorical variables such as gender or exam. In that case, one of the possible values is chosen to be the default[17]. For the equations the default can be regarded as zero. All other values are evaluated according to the β coefficients. If $\beta = 2$ for the exam variable, $\hat{y}/lg(\hat{y})$ rises by two when the text was written under exam conditions (Winkelmann 2000, 72). For the L1 variable, which has been dummy coded[18] the same holds true if a specific L1 is compared to the default.

If the L1 variable was treated like the other categorical variables, only a comparison between the different L1s and the default would have been possible, but not between the single L1s. Thus, the R package *factorplot* will be used (Armstrong 2014). It automatically swaps the defaults, so that a comparison between all available L1 combinations is possible. The results are presented in a matrix that contains the β and standard error differences. To enhance the legibility of the matrix, significant differences between the

[17] I have selected 'male' and 'no' to be the defaults in the upcoming models. The non-defaults are given in brackets behind the variable names.

[18] The number of L1 be k. So, k variables are created where each L1 has its own variable. Each variable is coded with one at the position x_i if $x_i =$ L1 or zero if $x_i \neq$ L1.

L1s are coded with different colours.

The regression type will always be stated underneath the tables. The R codes for the linear models and for the generalised linear models can be found in the appendix.

3.5 The Corpora

In order to realistically describe the usage of the passive by NNS EFL learners, realistic texts of these learners have to be found. Furthermore, additional and detailed information about the speakers is required in order to run the regressions. A text corpus can fulfil both of these requirements. It usually consists of a collection of texts written by a specific kind of speaker and of a database describing these speakers.

There are plenty of text corpora that could be used. To assure comparability only one corpus will be used. Therefore, the corpus needs to be large enough to allow for a statistical analysis[19]. Only three corpora fulfil these requirements.

The International Corpus Network of Asian Learners of English (ICNALE) is a recent corpus consisting of 2,600 NNS and 200 NS texts with about 1.3 Million tokens (Ishikawa 2013). These 2,600 NNS speakers come from 10 different Asian countries and speak English as their second or as a foreign language. The text corpus is accompanied by a large database with information on the students and their performance in English. Since the corpus is only focussed on Asian countries and my knowledge of their grammatical structures and languages is too limited, it will not be included into this master's thesis.

In the European framework, Sylviane Granger and the University of Louvain are pioneers in EFL corpus linguistics. Her latest corpus project, in which the University of Hanover also participates, is the Longitudinal Database of Learner English (LONGDALE). The first data is available for researchers within the universities participating in the project, yet as it is longitudinal and not yet accessible for public use, it will not be used either.

Being one of the oldest EFL corpora the International Corpus of Learner English (ICLE) seems most appropriate. The first version was published in 2002 (Granger 2003). In 2009 an extended second version followed. This second version consists of 6085 texts with 17 L1 and 3,753,030 tokens, which will be the basis for the analysis of the passive in this paper. To avoid redundancy, the variables available in ICLE are addressed in Chapter 3.7.

[19] There are plenty of smaller corpora but as there are no statistical benchmarks available I want to base my analyses on a basis as broad as possible.

3.6 The Data Collection Process

To use a passive construction as the dependent variable in an estimation model, the frequency of these constructions per text needs to be known. Because the reduced ICLE corpus which only contains texts with complete profiles, still consists of 2,720,365 words, it is too error-prone and time consuming to count and classify all the different tense, aspect, voice and mode constructions. Therefore, a computer aided system is required. In a first step, grammatical information is added to each single text using the Stanford Natural Language Processing parser, as it is implemented in Erwin Komen's Cesax programme (see http://erwinkomen.ruhosting.nl). Reformatted to xml files, the output can be analysed in a second step, where BaseX, an X-Query tool, is applied to search and count the different constructions per text. The following two sub-sections will guide through the single steps of this procedure.

3.6.1 Stanford NLP

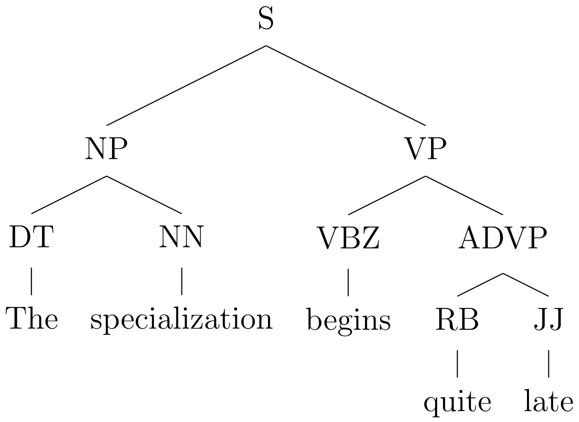
The texts of the second version of the ICLE corpus are delivered as plain 'txt' files. These text files do not yet contain any syntactical information which is needed to automatically search for grammatical constructions. In order to include these information, a linguistic parser can be used. One of the most common tools is the Stanford Natural Language Processing (Stanford NLP) parser (Manning 1999). Its advantages are that it can be obtained free of charge and that it is open source, i.e. the source codes are freely available and thereby can be adjusted to specific needs of the research project. Yet, it is not possible to save the parsed output as 'xml' files, which are required for an automatic analysis. Erwin Komen, a programmer associated with the Radboud Universiteit Nijmegen, the Netherlands, created the Cesax tool, which is able to save this parsed output in a non-plain text file format. Yet, Cesax' output is stored either in the 'psd' or the 'psdx' file format. 'psd' is the proprietary file format of Adobe's Photoshop. Opening these files with a simple text editor reveals that they are actually 'xml' files with a changed file ending. Therefore, simply renaming the files (see Source Code 1) allows for an analysis with an X-Query tool.

Source Code 1: DOS Code to Replace File Endings

```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation.
Alle Rechte vorbehalten.

D:\ICLE\parsed> rename *.psdx *.xml
```

These 'xml' files can easily be opened with any web browser or 'xml' tool. 'xml' is a language meant for computers and is therefore hardly readable. The main clause of the fifth sentence of the ICLE text BGSU1002 shall function as an example in this and the following chapter. The clause reads as follows: "The specialization begins quite late, [...]". Although this main clause consists only of five words and a comma, the corresponding 'xml' code is more than one page long. The actual code can be found in the Appendix (Source Code 5 on page 70). The following tree diagram shows the same code reformatted into a tree diagram.



The following chapter will show how search algorithms formulated in X-Query can be used to convert these tree diagrams into numbers.

3.6.2 X-Query

The addition of syntactical information to a text does not lead to new insights since adding these tags is merely time consuming and error prone. By teaching the computer how to interpret these tags the analytical process can be automatised and thereby fastened.

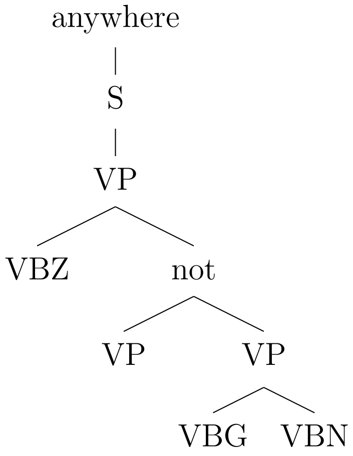
Since a text corpus is, in terms of its structure, not very different from any other kind of database, it can be searched with the corresponding tools. The developers of the xml standard, with which the texts are stored during their parsing, analogously created a query language (see www.w3c.org). The latest xml and X-Query standards the "XQuery and XPath Data Model 3.0" can be obtained from the W3C website (http://www.w3.org/TR/2014/REC-xpath-datamodel-30-20140408/). X-Query works on a user oriented level with FLWOR expressions. FLWOR stands for 'for', 'let', 'where', 'order by' and 'return' and is thereby similar to function definitions in programming languages such as Java or R. To correctly count the main clause of the example of the last chapter as present simple active third person singular, the following bits of code are required (Source Code 2).

Since 'xml' is a hierarchical language and therefore looks like the syntactical tree shown in the previous chapter, X-Query runs through the different nodes and returns all nodes that fit the definition. Thus, the first line tells X-Query to search in all files of the database for sentences that fit the definition given in line two to six. Instead of returning all findings, i.e. each sentence formulated in the third person singular present simple indicative active, X-Query is told to return only the base-uniform resource identifier or short base-uri.

Source Code 2: Present Simple Active (3^{rd} Person Singular)

```
1 for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
2 where $d/eTree/@Label=”VBZ”
3 and not($d/eTree/@Label=”VP”)
4 and not($d/eTree/eTree/@Label=”VP”)
5 and not($d/eTree/eTree/eTree/@Label=”VBG”)
6 and not($d/eTree/eTree/eTree/@Label=”VBN”)
7 return base−uri($d)
```

With the definition between line two and six, X-Query creates its own tree and returns the address per hit, which is, in this case, the name of the file. This tree looks similar to the one in the last chapter except for the fact that tags are used and not specific words. Moreover, this tree partly consists of a negative definition. This negative definition is required to prevent X-Query from mistakenly returning more complex constructions than actually searched. Sometimes it is necessary to specify certain words in order to correctly define a grammatical construction. Accordingly another eLeaf child node is included into the definition.



The result of this query is a list containing the file name per hit. This means if the searched construction appears two times in text one then BaseX returns: ”text one”, ”text on”. Because the number of specific constructions per text are required these results need to be transformed. To do so the results are copied to a spreadsheet software like Microsoft’s Excel or Open Office Calc. The results need to be compared with a list of all the files. Consequently, a list of files needs to be created using Microsoft’s CMD tool:

Source Code 3: Creating a File List

```
Microsoft Windows [Version 6.1.7601]
Copyright (c) 2009 Microsoft Corporation.
Alle Rechte vorbehalten.

D:\ICLE\parsed> dir \B > file_list.txt
```

\B stands for 'bare' meaning that only the file name is returned and no further information, such as file size etc..

To compare both lists the $= \mathit{COUNTIF}(\mathit{searched\ term}, \mathit{search\ term})$ function of either Excel or Calc can be used. A thereby created list of constructions per text can be loaded into R for further analysis.

Since there is no tool or algorithm available to fulfil the task needed for this paper, I had to programme it by myself. All the algorithms can be found in the Appendix.

To assure the correctness, I have controlled the first ten texts finding only three incorrect forms. These incorrect forms are due to wrongly tagged words. Especially NNS writing and texts with many errors and mistakes cause wrongly assigned tags (Manning 1999).

3.6.3 KoRpus

Some of the independent variables (see Chapter 3.7) have been computed using R's KoRpus package (Michalke 2014). This package analyses the single ICLE texts and returns specific linguistic values such as the number of sentences, the average length of them and some more. Problems arise once the texts contain non-ASCII characters like accents or language specific characters likes the German 'ö' or the Danish 'ø'. The KoRpus package cannot read these characters and returns an error. Since there are 757 files containing special characters, this problem cannot be neglected. The following Perl script removes all non-ASCII characters from the original files and saves them in a new folder. Mostly there is only one character removed, so the corruption of the results is reduced to a minimum.

Source Code 4: Replacing non-ASCII Characters

```
perl -pe's/[[:^ascii:]]//g' < /.../ICLE_id.txt > /.../
    ICLE_id_cleaned.txt
```

3.7 The Independent Variables

The independent variables belong to three categories. The first category are variables which give further information on the administrative environment in which the texts have been written. The second category provides social information on the students who have written the texts. The third category is made up of variables with linguistic content. A detailed description of the parameter values will be given in Chapter 4.1.

For the model estimation the following variables are used[20]:

Variables	
Administrative Variables	
Type	Conditions
Reference Tools	Exam
Social Variables	
Age	Sex
1^{st} Language	Languages at Home
English at School	English at University
Months Abroad	N other Languages
Linguistic Variables	
N Words	N Sentences
Avg. Words/Sentence	Flesch Reading Ease Score

Table 2: Variables.

3.7.1 Administrative Variables

The ICLE database provides five variables of the administrative type. Four of them have been used. The variable 'Exam' is a binary variable with the possible parameter values 'yes' and 'no'. It states whether the text was graded as an examination or not. The binary variable 'Reference Tools' states whether the students were allowed to use dictionaries or not. The variable 'Timing' expresses whether there was a time limit for writing the essays. To a certain extent this variable remains dubious since 'no' cannot mean that there is no time limit at all. With the parameter values 'argumentative' or 'literary' the 'Type' variables shows the text type. Depending on the text type the 'Title' variable gives the title of the essays. Possible titles have mostly been given by the ICLE organisers. Unfortunately, there are too many different formulations of similar titles to include the information into the analysis.

[20] Caused by the different levels of measurement some of the variables had be recoded. Therefore, the number of variables displayed with the different estimation models is higher than the number of variables shown in this table.

3.7.2 Social Variables

There are eight variables giving further social information on the students. Two of them are on a nominal and after reformatting six are on a metric scale. The 'Sex'[21] variable is coded with 'female' and 'male'. In the original data base the 'First Language' variable is coded on a nominal scale. In order to prepare the data base for the multivariate estimation models the 'First Language' variable has been recoded using dummy variables[22]. The 'Age' variable is a metric variable giving the age of the students in years. The 'Home Language' refers to the languages spoken by the students at home. Since I am not particularly interested in the specific languages, but only in the fact if and how many further languages the students speak at home, the variable has been recoded with possible values ranging from zero, for no further languages spoken at home, to three standing for three additional languages[23]. The additionally learnt languages are treated in the same way[24]. The remaining three metric variables refer to the length of the student's language learning history. They include the time the student has spent learning English at school and at university in years and the time she[25] has spent abroad in months.

3.7.3 Linguistic Variables

There are six linguistic variables included in the data base of which one has been provided by the ICLE organisers and the other five have been added later on. The length of the texts in words is given by the original data base. The other five have been computed using the *KoRpus* R package (Michalke 2014) or have been calculated by hand. The first variable is the number of sentences where a sentence is defined as a string of words terminated by '.', '!','?' and ';'. The second variable is the average number of words per sentence which is calculated by dividing the number of words by the number of sentences. The other three variables are meant to measure the complexity of the texts. They have been developed by Rudolf Flesch and J. Peter Kincaid. The Flesch-Kincaid Grade Level measures the complexity of a text and returns the approximated grade in which a student should be able to write such a text (Flesch 1948). Because Flesch and Kincaid worked in the United States, these grades refer to the American educational system. In order to make this grade score legible for non-U.S. readers it has been recalculated to the Flesch

[21] The ICLE data base calls this variable 'sex' referring to the biological sex. I will stick to this terminology being aware of the fact that this terminology is debated.

[22] Dummy coding means that for each parameter value a new binary variable is created where the respective original value, in this case the first language, is coded with one and all others values are coded with zero.

[23] The performance of the various languages remains questionable if three additional languages are spoken at home.

[24] The respective scripts can be found in the Appendix.

[25] Despite using only the female gender to assure readability, all genders are meant without the intent of discriminating either of them.

Reading Ease Score which can range from 0 to 100. The following formulae have been used:

Flesch-Kincaid Grade:

$$\text{Flesch-Kincaid Grade} = 0.39\frac{N\ words}{N\ sentences} + 11.8\frac{N\ syllables}{N\ words} - 15.59 \tag{31}$$

Flesch-Kincaid Grade solved for N syllables:

$$\text{N syllables} = \frac{(Flesch - Kincaid\ Grade - 0.39\frac{N\ words}{N\ sentences} + 15.59) \cdot N\ words}{11.8} \tag{32}$$

Flesch Reading Ease Score:

$$\text{Flesch Reading Ease Score} = 206.835 - 1.015\frac{N\ words}{N\ sentences} - 84.6\frac{N\ syllables}{N\ words} \tag{33}$$

As can be seen from the formulae, the Flesch-Kincaid Grade and the Flesch Reading Ease Score only differ in respect to the constants. These constants also change the direction of the scores. While a higher figure, in case of the Flesch-Kincaid Grade, stands for a higher complexity, i.e. more words per sentence and longer words, the Flesch Reading Ease Score works the other way around with a lower score standing for a higher complexity.

4 Results

4.1 Descriptive Statistics

4.1.1 Independent Variables

The mathematical foundation for this chapter has been established in Chapter 3.4.1. The respective formulae can be found there. The following results refer to the reduced sample, i.e. the complete cases of the ICLE data base (N=4424). In order to maintain legibility, the variables are presented according to the systematisation outlined in Chapter 3.7.

Timing		Exam		Reference Tools		Type	
Yes	No	Yes	No	Yes	No	Argumentative	Literary
1576	2848	1462	2962	2456	1968	4204	220
χ^2-value	p-value	χ^2-value	p-value	χ^2-value	p-value	χ^2-value	p-value
365.7288	≈0	508.5895	≈0	53.83	≈0	3587.761	≈0

Table 3: Descriptive Statistics of Administrative Variables.

All of the administrative variables are binary variables. 'Timing', 'exam' and 'reference tools' can each have the parameter value 'yes' or 'no'. 'Type' is coded either with 'argumentative' or 'literary'. The distribution of each variable is initially shown and then tested for a significant imbalance.

As can be seen in Table 3, the administrative variables are unevenly distributed. The χ^2-test supports this observation by stating that the differences in absolute frequencies are significant. Only the usage of reference tools is an exception. In the unreduced sample, i.e. the original ICLE data base, the usage of reference tools is evenly distributed (χ^2=0.3447, p=0.5571).

In contrast to the administrative independent variables the social variables are, concerning their structure and level of measurement, more diverse.

'Sex' is the simplest of the eight variables. It can either have the parameter value 'female' or 'male'. The sample contains 3342 women and 1082 men, which is a significant imbalance (χ^2=1154.521, p≈0).

L1	Frequency	Diff. to ICLE db.	L1	Frequency	Diff. to ICLE db.
Bulgarian	287	13	Japanese	321	45
Chinese	118	50	Norwegian	250	66
Cantonese	780	34	Polish	294	72
Czech	168	73	Russian	150	124
Dutch	190	72	Spanish	132	118
Finnish	180	81	Swedish	296	176
French	244	70	Tswana	399	120
German	224	221	Turkish	257	19
Italian	134	264			

Table 4: Distribution of L1.

Like ’sex’ ’First Language’ is a nominally scaled variable. Yet, it can have 17 different values, i.e. there are 17 different mother tongue backgrounds. Even though the absolute frequencies already suggest an uneven distribution, a χ^2 test is computed. As expected, the χ^2-test supports this observation with a χ^2-value of 1470.097.

Table 4 also indicates the difference to the unreduced ICLE data base. Some German, Italian and Russian students seem to have had difficulties or other problems with properly filling in the questionnaire, since their number tremendously drops as the NAs are omitted.

From all students asked if they spoke another language at home than at university, 4376 (98.9%) answered that they speak the same language at home as at university, 23 (0.5%) are bilingual and 25 (0.6%) are trilingual[26].

1699 students or 38% of the sample have not learnt another language besides English at school[27]. 937 (21%) stated that they have learnt one further language at school. Two or even three additional languages are learnt by 1118 (25%) and 670 (15%) students.

Variable	Min	Max	Mean	Variance	Standard Deviation	Skew	Kurtosis
Age	17	71	21.896	13.318	3.649	4.163	32.014
Eng. at School	1	20	8.666	9.583	3.096	0.032	2.097
Eng. at Uni.	0	7	2.137	2.423	1.557	0.226	2.180
Months Abroad	0	216	2.585	86.122	9.280	9.288	128.848

Table 5: Descriptive Statistics of Metric Social Variables.

Unsurprisingly, the majority of the students is in their early twenties. Because of the high kurtosis and the relatively low skew, more than the usually expected 68% can be found within one standard deviation (σ). The mentioned skew and kurtosis hint towards a non-normal distribution of ’age’.

With the age ranging between 17 and 71 years, most students have had between five and eleven years of English at school. The maximum of twenty years of English at school looks like a typo, yet, there are three students claiming to have had those twenty years of English and another eight claim to have had between sixteen and nineteen years. Perhaps these students originate from countries where English is taught from Kindergarten onwards. Otherwise such a long history of institutionalised English classes cannot be explained. The kurtosis is slightly lower than expected for a normal distribution.

Moreover, the students were asked for how long they already have studied English at university. They were allowed to answer in half-year steps. 855 stated that they have just started studying. Therefore, the mean has moved towards the left and now lies around the beginning of the third year. Because of the small range of the variable, the standard

[26] The performance in each further language has not been tested.

[27] ’None’ obviously means in this case L1 and English.

deviation is also comparatively small. Skew and kurtosis are similar to the previous variable.

Since the majority of students is in their early twenties and has just started studying, the time they have spent abroad is rather short. Hence, 2793 (63%) have not been abroad yet. Due to the extreme outliers the distribution has a high positive skew and is extremely steep (kurtosis=128.85).

Variable	Normal (W)	Cauchy (D)	Exponential (D)	Log. Normal (D)	Poisson (D)
Age	0.669	0.981	1.000	0.998	0.287
English at School	0.948	0.903	0.962	0.898	0.154
English at University	0.933	0.535	0.467	0.358	0.252
Months Abroad	0.277	0.500	0.631	0.631	0.615

Table 6: Test Statistics of Metric Social Variables.

Table 6 shows the test statistics of the four metric social variables. In order to test for normal distribution, the Shapiro-Wilk test has been used, whereas the tests for all other distributions have been conducted using the Kolmogorov-Smirnov test (see Chapter 3.4). In each case the p-value is zero and has therefore been omitted. Instead of the p-values the respective test statistics are provided. The Poisson distribution has the best fit for the first three variables and the distribution of 'Months Abroad' comes closest to a normal distribution. Yet, as it has been shown in Table 5 more than half have not been abroad and with about 129 the kurtosis is far too steep for a normal distribution. Therefore, I assume that the variable 'Months Abroad' does not follow a standard distribution.

Figure 9 (page 38) and Table 7 give insights into the inter-relation of the single social variables. Instead of plotting each metric social variable combination in a single graph, a so called scatter-plot is used. It is set up like a matrix with the variable names on the

	Age	Eng. at School	Eng. at Uni.	Months Abroad
Age	1	-0.110	0.230	0.183
Eng. at School	-0.110	1	-0.571	-0.003
Eng. at Uni.	0.230	-0.571	1	0.019
Months Abroad	0.183	-0.003	0.019	1

Table 7: Correlation Matrix of Metric Independent Social Variables.

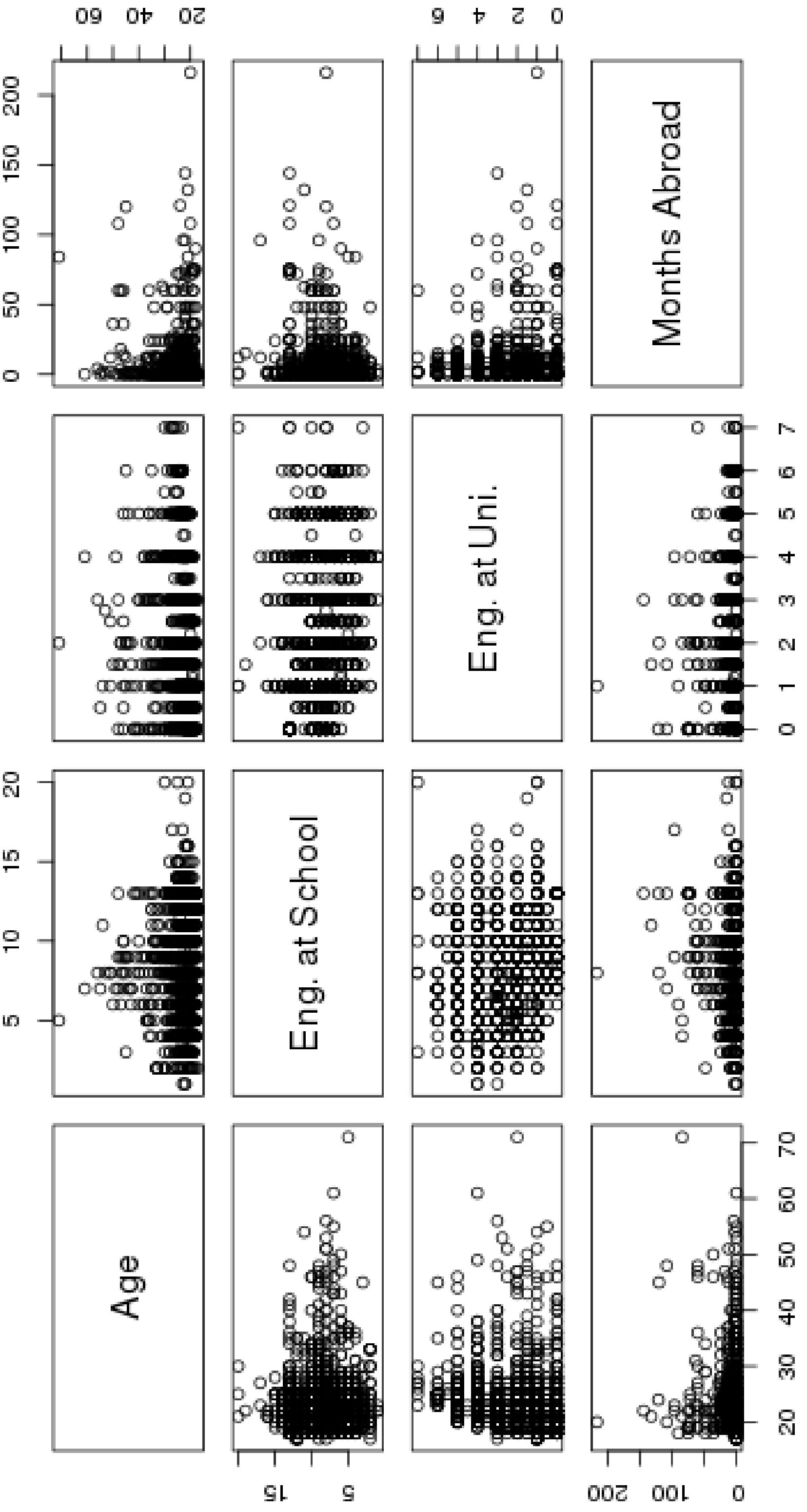


Figure 9: Scatterplot Matrices of Metric Independent Social Variables.

diagonal axis. The better the single points form a diagonal line, the higher the connection, i.e. correlation, between the two variables. If the points form a cloud or curve that is bent too far, there is little to no correlation. In order to support this visual impression, the correlation coefficient has also been calculated.

Judging from the scatter-plot, there seems to be no correlation between the variables. The points either form a cloud or cumulate in one of the corners or sides of the plot. The correlation matrix in Table 7 (page 37) supports this first impression. The only slightly stronger connection can be found between the years of English at school and the years of studying English at university. The only explanation imaginable for this correlation is that the essays have been written in an applied linguistics course and that students with less English at school tend to follow these courses earlier in their studies.

Variable	Min	Max	Mean/Median	Variance	Standard Deviation	Skew	Kurtosis
Length	92	4, 139	614.911(μ)	69, 635.410	263.885	2.720	20.991
N. Sentences	3	182	35.523(μ)	285.296	16.891	1.964	11.064
Avg. Sentence Length	6.702	164.333	18.352(μ)	30.165	5.492	6.509	132.788
Flesch-Kincaid Grade	3.363	14, 402	9.84985(m)	260, 086.000	509.986	20.754	451.769
F.-K. Reading Ease	-103, 101.400	86.189	56.23077(m)	13, 367, 701.000	3, 656.187	-20.753	451.690

Table 8: Descriptive Statistics of Linguistic Variables.

Next to the social variables listed and explained above, there are five linguistic variables which have to be examined in more detail. Because absolute frequencies are used instead of relative frequencies text length is one of the most obvious linguistic variables. Although a limit of 500 to 700 words was given for the majority of students, with mean of about 614 words implying that many students followed this restriction, the text length ranges from 92 to 4,139 words. Nonetheless, the small standard deviation and the high kurtosis imply that the vast majority obeyed the rule.

The number of sentences per text is the next linguistic variable. As defined in Chapter 3.7, the number of sentences ranges from three to 182 with a mean of 35.5. Because of the natural correlation between the number if words and sentences a word limit also directly influences the number of sentences. Consequently, the number of sentences has a high kurtosis of 11.06, as can be seen in Table 8.

The Flesch-Kincaid Reading Ease Score and the Flesch-Kincaid Grade evaluate the complexity of a text on the basis of the number of words per sentence and the number of syllabi per word. Several constants adjust these scores to a reference system. The Reading Ease Score is defined between 100 and zero, with 100 being on the easy end of the continuum. In order to make this score more intelligible, it can be transformed to a grade level, stating the minimum required grade to understand the text. Since Flesch and Kincaid invented this score for the U.S. military, the grade level refers to the American educational system (Flesch 1948).

The number of words per sentence and, theoretically, also the number of syllabi per words are not restricted, therefore it is imaginable that scores go beyond the above mentioned boundaries. Mistakes made by the students can also change the score. Yet, as I have not corrected wrongly used tenses or typos, which can cause a wrong identification, wrong punctuation will neither be corrected. Unfortunately, this results in some outliers being included in the summary statistics in Table 8. For the Flesch-Kincaid Grade and Reading Ease Score the outlier robust median, which will be marked with 'm', will be given instead of the outlier influenced mean, which will be marked with 'μ'. The most extreme example for an outlier is JPWA1001, who uses no punctuation at all. I might have excluded this text right from the beginning, but since the transition from non-outlier to outlier is smooth a clear boundary between both cannot be defined.

Despite these difficulties, both scores show the expected linguistic behaviour. The NNS students of English are, according to the Flesch-Kincaid scores, about six to seven years behind NS students in terms of their English proficiency.

Variable	Normal	Cauchy	Exponential	Log. Normal	Poisson
Length	0.822	0.997	1	1.000	0.517
N. Sentences	0.873	0.958	0.998	0.979	0.249
Avg. Sentence Length	0.741	0.963	0.999	0.984	0.105
Flesch-Kincaid Grade	0.970	0.930	0.987	0.941	0.149
Flesch-Kincaid Reading Ease	0.834	0.985	1.000	0.998	0.098

Table 9: Test Statistics for Metric Linguistic Variables.

Like the social variables, the linguistic variables show the best fit with a Poisson distribution. Yet, the mean of the Flesch-Kincaid Reading Ease Score (56.42) is smaller than its variance (143.4). This means that the Reading Ease Score is over-dispersed and thus cannot follow a Poisson distribution. Consequently, the fit for a negative binomial distribution is better (Kolmogorov-Smirnov D = 0.0627), but still significant (p = 1.665e-15).

The scatterplot and the correlation matrix show the linguistically expected results. Since the number of syllabi per word is morphologically restricted, there is a high correlation between the number of syllabi and the number of words per text. Also the number of words per sentence seems to have a natural boundary. Therefore, the number of words and sentences per text also has a high correlation. Because the sample fits these expectations the scatterplot (Figure 16) and the correlation matrix (Table 23) have been moved to the Appendix (page 86 and 87).

4.1.2 Tenses and Aspects

The easiest and most accessible way to understand the distribution of a variable is by plotting the histogram. Problems arise as soon as there are more variables than possibly can be plotted without exceeding the spacial boundaries of a paper. Therefore, the quantitative description of tenses and aspects will be achieved using tables instead of graphs.

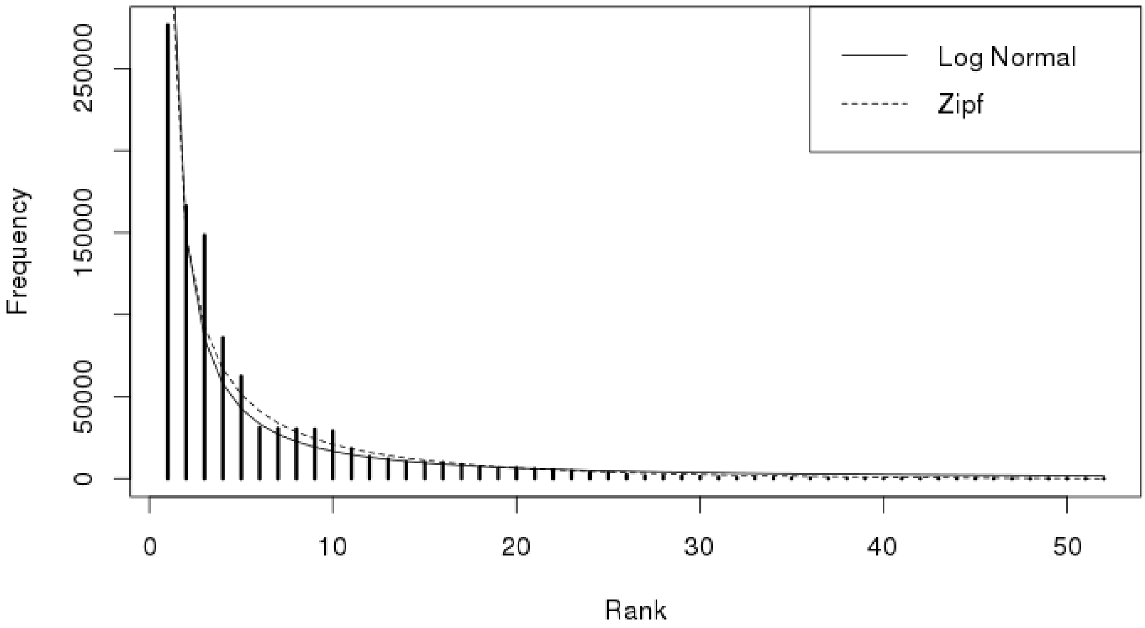


Figure 10: Observed Frequency of Tense and Aspects and theoretical log normal and Zipf values.

To get an impression of the quantitative differences, the sum of the tenses and aspects will be plotted. Figure 10 shows the tenses and aspects sorted by their rank on the x-axis and their frequency within the corpus on the y-axis. As can be seen, the most frequent construction is almost twice as frequent as the second most frequent construction. After around the 25^{th} rank the frequency approximates the x-axis.

This behaviour does not come as a surprise, since the underlying principles have been described for the distribution of words within natural language. Therefore, they can be assumed to hold true for the distribution of grammatical constructions. Newman (2004), Evert & Baroni (2007) and Briscoe (2008) have shown that this principle, called Zipf's Law, shapes the relation between rank and frequency of many observable phenomena, including the distribution of tenses and aspects[28]. Zipf's Law is mentioned and fitted at

[28] Zipf's Law is frequently used by biologists to show the quantitative relation between different beings

this point because it allows for a coverage of the vast majority of cases by simply analysing the first couple of grammatical constructions.

As has been shown in the preceding chapter (Table 8, page 39), the average ICLE text consists of around 35 sentences. In order to formulate a grammatically correct sentence, at least one inflected verb is required. On average 47.7 inflected verb constructions are used per text. Of these 47.7 verb phrases 27.389 or 57.44% refer to the present in an active voice, while 4.872 or 10.21% refer to the past and 2.339 or 4.9% to the future. 5.127 or 10.74% are formulated in the conditional mode. The remaining 16.7% are formulated in the passive voice.

For each construction the mean is higher than the median. This indicates an uneven distribution, with extreme values pulling up the mean. In combination with skew and kurtosis the shape of the distribution becomes clearer. The skew of each variable is positive, which implies a distribution leaning to the left. The tremendously high kurtosis shows that the values cumulate around the mean and the distribution is far steeper than expected for a normal distribution. For the more frequent constructions the standard deviation is smaller than the mean, which corresponds to the high kurtosis. In those cases where the standard deviation (σ) exceeds the mean the frequency is so low that the construction does not appear in the vast majority of texts.

Taking into account that neither the Poisson nor the negative binomial regression are intended for dependent variables with a large amount of zeros, it makes sense to further analyse only those constructions that have a mean of above one. There are of course regression types for distributions with a high number of zeros, such as the zero-inflated or Hurdle models (Cameron & Trivedi 2005, 678f.), but applying them would go beyond the scope of this thesis.

Table 10 (page 43) shows the characteristics of the distribution of the tenses and aspects. The high skew and kurtosis suggest that they are not normally distributed, since they are too far off the parameter values which are typical for a normal distribution. In order to restrict the amount of models which need to be formulated and tested, the number of possible regression types has to be reduced. Significance tests can be used to evaluate whether there is a significant difference between the observed and the theoretically expected empirical distribution. The Shapiro-Wilk test (Shapiro & Wilk 1965, Shapiro, Wilk & Chen 1968) is only one of many tools to test for normal distribution. Yet, as this test can only be applied for a specific distribution, a non-parametric test needs to employed, as it returns comparable results for various kinds of empirical distributions. For this reason this thesis uses the Kolmogorov-Smirnov test (Sheskin 2000), which has already been described in Chapter 3.4.1. It searches for the maximal difference between

within an ecosystem. The corresponding *R* package to fit Zipf's Law is also programmed by biologists (Oksanen, Blanchet, Kindt, Legendre, Minchin, O'Hara, Simpson, Solymos, Stevens & Wagner 2013).

Tense and Aspect	sum	min	max	median	mean	variance	σ	skew	kurtosis
Present Simple 3ps	62,654	0	146	12	14.162	98.151	9.907	3.338	28.103
Present Simple non 3ps	45,656	0	58	9	10.320	60.444	7.775	0.897	4.277
Present Simple	108,310	0	194	22	24.482	209.018	14.457	2.178	15.225
Present Perfect Have	3,910	0	15	0	0.884	1.870	1.367	2.885	17.339
Present Perfect Has	3,343	0	11	0	0.756	1.534	1.239	2.630	12.992
Present Perfect	7,253	0	22	1	1.639	4.316	2.078	2.503	13.740
Present Progressive are	2,669	0	16	0	0.603	1.125	1.061	2.964	19.906
Present Progressive is	2,604	0	21	0	0.589	1.128	1.062	4.392	49.164
Present Progressive	5,273	0	26	1	1.192	2.663	1.632	2.918	22.343
Present Perfect Progressive Have	219	0	4	0	0.050	0.058	0.241	5.750	46.185
Present Perfect Progressive Has	153	0	3	0	0.035	0.039	0.198	6.356	49.772
Present Perfect Progressive	372	0	4	0	0.084	0.104	0.322	4.642	31.278
Present Active	121,208	0	227	25	27.398	240.711	15.515	2.270	17.238
Past Simple	20,465	0	133	3	4.626	44.594	6.678	4.566	48.009
Past Perfect	986	0	18	0	0.223	0.657	0.811	8.134	112.955
Past Progressive	55	0	2	0	0.012	0.013	0.113	9.268	92.091
Past Perfect Progressive	48	0	3	0	0.011	0.013	0.114	12.535	201.385
Past Active	21,554	0	148	3	4.872	51.359	7.166	4.808	53.349
Going to is	69	0	4	0	0.016	0.022	0.147	12.637	220.073
Going to are	390	0	4	0	0.088	0.115	0.339	4.610	28.818
Going to	459	0	7	0	0.104	0.146	0.382	5.189	46.267
Will Future Simple	9,244	0	21	1	2.090	7.127	2.670	2.291	10.258
Will Future Perfect	48	0	3	0	0.011	0.013	0.116	12.768	205.253
Will Future Progressive	136	0	3	0	0.031	0.038	0.195	7.171	61.926
Will Future Perfect Progressive	1	0	1	0	0.0002	0.0002	0.015	66.491	4,422.000
Future Active	10,347	0	21	1	2.339	8.114	2.848	2.157	9.479
Present Simple Passive Are	6,235	0	21	1	1.409	2.847	1.687	2.284	14.776
Present Simple Passive Is	7,093	0	22	1	1.603	3.970	1.992	2.607	15.246
Present Simple Passive	13,328	0	29	2	3.013	8.457	2.908	2.036	10.752
Present Perfect Passive have	721	0	10	0	0.163	0.223	0.473	4.803	54.875
Present Perfect Passive has	1,026	0	6	0	0.232	0.338	0.581	3.411	19.244
Present Perfect Passive	1,747	0	15	0	0.395	0.666	0.816	3.791	35.323
Present Progressive Passive are	179	0	4	0	0.040	0.051	0.227	6.987	67.060
Present Progressive Passive is	158	0	3	0	0.036	0.041	0.203	6.489	53.783
Present Progressive Passive	337	0	4	0	0.076	0.102	0.319	5.271	38.867
Present Perfect Progressive Passive has	2	0	2	0	0.0005	0.001	0.030	66.491	4,422.000
Present Passive	7,295	0	21	1	1.649	3.390	1.841	2.156	12.764
Past Simple Passive	4,651	0	41	0	1.051	4.344	2.084	5.125	54.086
Past Perfect Passive	1	0	1	0	0.0002	0.0002	0.015	66.491	4,422.000
Past Progressive Passive	0	0	0	0	0	0	0		
Past Perfect Progressive Passive	0	0	0	0	0	0	0		
Past Passive	4,652	0	41	0	1.052	4.346	2.085	5.122	54.030
Will Future Simple Passive	1,226	0	9	0	0.277	0.513	0.716	4.173	29.606
Will Future Perfect Passive	3	0	1	0	0.001	0.001	0.026	38.362	1,472.667
Will Future Progressive Passive	0	0	0	0	0	0	0		
Will Future Perfect Progressive.1	0	0	0	0	0	0	0		
Will Future Passive	1,229	0	9	0	0.278	0.514	0.717	4.164	29.511
Conditional Simple	21,969	0	46	4	4.966	14.960	3.868	1.494	7.801
Conditional Perfect	469	0	10	0	0.106	0.174	0.418	7.218	102.709
Conditional Progressive	237	0	4	0	0.054	0.064	0.254	5.808	47.857
Conditional Perfect Progressive	9	0	1	0	0.002	0.002	0.045	22.103	489.558
Conditional Active	22,684	0	49	4	5.127	15.588	3.948	1.518	8.356
Conditional Simple Passive	4,278	0	15	0	0.967	2.224	1.491	2.543	12.618
Conditional Perfect Passive	83	0	2	0	0.019	0.021	0.144	8.202	76.565
Conditional Progressive Passive	0	0	0	0	0	0	0		
Conditional Perfect Progressive Passive	0	0	0	0	0	0	0		
Conditional Passive	22,052	0	46	4	4.985	15.023	3.876	1.492	7.772

Table 10: Descriptive Statistics of Tense and Aspect.

the observed and expected distribution. The lower the found difference, the closer are the two distributions. The distance, Kolmogorov-Smirnov D, can, like the χ^2 values, be transformed into a p-value. In the case of tenses and aspects the p-value is always close

Tense and Aspect	Normal	Cauchy	Exponential	Log. Normal	Poisson	Neg. Binomial	Best Distribution
Present Simple 3ps	0.974	0.879	0.939	0.876	0.237	0.067	Neg. Binomial
Present Simple non 3ps	0.835	0.731	0.784	0.715	0.237	0.060	Neg. Binomial
Present Simple	0.992	0.917	0.973	0.933	0.267	0.043	Neg. Binomial
Present Perfect Have	0.500	0.500	0.533	0.533	0.413	0.443	Poisson
Present Perfect Has	0.500	0.500	0.586	0.586	0.470	0.458	Neg. Binomial
Present Perfect	0.500	0.500	0.358	0.358	0.194	0.332	Poisson
Present Progressive are	0.500	0.500	0.639	0.639	0.547	0.505	Normal
Present Progressive is	0.500	0.500	0.633	0.633	0.555	0.528	Normal
Present Progressive	0.500	0.500	0.436	0.436	0.304	0.397	Poisson
Present Perfect Progressive Have	0.500	0.500	0.955	0.955	0.952	0.698	Normal
Present Perfect Progressive Has	0.500	0.500	0.968	0.968	0.966	0.706	Normal
Present Perfect Progressive	0.500	0.500	0.927	0.927	0.919	0.679	Normal
Present Active	0.995	0.930	0.981	0.950	0.269	0.042	Neg. Binomial
Past Simple	0.658	0.567	0.508	0.399	0.338	0.217	Neg. Binomial
Past Perfect	0.500	0.500	0.868	0.868	0.800	0.515	Normal
Past Progressive	0.500	0.500	0.988	0.988	0.988	0.725	Normal
Past Perfect Progressive	0.500	0.500	0.990	0.990	0.989	0.705	Normal
Past Active	0.665	0.574	0.519	0.410	0.350	0.214	Neg. Binomial
Going to is	0.500	0.500	0.987	0.987	0.985	0.676	Normal
Going to are	0.500	0.500	0.926	0.926	0.916	0.664	Normal
Going to	0.500	0.500	0.915	0.915	0.901	0.651	Normal
Will Future Simple	0.512	0.500	0.330	0.330	0.206	0.285	Poisson
Will Future Perfect	0.500	0.500	0.990	0.990	0.989	0.696	Normal
Will Future Progressive	0.500	0.500	0.973	0.973	0.970	0.686	Normal
Will Future Perfect Progressive	0.500	0.500	1.000	1.000	1.000	0.732	Normal
Future Active	0.538	0.500	0.362	0.303	0.207	0.259	Poisson
Present Simple Passive Are	0.500	0.500	0.370	0.370	0.244	0.361	Poisson
Present Simple Passive Is	0.503	0.500	0.339	0.339	0.201	0.351	Poisson
Present Simple Passive	0.684	0.593	0.513	0.404	0.154	0.222	Poisson
Present Perfect Passive have	0.500	0.500	0.867	0.867	0.850	0.651	Normal
Present Perfect Passive has	0.500	0.500	0.824	0.824	0.793	0.620	Normal
Present Perfect Passive	0.500	0.500	0.733	0.733	0.674	0.568	Normal
Present Progressive Passive are	0.500	0.500	0.965	0.965	0.960	0.683	Normal
Present Progressive Passive is	0.500	0.500	0.967	0.967	0.965	0.705	Normal
Present Progressive Passive	0.500	0.500	0.936	0.936	0.927	0.662	Normal
Present Perfect Progressive Passive has	0.500	0.500	1.000	1.000	1.000	0.681	Normal
Present Passive	0.531	0.500	0.321	0.311	0.198	0.336	Poisson
Past Simple Passive	0.500	0.500	0.570	0.570	0.349	0.386	Poisson
Past Perfect Passive	0.500	0.500	1.000	1.000	1.000	0.732	Normal
Past Progressive Passive	0.500	0.500	1	1	1	0	Neg. Binomial
Past Perfect Progressive Passive	0.500	0.500	1	1	1	0	Neg. Binomial
Past Passive	0.500	0.500	0.570	0.570	0.349	0.386	Poisson
Will Future Simple Passive	0.500	0.500	0.812	0.812	0.758	0.568	Normal
Will Future Perfect Passive	0.500	0.500	0.999	0.999	0.999	0.740	Normal
Will Future Progressive Passive	0.500	0.500	1	1	1	0	Neg. Binomial
Will Future Perfect Progressive.1	0.500	0.500	1	1	1	0	Neg. Binomial
Will Future Passive	0.500	0.500	0.812	0.812	0.757	0.567	Normal
Conditional Simple	0.811	0.687	0.699	0.590	0.160	0.120	Neg. Binomial
Conditional Perfect	0.500	0.500	0.917	0.917	0.899	0.633	Normal
Conditional Progressive	0.500	0.500	0.952	0.952	0.948	0.698	Normal
Conditional Perfect Progressive	0.500	0.500	0.998	0.998	0.998	0.734	Normal
Conditional Active	0.820	0.696	0.708	0.599	0.158	0.116	Neg. Binomial
Conditional Simple Passive	0.500	0.500	0.531	0.531	0.380	0.405	Poisson
Conditional Perfect Passive	0.500	0.500	0.982	0.982	0.981	0.715	Normal
Conditional Progressive Passive	0.500	0.500	1	1	1	0	Neg. Binomial
Conditional Perfect Progressive Passive	0.500	0.500	1	1	1	0	Neg. Binomial
Conditional Passive	0.813	0.688	0.700	0.591	0.160	0.120	Neg. Binomial

Table 11: Kolmogorov-Smirnov D-value per Distribution per Tense and Aspect.

to zero. Therefore, Table 11 only indicates the Kolmogorov-Smirnov D values.

When comparing the different Kolmogorov-Smirnov D values it can be seen that the values 0, 0.5 and 1 appear more than once. The tenses and aspects where the normal and Cauchy distribution have a value of 0.5 and the other distributions a value of 1, are those which have absolute frequencies close to zero. The difference in values is due to the shape of the expected empirical distribution. The tenses and aspects previously selected for a deeper analysis show the expected behaviour which means that the Poisson distribution has the best fit.

4.2 Inferential Statistics

Based on the methods presented in Chapter 3, the following sections will illustrate the results of the statistical analysis. In a first step, Chapter 4.2.1 will evaluate which model fits best to describe the different passive constructions. Based on this selection process, the following chapters will give detailed insights into selected passive constructions. Due to the limitations of this paper, not all constructions can be discussed. The results for the remaining passive constructions can be found in the Appendix.

4.2.1 Model Selection

Chapter 2 has shown that there are few linguistic papers using more than descriptive statistics and none using regression models to evaluate the usage of the passive. Although theoretical considerations and the distribution of the dependent variables (see Chapter 4.1.2) suggest the employment of count data regression models, it is better to recheck these considerations in order to be on the safe side. For each tense and aspect a linear, Poisson, negative binomial type I and type II regression model has been formulated and the corresponding AIC have been calculated. To have a more intuitive measure, the adjusted R^2 values have also been calculated.

Table 12 shows the results of these calculations. The first column contains the adjusted R^2 values, which can, multiplied with 100, be read as the share of variance of the dependent variable, explained on the basis of the linear model in per cent. The columns two to five indicate the AIC values for the different model types. The last column shows the name of the best model based on the AIC values. The lower the value, the better[29]. An 'x' in the last column indicates that a grammatical construction is excluded from analysis. This exclusion can have either of the following three reasons: the absolute frequency is too low to successfully estimate a model, one of the AIC values is negative or the grammatical construction does not appear in the corpus at all.

There are 57 tenses and aspects listed in Table 12. In 22 cases it was not possible to estimate a model. There are two cases in which the linear regression fits best.

In three cases the negative binomial regression type II has the best fit. The variance of the remaining 30 tenses and aspects could be modelled with a negative binomial type I regression. Interestingly, the AIC values suggest the use of a negative binomial type II regression only for the conditional active and passive. The adjusted R^2 values show that this allocation is not due to low frequencies which might obscure the results.

[29] AIC values can only be compared for the same underlying model composition, i.e. only row wise.

Tense and Aspect	Adj. R^2	Linear Model	Poisson	Neg. Binomial I	Neg. Binomial II	Best Model
Present Simple 3ps	0.602	28, 802.220	29, 914.890	27, 264.830	27, 603.220	Neg. Binomial I
Present Simple non 3ps	0.280	29, 283.080	37, 929.860	28, 890.910	29, 033.380	Neg. Binomial I
Present Simple	0.590	32, 278.870	37, 129.640	31, 602.890	31, 878.590	Neg. Binomial I
Present Perfect Have	0.144	14, 670.360	11, 385.710	10, 849.240	10, 883.600	Neg. Binomial I
Present Perfect Has	0.147	13, 777.540	10, 469.470	9, 986.398	10, 004.950	Neg. Binomial I
Present Perfect	0.209	18, 022.100	15, 480.050	14, 489.610	14, 552.830	Neg. Binomial I
Present Progressive are	0.061	12, 829.910	9, 633.025	9, 125.695	9, 155.508	Neg. Binomial I
Present Progressive is	0.131	12, 499.170	9, 189.686	8, 920.872	8, 943.692	Neg. Binomial I
Present Progressive	0.125	16, 333.080	13, 693.260	12, 929.300	12, 961.700	Neg. Binomial I
Present Perfect Progressive Have	0.008	-47.004	1, 778.297	1, 760.054	1, 760.176	X
Present Perfect Progressive Has	0.004	-1, 749.722	1, 368.025	1, 356.194	1, 356.519	X
Present Perfect Progressive	0.012	2, 509.906	2, 627.671	2, 586.537	2, 589.575	Linear Model
Present Active	0.623	32, 537.620	36, 997.760	31, 917.960	32, 233.110	Neg. Binomial I
Past Simple	0.275	27, 964.940	30, 921.350	21, 920.910	22, 225.490	Neg. Binomial I
Past Perfect	0.128	10, 125.690	4, 899.737	4, 379.155	4, 418.114	Neg. Binomial I
Past Progressive	0.005	-6, 738.925	608.637	610.518	0	X
Past Perfect Progressive	0.015	-6, 690.401	518.320	512.793	0	X
Past Active	0.280	28, 563.400	32, 131.790	22, 276.120	22, 615.310	Neg. Binomial I
Going to is	0.020	-4, 450.104	639.627	621.296	619.098	X
Going to are	0.022	2, 915.435	2, 672.695	2, 616.117	0	X
Going to	0.018	3, 994.021	3, 046.436	2, 952.219	122, 454.800	Neg. Binomial I
Will Future Simple	0.116	20, 734.580	19, 509.740	16, 671.040	16, 734.880	Neg. Binomial I
Will Future Perfect	0.008	-6, 509.354	527.165	515.627	0	X
Will Future Progressive	0.039	-2, 058.261	1, 138.545	1, 125.893	0	X
Will Future Perfect Progressive	-0.002	-24, 540.000	68.000	70.000	0	X
Future Active	0.122	21, 274.280	20, 479.800	17, 460.040	17, 511.080	Neg. Binomial I
Present Simple Passive Are	0.164	16, 424.800	14, 260.100	13, 725.340	13, 774.390	Neg. Binomial I
Present Simple Passive Is	0.280	17, 235.350	14, 915.360	14, 320.570	14, 348.750	Neg. Binomial I
Present Simple Passive	0.321	20, 323.520	19, 288.980	18, 328.030	18, 384.700	Neg. Binomial I
Present Perfect Passive have	0.041	5, 769.452	4, 104.196	4, 026.234	4, 033.059	Neg. Binomial I
Present Perfect Passive has	0.068	7, 475.624	5, 151.615	5, 031.292	5, 044.580	Neg. Binomial I
Present Perfect Passive	0.087	10, 387.790	7, 275.171	7, 029.930	7, 059.792	Neg. Binomial I
Present Progressive Passive are	0.016	-602.353	1, 475.084	1, 444.875	116, 700.400	X
Present Progressive Passive is	0.006	-1, 545.295	1, 381.639	1, 368.369	1, 365.252	X
Present Progressive Passive	0.018	2, 409.934	2, 410.161	2, 343.409	2, 346.593	Neg. Binomial I
Present Perfect Progressive Passive has	0	0	0	0	0	X
Present Passive	0.181	17, 108.010	15, 174.340	14, 624.100	14, 689.730	Neg. Binomial I
Past Simple Passive	0.221	17, 981.350	13, 565.370	11, 573.480	11, 628.080	Neg. Binomial I
Past Perfect Passive	0	0	0	0	0	X
Past Progressive Passive	0	0	0	0	0	X
Past Perfect Progressive Passive	0	0	0	0	0	X
Past Passive	0.221	17, 984.600	13, 569.770	11, 575.020	11, 629.200	Neg. Binomial I
Will Future Simple Passive	0.067	9, 332.043	5, 910.486	5, 567.960	5, 589.313	Neg. Binomial I
Will Future Perfect Passive	-0.002	-19, 680.550	91.415	0	0	X
Will Future Progressive Passive	0	0	0	0	0	X
Will Future Perfect Progressive	0	0	0	0	0	X
Will Future Passive	0.067	9, 336.791	5, 917.421	5, 576.026	5, 597.563	Neg. Binomial I
Conditional Simple	0.192	23, 614.980	24, 280.820	22, 163.710	22, 154.140	Neg..Binomial II
Conditional Perfect	0.045	4, 658.010	2, 933.747	2, 813.911	2, 812.113	Neg. Binomial II
Conditional Progressive	0.009	406.115	1, 874.778	1, 853.474	1, 850.127	Linear Model
Conditional Perfect Progressive	0	0	0	0	0	X
Conditional Active	0.197	23, 768.590	24, 508.190	22, 363.280	22, 354.790	Neg. Binomial II
Conditional Simple Passive	0.101	15, 655.560	12, 649.450	11, 662.850	11, 669.730	Neg. Binomial I
Conditional Perfect Passive	0.014	-4, 630.756	811.243	809.112	19, 423.280	X
Conditional Progressive Passive	0	0	0	0	0	X
Conditional Perfect Progressive Passive	0	0	0	0	0	X
Conditional Passive	0.194	23, 624.190	24, 294.020	22, 182.890	22, 173.310	Neg. Binomial II

Table 12: Adjusted R^2-Values and AIC-Values with Indicated Best Model.

Based on these findings, the following chapters will present selected models in detail and will indicate underlying linguistic trends. The remaining results can be found in the Appendix.

4.2.2 Present Passive

The present passive combines all aspects of the present indicative. As Table 12 (page 46) shows, the negative binomial regression has the best fit for the present indicative passive. Although the AIC value of 14,622.15 for the negative binomial regression is 17.13 % lower than the linear regression, it has nonetheless been computed, in order to have a point of reference. The parameter values have to be interpreted as stated in Chapter 3.4.3.

In total, Table 13 (page 48) contains the β coefficients, standard errors and asterisks indicating the significant p-values of 16 variables[30]. In case of the nominally scaled dichotomous variables, the non-default values are attached to the variable names.

10 out of 16 variables are equally significant for the negative binomial model as well as for the linear one. In the negative binomial (linear) model eight (seven) variables have no significant influence on the frequency of the present passive, while none (two) is (are) significant at the 10% level. Three (two) are significant at the 5% level and five (five) at the 1% level. Except for the exam situation all of the administrative variables have a significant influence on the frequency of the passive. The text type has the strongest influence. Of the social variables none is significant, except for the months the students spent abroad. For each month spent abroad the logarithm of the number of present passives drops by -0.004. Quite naturally all variables containing the number of words per text, i.e. length, average sentence length and Flesh-Kincaid Reading Ease score, significantly influence the number of present passives. The number of sentences per text has a significant influence in the linear model but not in the negative binomial model.

Table 13 does not contain the mother tongue variable. It has been omitted on purpose. For multi-nominal or dummy-coded variables usually one parameter value is chosen to be the default and the other parameter values are compared to it. In the classical CIA this is not a severe problem, since the NS group can be selected to be the default and then all NNS groups are compared to it. Yet, because there is no NS group, the default needs to be altered until the difference between all NNS groups has been calculated. Figure 11 (page 50) presents the pairwise differences between the various L1. The L1 varieties on the right of the figure are the reference or default varieties.

For each L1 combination three pieces of information are given. The upper number indicates the β coefficient, while the lower number stands for the standard error. The colour shows whether the difference is significant and if the difference is due to under-

[30] As explained in Chapter 3.7.

	Present Passive	
	Negative Binomial Regression	Linear Regression
Text Type (Literary)	−0.232***	−0.431***
	(0.080)	(0.130)
Timing (Yes)	−0.121**	−0.152
	(0.061)	(0.097)
Reference Tools (Yes)	0.096**	0.157*
	(0.049)	(0.0084)
Exam (Yes)	0.078	0.163*
	(0.059)	(0.093)
Age	0.004	0.016**
	(0.005)	(0.008)
Sex (Male)	−0.049	0.094
	(0.037)	(0.061)
Other Home Languages	0.018	0.001
	(0.092)	(0.154)
English at School	0.003	0.008
	(0.009)	(0.015)
English at University	0.016	0.028
	(0.018)	(0.031)
Months Abroad	−0.004**	−0.006**
	(0.002)	(0.003)
N. other Languages	−0.003	−0.02
	(0.018)	(0.032)
Length	0.001***	0.001***
	(0.0002)	(0.0003)
N. Sentences	0.003	0.016***
	(0.003)	(0.005)
Avg. Sentence Length	−0.023***	−0.001
	(0.006)	(0.007)
Flesch-Kincaid Reading Ease	−0.015***	−0.015***
	(0.002)	(0.003)
Constant	0.987***	0.776***
	(0.231)	(0.299)
N	4,424	
Log. Likelihood	−7,279.077	−8,530.132
θ	2.967*** (0.191)	
Akaike Inf. Crit.	14,622.150	17,126.26
Notes:		***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 13: Summary of Negative Binomial Regression Type I and Linear Regression of the Present Passive with Indicated Changes from the Defaults and Model Statistics

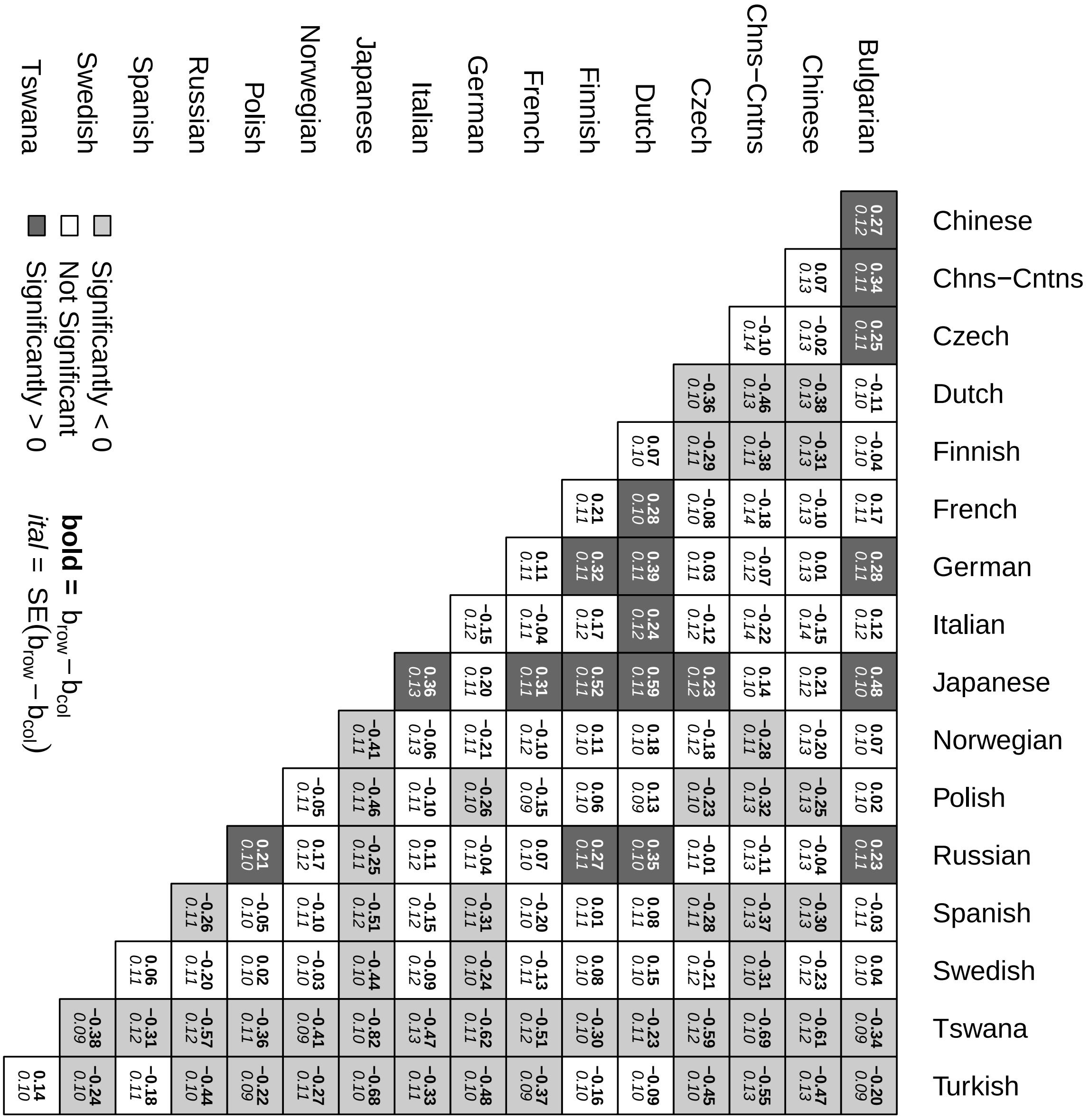
or overuse. The box for the Chinese Bulgarian combination is for instance dark grey and contains the numbers 0.27 and 0.12. This means that native speakers of Chinese use significantly less present passives per text than a native speaker of Bulgarian. The difference can be specified as it adds 0.27 to the logarithm of the present passive count. At this point the standard errors are given to assure transparency. They can be used to calculate the t-statistics and thereby the p-value.

Two NNS varieties immediately arouse the attention since they use the present passive in almost every case more often than the other NNS varieties. Speakers of Tswana use the present passive in 15 out of 16 cases significantly more often. With 13 out 16 Turkish is close behind, whereas in 13 out of 16 cases native speakers of Japanese use the present passive significantly less. The other 14 L1s do not show such extreme behaviour (see Table 14).

	sig+	sig-	insig
Bulgarian	6	2	8
Chinese	0	7	9
Chinese-Cantonese	0	9	7
Czech	1	7	8
Dutch	8	1	7
Finnish	6	1	9
French	1	3	12
German	0	8	8
Italian	1	3	12
Japanese	0	13	3
Norwegian	2	2	12
Polish	6	2	8
Russian	1	7	8
Spanish	6	1	9
Swedish	3	2	11
Tswana	15	0	1
Turkish	12	0	4

Table 14: Pairwise Differences in L1 Parameter Values for Present Passive Neg. Binomial Type I Regression.

Although not apparent at first glance, there are clear and specific patterns for the language families. Within the Germanic languages significant differences can be found between speakers of Dutch and Germans, with an imbalance towards the Dutch. At the same time Swedes use more present passives than Germans. The difference between Swedes, Norwegians and the Dutch is not significant. The discrepancy between the three Asian languages is not significant, which is remarkable since the Japanese under-use the passive most frequently. Between native speakers of Romance languages no significant



	Bulgarian	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana
Chinese	■ **0.27** *0.12*															
Chns−Cntns	■ **0.34** *0.11*	□ **0.07** *0.13*														
Czech	■ **0.25** *0.11*	□ **−0.02** *0.13*	□ **−0.10** *0.14*													
Dutch	□ **−0.11** *0.10*	▨ **−0.38** *0.13*	▨ **−0.46** *0.13*	▨ **−0.36** *0.10*												
Finnish	□ **−0.04** *0.10*	▨ **−0.31** *0.13*	▨ **−0.38** *0.11*	▨ **−0.29** *0.11*	□ **0.07** *0.10*											
French	□ **0.17** *0.11*	□ **−0.10** *0.13*	□ **−0.18** *0.14*	□ **−0.08** *0.10*	■ **0.28** *0.10*	□ **0.21** *0.11*										
German	■ **0.28** *0.11*	□ **0.01** *0.13*	□ **−0.07** *0.12*	□ **0.03** *0.11*	■ **0.39** *0.11*	■ **0.32** *0.11*	□ **0.11** *0.11*									
Italian	□ **0.12** *0.12*	□ **−0.15** *0.14*	□ **−0.22** *0.14*	□ **−0.12** *0.12*	■ **0.24** *0.12*	□ **0.17** *0.12*	□ **−0.04** *0.11*	□ **−0.15** *0.12*								
Japanese	■ **0.48** *0.10*	□ **0.21** *0.12*	□ **0.14** *0.10*	■ **0.23** *0.12*	■ **0.59** *0.11*	■ **0.52** *0.11*	■ **0.31** *0.11*	□ **0.20** *0.11*	■ **0.36** *0.13*							
Norwegian	□ **0.07** *0.10*	□ **−0.20** *0.13*	▨ **−0.28** *0.11*	□ **−0.18** *0.12*	□ **0.18** *0.10*	□ **0.11** *0.10*	□ **−0.10** *0.12*	□ **−0.21** *0.11*	□ **−0.06** *0.13*	▨ **−0.41** *0.11*						
Polish	□ **0.02** *0.10*	▨ **−0.25** *0.13*	▨ **−0.32** *0.13*	▨ **−0.23** *0.10*	□ **0.13** *0.09*	□ **0.06** *0.10*	□ **−0.15** *0.09*	▨ **−0.26** *0.10*	□ **−0.10** *0.11*	▨ **−0.46** *0.11*	□ **−0.05** *0.11*					
Russian	■ **0.23** *0.11*	□ **−0.04** *0.13*	□ **−0.11** *0.13*	□ **−0.01** *0.11*	■ **0.35** *0.10*	■ **0.27** *0.11*	□ **0.07** *0.10*	□ **−0.04** *0.11*	□ **0.11** *0.12*	▨ **−0.25** *0.11*	□ **0.17** *0.12*	■ **0.21** *0.10*				
Spanish	□ **−0.03** *0.11*	▨ **−0.30** *0.13*	▨ **−0.37** *0.13*	▨ **−0.28** *0.11*	□ **0.08** *0.11*	□ **0.01** *0.11*	□ **−0.20** *0.10*	▨ **−0.31** *0.11*	□ **−0.15** *0.12*	▨ **−0.51** *0.12*	□ **−0.10** *0.11*	□ **−0.05** *0.10*	▨ **−0.26** *0.11*			
Swedish	□ **0.04** *0.10*	□ **−0.23** *0.12*	▨ **−0.31** *0.10*	□ **−0.21** *0.12*	□ **0.15** *0.10*	□ **0.08** *0.10*	□ **−0.13** *0.11*	▨ **−0.24** *0.10*	□ **−0.09** *0.12*	▨ **−0.44** *0.10*	□ **−0.03** *0.10*	□ **0.02** *0.10*	□ **−0.20** *0.11*	□ **0.06** *0.11*		
Tswana	▨ **−0.34** *0.09*	▨ **−0.61** *0.12*	▨ **−0.69** *0.10*	▨ **−0.59** *0.12*	▨ **−0.23** *0.11*	▨ **−0.30** *0.10*	▨ **−0.51** *0.12*	▨ **−0.62** *0.11*	▨ **−0.47** *0.13*	▨ **−0.82** *0.10*	▨ **−0.41** *0.09*	▨ **−0.36** *0.11*	▨ **−0.57** *0.12*	▨ **−0.31** *0.12*	▨ **−0.38** *0.09*	
Turkish	▨ **−0.20** *0.09*	▨ **−0.47** *0.13*	▨ **−0.55** *0.13*	▨ **−0.45** *0.10*	□ **−0.09** *0.10*	□ **−0.16** *0.10*	▨ **−0.37** *0.09*	▨ **−0.48** *0.10*	▨ **−0.33** *0.11*	▨ **−0.68** *0.10*	▨ **−0.27** *0.11*	▨ **−0.22** *0.09*	▨ **−0.44** *0.10*	□ **−0.18** *0.11*	▨ **−0.24** *0.10*	□ **0.14** *0.10*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 11: Pairwise Differences in L1 Parameter Values for Present Passive Neg. Binomial Type I Regression.

differences can be found. Due to the significant differences between Russian and Bulgarian, Polish and Russian, and Czech and Polish the Slavic languages do not form a uniform block. There is a Romance and to a certain degree also Germanic language block and thus the present passive seem to have a tendency to be organised according to the grammatical rules of language families rather than the rules of single languages. Yet, Slavic languages seem to be an exception with L1 specific frequencies.

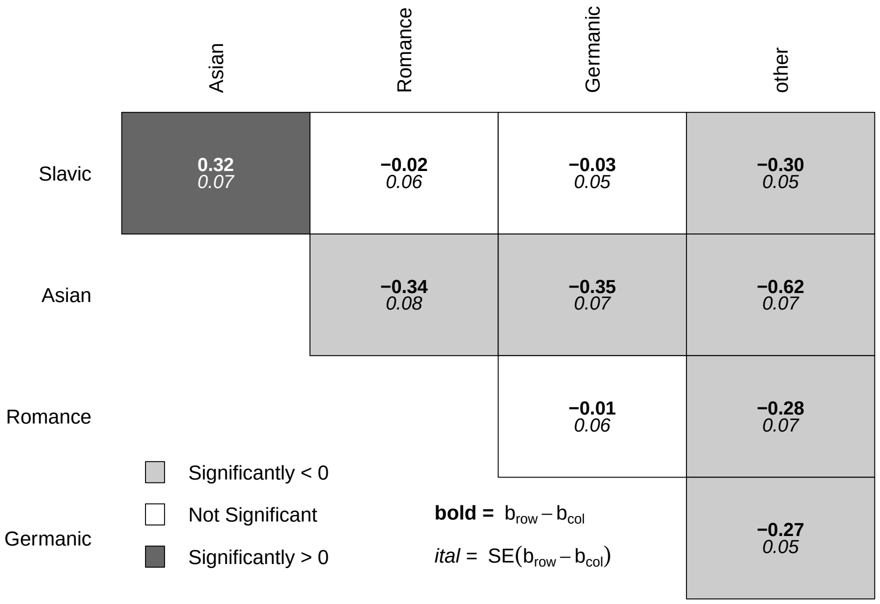
In order to evaluate if the statements about the language families are true, the single L1s have been combined according to their respective language families. Russian, Polish, Bulgarian and Czech form the Slavic block, whereas German, Dutch, Swedish and Norwegian are Germanic, French, Italian and Spanish are Romance languages and Chinese, Cantonese and Japanese are Asian languages. The remaining L1s are summarised under 'other'. This categorisation may not withstand a closer checking by experts in the respective languages yet for a superficial analysis of language family specific usage of the present passive it should be enough. The following figure shows a summary of the previous Figure 11.

The observation that Asian languages use less present passives than the other language families is supported by the summarised plot in Figure 12. While there are no significant differences between Romance, Germanic and Slavic languages, Asians use significantly less and the remaining languages significantly more present passives than Germanic, Romance and Slavic languages.

Up to now the regression has shown which variables significantly influence the frequency of the present passive. Yet, due to the structure of the L1 variable which is multilevel, dummy-coded and nominal-scaled, it was not possible to show if the L1 variable has a significant influence on the frequency of the present passive. In order to evaluate the relative importance of the single variables, the overall fit of the model needs to be decomposed. Since this decomposition is based on R^2, it cannot be applied to any other regression type than the linear regression. However, as mentioned above, the results of the linear regression are in terms of quality, i.e. direction and significance of the β coefficients, not too far away from the negative binomial regression. This justifies another look at linear regressions.

Table 15 (page 52) shows the R^2 decomposition as proposed by Lindeman's, Merenda's and Gold (Grömping 2006). Decomposed for each variable, the single values stand for the share of the explained variance[31], on the basis of the linear model. The single values add up to the total R^2. In case of the present passive, 18.29% of the variance can be explained with the linear model. 11.1% are caused by the length of the text, the number of sentences and the variable calculated on their basis.

[31] Multiplied with 100 they can be read as percentages.



	Asian	Romance	Germanic	other
Slavic	**0.32** *0.07*	**−0.02** *0.06*	**−0.03** *0.05*	**−0.30** *0.05*
Asian		**−0.34** *0.08*	**−0.35** *0.07*	**−0.62** *0.07*
Romance			**−0.01** *0.06*	**−0.28** *0.07*
Germanic				**−0.27** *0.05*

Figure 12: Pairwise Differences in Language Family Parameter Values for Present Passive Neg. Binomial Type I Regression.

Statistic	Relative Importance
L1	0.041
Text Type	0.002
Conditions	0.007
Reference Tools	0.005
Exam	0.003
Age	0.001
Sex	0.001
Other Home Languages	0.00001
English at School	0.002
English at University	0.003
Months Abroad	0.001
N. other Languages	0.005
Length	0.057
N. Sentences	0.043
Avg. Sentence Length	0.002
Flesch-Kincaid Reading Ease	0.009
Proportion of variance explained by model:	18.29%

Table 15: Lindeman's, Merenda's and Gold's Relative Importance for the Linear Model of the Present Passive.

Again, this does not come as a surprise, since students primarily need more words to be able to use more passives, rather than another month abroad or another year of English classes. 57% of the remaining 7.19% of the explained variance are due to the L1. This means that 57% of real linguistic variance, i.e. variance caused by language transfer and extra-linguistic influences, is due to L1 specific transfer.

4.2.3 Past Passive or the Past Simple Passive

In contrast to the present passive the past passive has only around 64% of the present passive's absolute frequency. Nonetheless, the past passive has an R^2 value which is 4 per cent points higher than that of the present passive. For the past passive, the negative binomial regression has a 36% lower AIC value than the corresponding linear regression. This difference is higher than for the respective regression types for the present passive. The better fit of the past passive models is due to the difference in the absolute frequency composition. While there are passive constructions for all aspects in the present tense, only the simple is used in the past tense. Therefore, the low fits of the perfect, the progressive and the perfect progressive do not diminish the overall fit of the past passive.

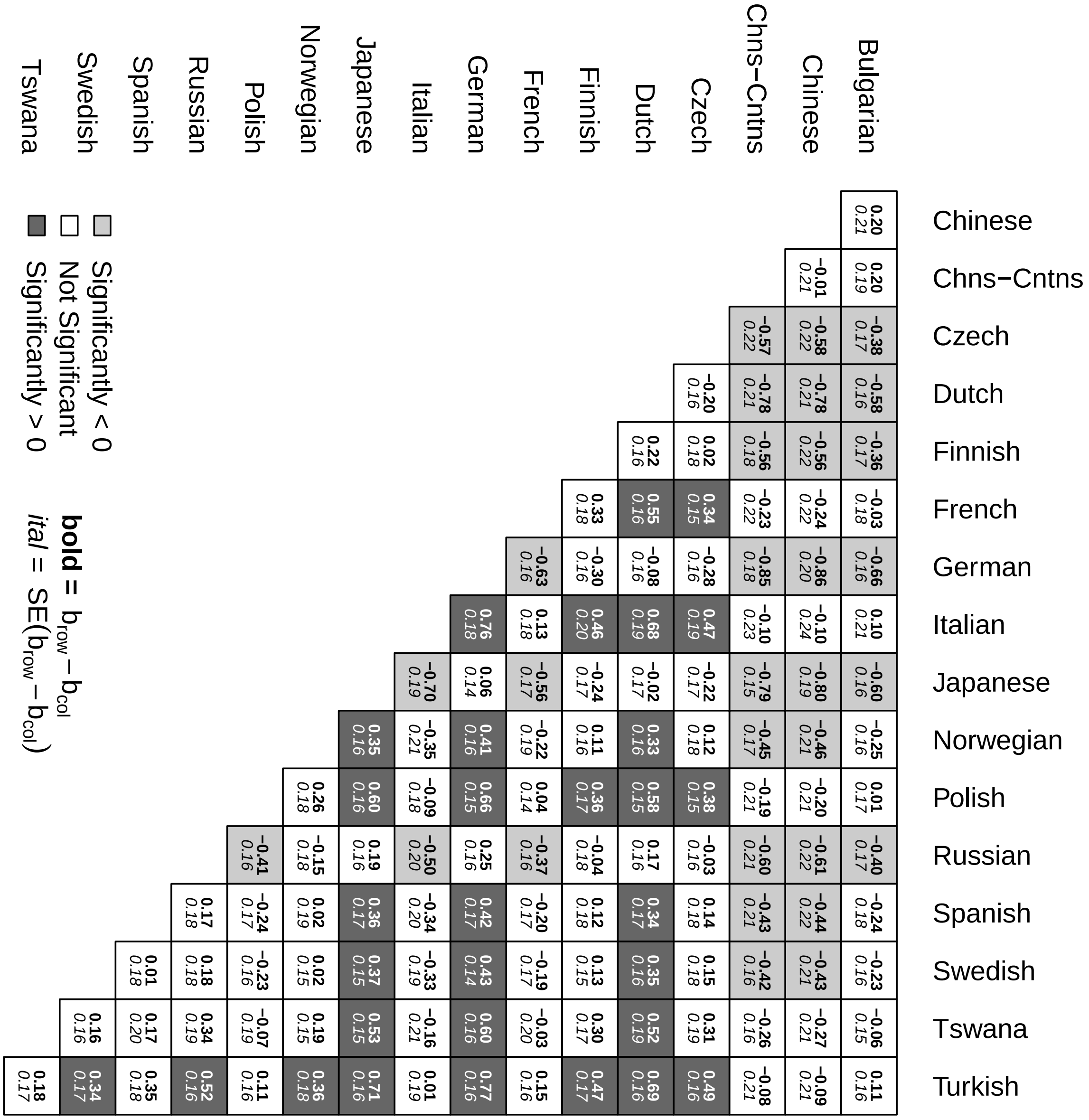
Like in Chapter 4.2.2, the following Table 16 (page 54) shows the β coefficients with their significance levels indicated and the standard errors for the negative binomial and linear regression. The L1 variable will again be discussed at a later stage of this chapter.

At a first glance, the significance levels are differently distributed between the linear and the negative binomial regression in case of the past passive compared to the present passive. There are four additional variables significant at the 5% level in the linear regression. In the end three variables have a significant influence in the linear regression that are not significant in the negative binomial regression. Age, the number of years of English at school and the number of sentences are significant in the linear model but not in the negative binomial model. Nonetheless, a similar pattern as in the present passive can also be found with the results for the past passive. Almost every linguistic variable has a significant influence as well as the text genre and the exam situation, whereas the majority of the social variables is not significant. Interestingly, students who speak another language at home tend to use less past passives than mono-lingual students, which seems to support the idea that a higher quantity of spoken languages might drop their respective language performances.

There seems to be a disparity between northern and southern Germanic languages. While speakers of Dutch and German, and of Swedish and Norwegian do not use the past passive differently, the southern NNS varieties apply significantly more past passive constructions than the northern varieties. The Asian speakers of English behave like they have done with the present passive. There is no difference between native speakers of

	Past Passive	
	Negative Binomial Regression	Linear Regression
Text Type (Literary)	0.337***	0.500***
	(0.114)	(0.143)
Timing (Yes)	−0.061	−0.164
	(0.089)	(0.107)
Reference Tools (Yes)	0.046	0.091
	(0.078)	(0.092)
Exam (Yes)	0.139*	0.334***
	(0.084)	(0.103)
Age	0.004	0.018**
	(0.008)	(0.009)
Sex (Male)	0.107*	0.067
	(0.059)	(0.068)
Other Home Languages	−0.430**	−0.269
	(0.180)	(0.170)
English at School	0.006	0.030*
	(0.014)	(0.016)
English at University	0.031	0.091**
	(0.028)	(0.034)
Months Abroad	0.002	0.002
	(0.003)	(0.003)
N. other Languages	−0.037	−0.072**
	(0.030)	(0.035)
Length	0.002***	0.004***
	(0.0003)	(0.0003)
N. Sentences	−0.006	−0.016***
	(0.005)	(0.006)
Avg. Sentence Length	−0.017**	−0.022***
	(0.008)	(0.008)
Flesch-Kincaid Reading Ease	0.006**	0.008***
	(0.003)	(0.003)
Constant	−1.637***	−2.007***
	(0.293)	(0.329)
N	4,424	
Log Likelihood	−5,756.285	−8, 958.447
θ	0.739*** (0.036)	
Akaike Inf. Crit.	11,576.570	17,982.89
Notes:		***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 16: Summary of Negative Binomial Regression Type I and Linear Regression of the Past Passive with Indicated Changes from the Defaults and Model Statistics



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	□ **0.20** *0.21*	□ **0.20** *0.19*	▒ **−0.38** *0.17*	▒ **−0.58** *0.16*	▒ **−0.36** *0.17*	□ **−0.03** *0.18*	▒ **−0.66** *0.16*	□ **0.10** *0.21*	▒ **−0.60** *0.16*	□ **−0.25** *0.16*	□ **0.01** *0.17*	▒ **−0.40** *0.17*	□ **−0.24** *0.18*	□ **−0.23** *0.16*	□ **−0.06** *0.15*	□ **0.11** *0.16*
Chinese		□ **−0.01** *0.21*	▒ **−0.58** *0.22*	▒ **−0.78** *0.21*	▒ **−0.56** *0.22*	□ **−0.24** *0.22*	▒ **−0.86** *0.20*	□ **−0.10** *0.24*	▒ **−0.80** *0.19*	▒ **−0.46** *0.21*	□ **−0.20** *0.21*	▒ **−0.61** *0.22*	▒ **−0.44** *0.22*	▒ **−0.43** *0.21*	□ **−0.27** *0.21*	□ **−0.09** *0.21*
Chns−Cntns			▒ **−0.57** *0.22*	▒ **−0.78** *0.21*	▒ **−0.56** *0.18*	□ **−0.23** *0.22*	▒ **−0.85** *0.18*	□ **−0.10** *0.23*	▒ **−0.79** *0.15*	▒ **−0.45** *0.17*	□ **−0.19** *0.21*	▒ **−0.60** *0.21*	▒ **−0.43** *0.21*	▒ **−0.42** *0.16*	□ **−0.26** *0.16*	□ **−0.08** *0.21*
Czech				□ **−0.20** *0.16*	□ **0.02** *0.18*	■ **0.34** *0.15*	□ **−0.28** *0.16*	■ **0.47** *0.19*	□ **−0.22** *0.17*	□ **0.12** *0.18*	■ **0.38** *0.15*	□ **−0.03** *0.16*	□ **0.14** *0.18*	□ **0.15** *0.18*	□ **0.31** *0.19*	■ **0.49** *0.16*
Dutch					□ **0.22** *0.16*	■ **0.55** *0.16*	□ **−0.08** *0.16*	■ **0.68** *0.19*	□ **−0.02** *0.17*	■ **0.33** *0.16*	■ **0.58** *0.15*	□ **0.17** *0.16*	■ **0.34** *0.17*	■ **0.35** *0.16*	■ **0.52** *0.19*	■ **0.69** *0.16*
Finnish						□ **0.33** *0.18*	□ **−0.30** *0.16*	■ **0.46** *0.20*	□ **−0.24** *0.17*	□ **0.11** *0.16*	■ **0.36** *0.17*	□ **−0.04** *0.18*	□ **0.12** *0.18*	□ **0.13** *0.15*	□ **0.30** *0.17*	■ **0.47** *0.17*
French							▒ **−0.63** *0.16*	□ **0.13** *0.18*	▒ **−0.56** *0.17*	□ **−0.22** *0.19*	□ **0.04** *0.14*	▒ **−0.37** *0.16*	□ **−0.20** *0.17*	□ **−0.19** *0.17*	□ **−0.03** *0.20*	□ **0.15** *0.16*
German								■ **0.76** *0.18*	□ **0.06** *0.14*	■ **0.41** *0.16*	■ **0.66** *0.15*	□ **0.25** *0.16*	■ **0.42** *0.17*	■ **0.43** *0.14*	■ **0.60** *0.16*	■ **0.77** *0.16*
Italian									▒ **−0.70** *0.19*	□ **−0.35** *0.21*	□ **−0.09** *0.18*	▒ **−0.50** *0.20*	□ **−0.34** *0.20*	□ **−0.33** *0.19*	□ **−0.16** *0.21*	□ **0.01** *0.19*
Japanese										■ **0.35** *0.16*	■ **0.60** *0.16*	□ **0.19** *0.16*	■ **0.36** *0.17*	■ **0.37** *0.15*	■ **0.53** *0.15*	■ **0.71** *0.16*
Norwegian											□ **0.26** *0.18*	□ **−0.15** *0.18*	□ **0.02** *0.19*	□ **0.02** *0.15*	□ **0.19** *0.15*	■ **0.36** *0.18*
Polish												▒ **−0.41** *0.16*	□ **−0.24** *0.17*	□ **−0.23** *0.16*	□ **−0.07** *0.19*	□ **0.11** *0.16*
Russian													□ **0.17** *0.18*	□ **0.18** *0.18*	□ **0.34** *0.19*	■ **0.52** *0.16*
Spanish														□ **0.01** *0.18*	□ **0.17** *0.20*	□ **0.35** *0.18*
Swedish															□ **0.16** *0.16*	■ **0.34** *0.17*
Tswana																□ **0.18** *0.17*

▒ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 13: Pairwise Differences in L1 Parameter Values for Past Passive Neg. Binomial Type I Regression.

Mandarin and Cantonese, but the Japanese use significantly less past passives. For the Romance languages there is an imbalance between French and Italian towards French, but none between French and Spanish, and Italian and Spanish. Czech and Russian students use more past passives than Polish ones and there is no significant difference between Czech and Russian speakers of English. Native speakers of Bulgarian use about as many past passives as native speakers of Polish. In contrast to the present passive Polish students have overtaken their fellow Russian and Czech students.

	sig+	sig-	insig
Bulgarian	0	6	10
Chinese	0	8	8
Chinese-Cantonese	0	9	7
Czech	7	0	9
Dutch	9	0	7
Finnish	6	0	10
French	0	5	11
German	11	0	5
Italian	0	6	10
Japanese	11	0	5
Norwegian	2	2	12
Polish	0	6	10
Russian	7	0	9
Spanish	1	2	13
Swedish	2	3	11
Tswana	0	3	13
Turkish	0	6	10

Table 17: Pairwise Differences in L1 Parameter Values for Past Passive Neg. Binomial Type I Regression.

In contrast to the present passive the summary of Figure 13 draws a different picture. The vast majority of L1s are statistically insignificant in terms of the frequency of their past passive use. In more than half of the cases, only the southern Germanic languages and Japanese use significantly more past passives than the language they are compared to. There are in total fewer significant differences between the various L1s.

The summary according to the language families supports this first impression. Speakers of Asian languages apply significantly more past passives than students with Romance native language backgrounds. Although native speakers of Chinese and Cantonese use far less past passives than speakers of most other L1s, the difference towards Romance languages is only marginal. The overuse of the past passive by the Japanese is so high that it moves the Asian languages towards a significant imbalance. The strong overuse of the past passive by native speakers of Dutch and German and the indifferent usage by Swedes and Norwegians causes the significant overuse of the past passive by the Germanic languages in contrast to all other language families. The remaining L1 family contrasts are insignificant.

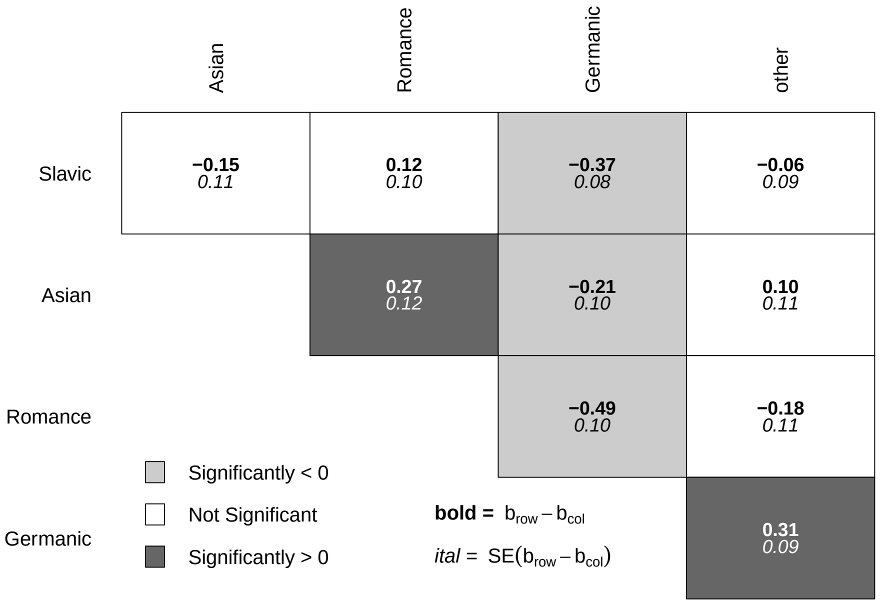


Figure 14: Pairwise Differences in Language Family Parameter Values for Past Passive Neg. Binomial Type I Regression.

Statistic	Relative Importance
L1	0.046
Text Type	0.005
Conditions	0.001
Reference Tools	0.003
Exam	0.002
Age	0.001
Sex	0.0002
Other Home Languages	0.001
English at School	0.002
English at University	0.003
Months Abroad	0.0004
N. other Languages	0.001
Length	0.096
N. Sentences	0.06
Avg. Sentence Length	0.003
Flesch-Kincaid Reading Ease	0.002
Proportion of variance explained by model:	22.68%

Table 18: Lindeman's, Merenda's and Gold's Relative Importance for Linear Model of Past Simple Passive.

The linear model of the past passive has a fit of 22.68% (see Table 18/page 57). The greatest share of this overall fit is due to the text length and it's derivatives. 70.99% of the 22.68% are caused by linguistic variables. Another 20.28% are due to the L1 variable, i.e. the second most important variable category is the L1. The remaining 8.73% of the overall fit are split among the administrative and social variables. Apparently a lower absolute frequency seems to diminish the influence of the L1 variable.

4.2.4 Passive Indicative

Chapter 3.3 defined several grammatical constructions. Defining the conditional and thereby including it into in depth analysis would go beyond the constraints of this thesis. Therefore, I will do without it and will consequently call the summary of the combined passive constructions 'Passive Indicative'. This category does not appear in the previous discussion, because I consider it rather as a statistical construction under a working hypothesis than as a proper well defined grammatical category.

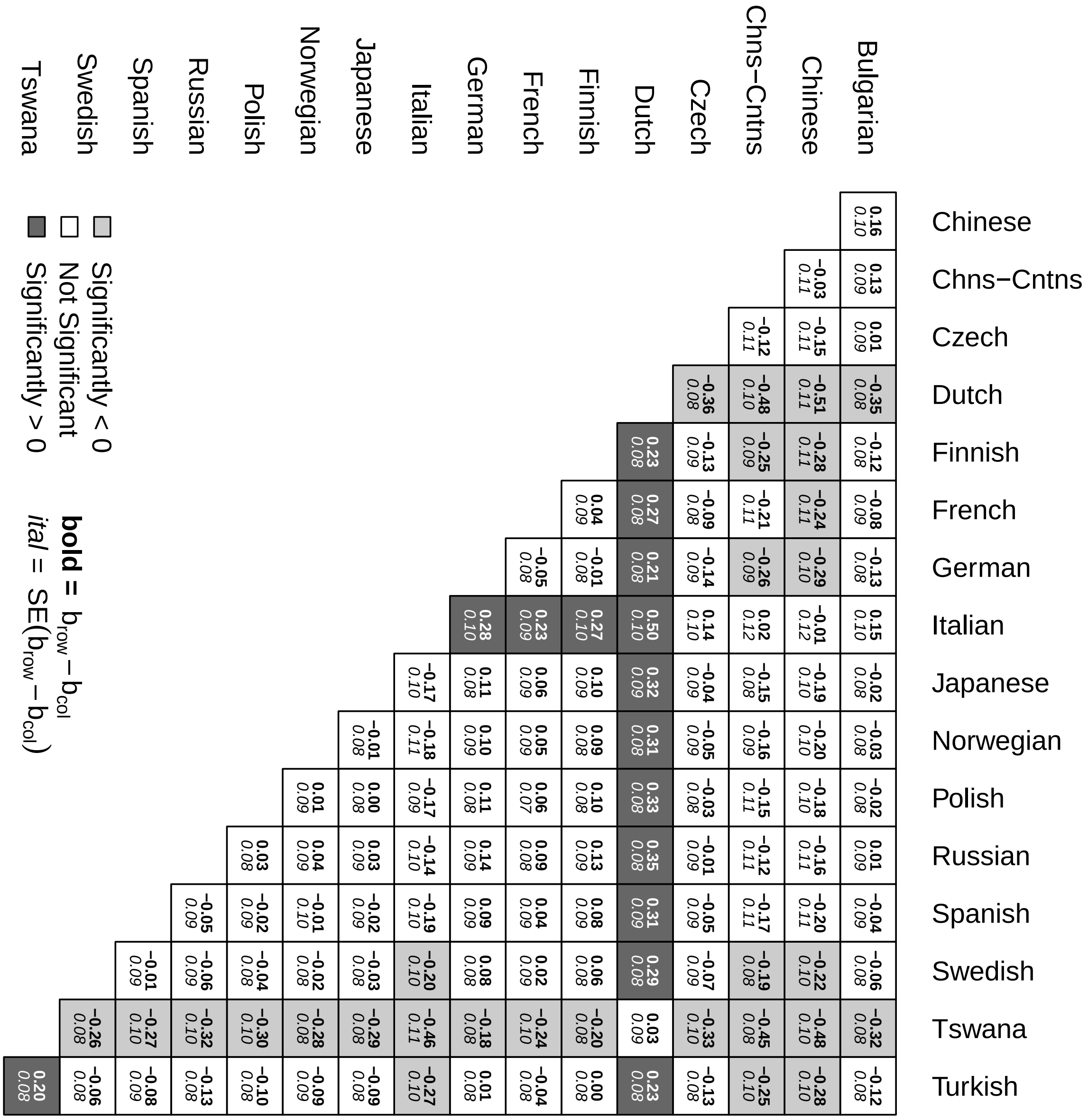
Table 19 (page 59) shows the results for the negative binomial and linear regression of the passive indicative. When compared to the previous regressions, relevant deviations can be seen. While the text type is significant for the present and past passive, it has no significant influence on the passive indicative. The remaining administrative variables are significant in contrast to the above mentioned regressions. Also, the social variables are not relevant for the frequency of the passive indicative. Concerning the linear regression they are significant. Interestingly the number of sentences is of no importance while the average sentence length is significant at the 5% level. The negative binomial regression has an AIC value 12.38% lower than that of the linear regression. Yet, compared to the previous regressions, the difference between the two regression types is far smaller.

Three clear claims can be made based on the pairwise differences in L1 parameter values. The first observation is the fact that native speakers of Dutch use significantly more indicative passives than all other L1s. The second observation is that students from Botswana also apply more indicative passives than the other L1s and even more than the Dutch. The last clear observation is that the remaining L1 differences are, with a few exceptions, not significant. With only few significant differences it makes little sense to combine the different L1s into language families.

The results of the R^2 decomposition on the basis of the β coefficients turned out as expected. Around 90% of the entire R^2 value is due to three variables. The text length and the number of sentences have a combined relative importance of 0.2446 or 73.78% of the total R^2. Another 0.0531 are caused by the L1, making it the most influential but not directly text based variable. The remaining variables have a negligible influence which corresponds to the results of Table 19 (page 59).

	Passive Indicative Negative Binomial Regression	Linear Regression
Text Type (Literary)	0.009	−0.001
	(0.061)	(0.193)
Timing (Yes)	-0.114^{**}	-0.304^{**}
	(0.048)	(0.144)
Reference Tools (Yes)	0.079^{**}	0.257^{**}
	(0.040)	(0.124)
Exam (Yes)	0.108^{**}	0.552^{***}
	(0.045)	(0.139)
Age	0.003	0.031^{***}
	(0.004)	(0.012)
Sex (Male)	0.004	−0.016
	(0.030)	(0.090)
Other Home Languages	−0.109	−0.298
	(0.080)	(0.229)
English at School	0.010	0.046^{**}
	(0.007)	(0.022)
English at University	0.019	0.104^{**}
	(0.015)	(0.046)
Months Abroad	−0.001	−0.004
	(0.001)	(0.004)
N. other Languages	−0.015	-0.086^{*}
	(0.015)	(0.047)
Length	0.001^{***}	0.006^{***}
	(0.0001)	(0.0005)
N. Sentences	0.004	0.003
	(0.002)	(0.008)
Avg. Sentence Length	-0.010^{**}	-0.023^{**}
	(0.005)	(0.011)
Flesch-Kincaid Reading Ease	-0.005^{***}	-0.009^{**}
	(0.001)	(0.004)
Constant	0.397^{**}	-1.302^{***}
	(0.176)	(0.444)
N	4,424	
Log Likelihood	−9,140.291	−10,275.01
θ	3.480^{***} (0.173)	
Akaike Inf. Crit.	18,344.580	20,616.02
Notes:		***Significant at the 1 percent level. **Significant at the 5 percent level. *Significant at the 10 percent level.

Table 19: Summary of Negative Binomial Regression Type I and Linear Regression of the Passive Indicative



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**0.16** *0.10*	**0.13** *0.09*	**0.01** *0.09*	**−0.35** *0.08*	**−0.12** *0.08*	**−0.08** *0.09*	**−0.13** *0.08*	**0.15** *0.10*	**−0.02** *0.08*	**−0.03** *0.08*	**−0.02** *0.08*	**0.01** *0.09*	**−0.04** *0.09*	**−0.06** *0.08*	**−0.32** *0.08*	**−0.12** *0.08*
Chinese		**−0.03** *0.11*	**−0.15** *0.11*	**−0.51** *0.11*	**−0.28** *0.11*	**−0.24** *0.11*	**−0.29** *0.10*	**−0.01** *0.12*	**−0.19** *0.10*	**−0.20** *0.10*	**−0.18** *0.10*	**−0.16** *0.11*	**−0.20** *0.11*	**−0.22** *0.10*	**−0.48** *0.10*	**−0.28** *0.10*
Chns−Cntns			**−0.12** *0.11*	**−0.48** *0.10*	**−0.25** *0.09*	**−0.21** *0.11*	**−0.26** *0.09*	**0.02** *0.12*	**−0.15** *0.08*	**−0.16** *0.09*	**−0.15** *0.11*	**−0.12** *0.11*	**−0.17** *0.11*	**−0.19** *0.08*	**−0.45** *0.08*	**−0.25** *0.10*
Czech				**−0.36** *0.08*	**−0.13** *0.09*	**−0.09** *0.08*	**−0.14** *0.09*	**0.14** *0.10*	**−0.04** *0.09*	**−0.05** *0.09*	**−0.03** *0.08*	**−0.01** *0.09*	**−0.05** *0.09*	**−0.07** *0.09*	**−0.33** *0.10*	**−0.13** *0.08*
Dutch					**0.23** *0.08*	**0.27** *0.08*	**0.21** *0.08*	**0.50** *0.10*	**0.32** *0.09*	**0.31** *0.08*	**0.33** *0.08*	**0.35** *0.08*	**0.31** *0.09*	**0.29** *0.08*	**0.03** *0.09*	**0.23** *0.08*
Finnish						**0.04** *0.09*	**−0.01** *0.08*	**0.27** *0.10*	**0.10** *0.09*	**0.09** *0.08*	**0.10** *0.08*	**0.13** *0.09*	**0.08** *0.09*	**0.06** *0.08*	**−0.20** *0.08*	**0.00** *0.08*
French							**−0.05** *0.08*	**0.23** *0.09*	**0.06** *0.09*	**0.05** *0.09*	**0.06** *0.07*	**0.09** *0.08*	**0.04** *0.09*	**0.02** *0.09*	**−0.24** *0.10*	**−0.04** *0.08*
German								**0.28** *0.10*	**0.11** *0.08*	**0.10** *0.09*	**0.11** *0.08*	**0.14** *0.09*	**0.09** *0.09*	**0.08** *0.08*	**−0.18** *0.08*	**0.01** *0.08*
Italian									**−0.17** *0.10*	**−0.18** *0.11*	**−0.17** *0.09*	**−0.14** *0.10*	**−0.19** *0.10*	**−0.20** *0.10*	**−0.46** *0.11*	**−0.27** *0.10*
Japanese										**−0.01** *0.08*	**0.00** *0.08*	**0.03** *0.09*	**−0.02** *0.09*	**−0.03** *0.08*	**−0.29** *0.08*	**−0.09** *0.08*
Norwegian											**0.01** *0.09*	**0.04** *0.09*	**−0.01** *0.10*	**−0.02** *0.08*	**−0.28** *0.08*	**−0.09** *0.09*
Polish												**0.03** *0.08*	**−0.02** *0.09*	**−0.04** *0.08*	**−0.30** *0.10*	**−0.10** *0.08*
Russian													**−0.05** *0.09*	**−0.06** *0.09*	**−0.32** *0.10*	**−0.13** *0.08*
Spanish														**−0.01** *0.09*	**−0.27** *0.10*	**−0.08** *0.09*
Swedish															**−0.26** *0.08*	**−0.06** *0.08*
Tswana																**0.20** *0.08*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 15: Pairwise Differences in L1 Parameter Values for Passive Indicative Neg. Binomial Type I Regression.

Statistic	Relative Importance
L1	0.0531
Text Type	0.0015
Timing	0.0041
Reference Tools	0.0069
Exam	0.0024
Age	0.0012
Sex	0.0003
Other Home Languages	0.0003
English at School	0.0032
English at University	0.0039
Months Abroad	0.0001
N. other Languages	0.0025
Length	0.1457
N. Sentences	0.0989
Avg. Sentence Length	0.0043
Flesch-Kincaid Reading Ease	0.0031
Proportion of variance explained by model:	33.16%

Table 20: Lindeman's, Merenda's and Gold's Relative Importance for Linear Model of Passive Indicative.

5 Discussion

Before comparing the results of Chapter 4 with these found in the literature (see Chapter 2) a brief summary shall give an overview of the findings so far.

The following Table 21 (page 62) contains the β coefficients for the negative binomial regression of the present and past passive and of the passive indicative. There are four variables that are significant for all three regression models: the length of the text, the average sentence length, the Flesch-Kincaid Reading Ease Score and the constant. Given the fact that one needs more words to be able to produce a construction of a certain type, it is no surprise that all four are significant. The remaining variables are not significant for all passive types. Yet, timing, text type and reference tools are significant in two out of three cases. Interestingly, the β coefficients for the text type change direction. Sex, the number of other languages spoken at home and the number of months spent abroad are only once significant. With -0.430 the number of other languages spoken at home has the strongest influence per unit of all observed variables. Although this β coefficient is the highest observed, its relative importance is to be neglected nonetheless.

Following Douglas Biber (1992), Granger (1983) and Weiner & Labov (1983) the text genre has a significant influence on the frequency of the passive. In two out of

	Summary of Negative Binomial Regressions		
	Present Passive	Past Passive	Passive Indicative
Text Type (Literary)	-0.232^{***}	0.337^{***}	0.009
Timing (Yes)	-0.121^{**}	-0.061	-0.114^{**}
Reference Tools (Yes)	0.096^{**}	0.046	0.079^{**}
Exam (Yes)	0.078	0.139^{*}	0.108
Age	0.004	0.004	0.003
Sex (Male)	-0.049	0.107^{**}	0.004
Other Home Languages	0.018	-0.430^{**}	-0.109
English at School	0.003	0.006	0.010
English at University	0.016	0.031	0.019
Months Abroad	-0.004^{**}	0.002	-0.001
N. other Languages	-0.003	-0.037	-0.015
Length	0.001^{***}	0.002^{***}	0.001^{***}
N. Sentences	0.003	-0.006	0.004
Avg. Sentence Length	-0.023^{***}	-0.017^{**}	-0.010^{**}
Flesch-Kincaid Reading Ease	-0.015^{***}	0.006^{**}	-0.005^{***}
Constant	0.987^{***}	-1.637^{***}	0.397^{**}

Notes: ***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Table 21: Summary of Negative Binomial Regression Type I of Present, Past and Combined Indicative Passive.

three cases this suggestion holds true. The problem is that for the present passive the frequency significantly decreases, while it significantly increases for the past passive. If all passives are combined, no significant effect can be found. This suggests that the genre dependency of the passive is dependent on the complexity of the passive and that Douglas' and Granger's theory cannot be generalised. Yet, as Chapter 4.1.1 shows, ICLE contains only two genres and those two are unevenly distributed. Thus, the partial rejection of Douglas' theory needs to be modified again. The data foundation of the ICLE corpus seems to be too weak to come to profound results in terms of the genre dependency. Unfortunately Weiner & Labov's (1983) findings cannot be transferred onto the given research, since they base their findings on spoken language and this thesis is based on written language.

Timing, the exam situation and the possible usage of dictionaries are irrelevant for four reasons. Neither of the three variables has previously been discussed in the literature, which is per se not a problem, but since the results of the regressions are not clear, a point of reference would be needed to better judge the findings. Also, neither variable is significant for all passives and none is significant at the one percent level. Chapter 4.2 has shown that the results of the linear and negative binomial regression deviate in terms of significance level and direction of β coefficients. The relative importance analysis based on the linear regression has revealed that this might be due to the comparatively low relative importance. In other words, the timing, exam and reference tool variables are so unimportant that neither of both regressions can cope with them and that they are ultimately irrelevant to the frequency of the passive.

Weiner & Labov (1983) indicate that the age of a speaker influences the frequency of the passives she is using. Descriptive statistics for the age variable again suggest that there is no profound foundation to compare my results with Labov's and Weiner's. As Chapter 4.1 shows, the vast majority of students is between 18 and 25 years old. Assuming that Weiner & Labov (1983) are right the effect needs to be strong in order to cause measurable effects within this seven-year frame. Again, the distribution of the variable does not allow for a generalisation beyond the statement that, in case of university students a few years more or less do not influence their grammatical behaviour.

Since the underlying corpus is based on speakers who are all within the educational system, the corresponding variables are of great interest. Although the educational systems of the various countries cannot be compared at this point, the general thesis is nonetheless that the more lessons a student receives, the greater are the chances that her English improves. Assuming on the basis of the linguistic literature (see Chapter 2) that a higher frequency of the passive is to be desired for academic writing, significant positive β coefficients are to be expected. The results of the regressions unfortunately indicate that neither more years of English at school, nor courses at the university or time abroad rise the frequency of the passive. From a didactic point of view this result is devastating, as it contradicts the idea of institutionalised education.

Two facts relativise this impression. First of all, the majority of students has not been abroad yet. Secondly, since the number of years within the educational system does not seem to have any kind of influence on the frequency of the passive, it can be assumed that other factors such as personal skills, the quality of the teachers and lecturers and the didactic concepts of the respective educational system are far more important than mere quantity of classes. In order to support this impression another type of corpora has to be used. Longitudinal data, i.e. multiple texts from the same students over time, would allow for an evaluation of the student's improvement.

The linguistic variables are a statistical necessity and therefore their concrete results are not important at this point. Since previous research has not taken these variables into account, a point of reference is missing. Judging from the results of Table 21 either the text length, the average sentence length or the Flesch-Kincaid Reading Ease score is enough to turn the absolute counts into relative frequencies. The Flesch-Kincaid Reading Ease score seems to be most promising as it contains the number of words, sentences and syllables and can easily be computed with available software (see Michalke 2014).

The key variable of this research is the L1 variable. Unfortunately there is no way to combine the results of the different regressions for the various L1/passive combinations into one table, because a 16 times 16 table with more than 100 items in each cell does not fit the page margins. Therefore, only the main findings will be discussed.

The first and most important finding is that the frequency of the passive is dependent on the language family, the specific L1 and the tense and aspect of the passive. This dependency is so specific that neither the uniform behaviour of language families nor the uniform behaviour of specific languages across either the tense or the aspect dimension is given. Up to this point my results go in line with Hinkel's (2004). Consequently, there can be no talk of passive voice as a uniform grammatical construction, but only of the passive with a specific tense/aspect combination. Additionally, the mutual influence of the tenses, aspects, modes and voices cannot be evaluated. Significant differences between different L1 combinations for one passive construction, but not for another might either be caused by the change in the passive construction or the change in tense and/or aspect. In the future a further analysis is therefore required.

The second finding is that the tense and aspect dependent results neutralise each other when combined. Only Dutch and Tswana students significantly deviate from this observation. At this point, the limitations of this paper become obvious. Actually, the results for all passive constructions need to be analysed in detail, to be able to come to generalisable conclusions. Due to the fact that other corpora and methods are used a more detailed comparison to Granger (2013) and Hinkel (2004) is not possible beyond the statement that the passive is L1 dependent and thereby influenced by transfer (Selinker 1972).

6 Summary and Outlook

This master's thesis is based on several maxims. There are grammatical categories which have been formulated centuries ago and that are still valid today. Extra-linguistic variables influence the frequency of linguistic phenomena. Corpora and statistics allow for a description of underlying general rules. They allow to predict, at least to a certain degree, the linguistic behaviour of people, if the above mentioned extra-linguistic variables are known. Based on these maxims, Chapter 3 has assembled a collection of methods and techniques to quantitatively describe the passive.

Chapter 4.1 has shown that methods and corpora have to be improved to come to meaningful results. Many variables are too narrow to achieve generalisable results that can be compared to contemporary linguistic literature (see Chapter 2). On the one hand the inclusion of the International Corpus Network of Asian Learners of English (ICNALE) would broaden the perspective beyond Europe, as well as it would allow for a comparison with NS (Ishikawa 2013). On the other hand, regression types, developed to deal with distributions with many zero values, can further improve the accuracy of the regression analyses.

Despite the use of computers, the collection of tense and aspect frequencies is still tedious. Since the single tools are already available, the development of an *R* package that automatically determines and counts various grammatical constructions, should be the next step.

According to the results of Chapter 4.2, the passive has to be characterised along three dimensions. For all passives, those analysed in this paper and those found in the appendix, the linguistic variables are the most important ones, followed by the L1 variable and the remaining variables. In each case the frequency and the fit decreases, as the complexity increases, i.e. the simple aspect is more frequent than the perfect one and the perfect aspect more frequent than the continuous. The L1 specific characteristics are dependent on the tense, aspect and mode combination of the passive and can therefore not be generalised. As a consequence, this paper does not claim completeness. It rather presents first impressions that need further investigations especially with those results that have not found their way into the paper, but can be found in the Appendix.

The quantitative analysis of linguistic phenomena needs to take another great leap forward. The path that has been started by collecting different types of corpora, has to be continued. The reason why I have not been able to compare my results with the current state of linguistics, is not due to a lack of literature on the passive, but caused by age of the literature that uses quantitative approaches.

As Chapter 3 has shown tools from economics, information technology, psychology, statistics and linguistics have to be combined to come to more precise findings. This paper has shown that mathematics and statistics can be beneficial for linguistic research to come to new insights.

References

Akaike, H. (1974), 'A new look at the statistical model identification', *Automatic Control, IEEE Transactions on* **19**(6), 716–723.

Anderwald, L. (2014), 'Measuring the success of prescriptivism: quantitative grammaticography, corpus linguistics and the progressive passive', *English Language and Linguistics* **18**(01), 1–21.

Armstrong, D. (2014), *factorplot: factorplot.* R package version 1.1-1.

Bauer, T. & Sinning, M. (2008), 'An extension of the Blinder Oaxaca decomposition to nonlinear models', *AStA Advances in Statistical Analysis* (49), 1–12.

Biber, D. (1992), 'On the complexity of discourse complexity: A multidimensional analysis', *Discourse Processes* **15**(2), 133–163.

Blevins, J. P. (2003), 'Passives and impersonals', *Journal of Linguistics* **39**(3), 473–520.

Bortz, J. (2010), *Statistik für Human- und Sozialwissenschaftler*, Springer, Berlin.

Briscoe, T. (2008), 'Language learning, power laws, and sexual selection', *Mind & Society* **7**(1), 65–76.

Burnham, K., Anderson, D. & Diggle, J. (2002), *Model Selection and Multimodel Inference*, Springer, New York.

Burnham, K. P., Anderson, D. R. & Huyvaert, K. P. (2010), 'AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons', *Behavioral Ecology and Sociobiology* **65**(1), 23–35.

Cameron, A. C. & Trivedi, P. K. (2005), *Microeconometrics: Methods and Applications*, Cambridge University Press, New York.

Darlington, R. (1968), 'Multiple Regression in Psychological Research and Practice', *Psychological Bulletin* **69**(3), 161–182.

Dobson, A. (2002), *An introduction to generalized linear models*, Chapman and Hall, Boca Raton.

Eddington, D. (2010), 'A comparison of two tools for analyzing linguistic data: logistic regression and decision trees', *Italian Journal of Linguistics* **2**(June 2009), 265–286.

Evert, S. & Baroni, M. (2007), 'zipfR: Word frequency distributions in R', *Proceedings of the 45th Annual Meeting of the ACL ...* (V).

Flesch, R. (1948), 'A new readability yardstick.', *Journal of Applied Psychology* **32**(3), 221–233.

Granger, S. (1983), *The be+ past participle construction in spoken English with special emphasis on the passive*, Elsevier Ltd, Amsterdam.

Granger, S. (1996), 'From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora'.

Granger, S. (1997), 'Automated retrieval of passives from native and learner corpora: precision and recall', *Journal of English Linguistics* **25**(4), 365–374.

Granger, S. (2003), 'The International Corpus of Learner English : A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research', *TESOL Quarterly* **37**(3), 538–546.

Granger, S. (2013), 'Communication à un colloque (Conference Paper) " The passive in learner English : Corpus insights and implications for pedagogical grammar " The passive in learner English Corpus insights and implications for pedagogical grammar'.

Gries, S. T. (2003), *Multifactorial Analysis in Corpus Linguistics*, Continuum, New York.

Gries, S. T. (2013), *Statistics for Linguistics with R*, Mouton de Gruyter, Boston, Berlin.

Grömping, U. (2006), 'Relative importance for linear regression in r: The package relaimpo', *Journal of Statistical Software* **17**(1), 1–27.

Hinkel, E. (2004), 'Tense, aspect and the passive voice in L1 and L2 academic texts', *Language Teaching Research* **8**(1), 5–29.

Huddleston, R. & Pullum, G. K. (2007), *A student's introduction to English grammar*, Vol. 84, Cambridge University Press, Cambridge.

Hundt, M. (2004), The Passival and the Progressive Passive: A Case Study of Layering in the English Aspect and Voice Systems, *in* H. Lindquist & C. Mair, eds, 'Corpus Approaches to Grammaticalization in English', John Benjamins, Amsterdam, pp. 79–120.

Ishikawa, S. (2013), 'The ICNALE and Sophisticated Contrastive Interlanguage Analysis of Asian Learners of English', *Learner Corpus Studies in Asia and the World* **1**, 91–118.

Jackman, S. (2012), *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University*, Department of Political Science, Stanford University, Stanford, California. R package version 1.04.4.

Jann, B. (2008), 'The Blinder-Oaxaca decomposition for linear regression models', *The Stata Journal* **8**(4), 453–479.

Kachru, B. (1992), 'World Englishes: Approaches, issues and resources', *Language teaching* (c).

Labov, W. (1972), '13. The Social Stratification of (R) in New York City Department Stores'.

Lancaster, T. (2004), *Introduction to Modern Bayesian Econometrics*, Wiley-Blackwell, Boston.

Langacker, R. & Munro, P. (1975), 'Passives and their meaning', *Language* **51**(4), 789–830.

Manning, C. D. (1999), *Foundations of Statistical Natural Language Processing*, Massachusetts Institute of Technology, Boston.

McCullagh, P. & Nelder, J. (1972), 'Generalized linear models.', *Journal of the Royal Statistical Society* **135**(3), 370–384.

McCullagh, P. & Nelder, J. (1989), *Generalized linear models.*, Chapman and Hall, London.

Michalke, M. (2014), *koRpus: An R Package for Text Analysis.* (Version 0.05-5).

Newman, M. E. J. (2004), 'Power laws, Pareto distributions and Zipf's law', (1), 28.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H. & Wagner, H. (2013), *vegan: Community Ecology Package.* R package version 2.0-10.

Puckica, J. (2009), 'Passive constructions in present-day English', *Groninger Arbeiten zur Germanistischen Linguistik* **49**(December), 215–235.

Pullum, G. K. & Huddleston, R. (2002), *The Cambridge Grammar of the English Language*, Cambridge University Press, Cambridge.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985), *A Comprehensive grammar of the English language*, Addison Wesley.

Rigby, R. A. & Stasinopoulos, D. M. (2005), 'Generalized additive models for location, scale and shape,(with discussion)', *Applied Statistics* **54**, 507–554.

Schmid, H.-J. (2011), *English Morphology and Word-formation: An Introduction*, Schmidt, Berlin.

Selinker, L. (1972), 'Interlanguage', *IRAL-International Review of Applied Linguistics in ...* **10**(3), 209–231.

Shapiro, S. & Wilk, M. (1965), 'An analysis of variance test for normality(complete samples).', *Journal of the American ...* **52**(3), 591–611.

Shapiro, S., Wilk, M. & Chen, H. (1968), 'A comparative study of various tests for normality', *Journal of the American ...* **63**(324), 1343–1372.

Sheskin, D. (2000), 'Parametric and nonparametric statistical procedures', *Boca Raton: CRC* .

Simpson, E. (1951), 'The interpretation of interaction in contingency tables', *Journal of the Royal Statistical Society. Series B (...* **13**(2), 238–241.

Tang, W., He, H. & Tu, X. (2012), *Applied Categorical and Count Data Analysis*, Chapman and Hall, Boca Raton.

Venables, W. N. & Ripley, B. D. (2002), *Modern Applied Statistics with S*, fourth edn, Springer, New York. ISBN 0-387-95457-0.

Warner, A. (1995), 'Predicting the progressive passive: Parametric change within a lexicalist framework', *Language* **71**(3), 533–557.

Weiner, E. & Labov, W. (1983), 'Constraints on the agentless passive', *Journal of linguistics* **19**(1), 29–58.

Winkelmann, R. (2000), *Count Data*, Springer.

Winkelmann, R. (2008), *Econometric Analysis of Count Data*, Springer Berlin Heidelberg, Berlin, Heidelberg.

Wooldridge, J. M. (2002), *Introductory Ecometrics*, South-Western.

Zeileis, A., Kleiber, C. & Jackman, S. (2008), 'Regression models for count data in R', *Journal of Statistical Software* **27**(8).

7 Appendix

7.1 Stanford NLP - Parser

Source Code 5: Example for a Parsed Sentence

```
<forest forestId="3" File="BGSU1002" TextId="BGSU1002"
Location="s5">
<div lang="eng">
 <seg/>
 </div>
 <div lang="org">
        <seg>The specialization begins quite late , the practice
        is inadequate and is not always up to the requirements
        that real work will set to the students after they leave
        the university .</seg>
      </div>
      <eTree Id="138" Label="S" from="1" to="180">
        <fs type="stat">
          <f name="prob" value="0,92" />
        </fs>
        <eTree Id="139" Label="NP" from="1" to="18">
          <fs type="stat">
            <f name="prob" value="1,00" />
          </fs>
          <eTree Id="140" Label="DT" from="1" to="3">
            <fs type="stat">
              <f name="prob" value="0,97" />
            </fs>
            <eLeaf Type="Vern" Text="The" />
          </eTree>
          <eTree Id="141" Label="NN" from="5" to="18">
            <fs type="stat">
              <f name="prob" value="0,99" />
            </fs>
            <eLeaf Type="Vern" Text="specialization" />
          </eTree>
        </eTree>
        <eTree Id="142" Label="VP" from="20" to="178">
```

```
<fs type="stat">
  <f name="prob" value="1,00" />
</fs>
<eTree Id="143" Label="VBZ" from="20" to="25">
  <fs type="stat">
    <f name="prob" value="0,93" />
  </fs>
  <eLeaf Type="Vern" Text="begins" />
</eTree>
<eTree Id="144" Label="ADVP" from="27" to="36">
  <fs type="stat">
    <f name="prob" value="1,00" />
  </fs>
  <eTree Id="145" Label="RB" from="27" to="31">
    <fs type="stat">
      <f name="prob" value="0,97" />
    </fs>
    <eLeaf Type="Vern" Text="quite" />
  </eTree>
  <eTree Id="146" Label="JJ" from="33" to="36">
    <fs type="stat">
      <f name="prob" value="0,69" />
    </fs>
    <eLeaf Type="Vern" Text="late" />
  </eTree>
</eTree>
<eTree Id="147" Label="," from="38" to="38">
  <fs type="stat">
    <f name="prob" value="0,99" />
  </fs>
  <eLeaf Type="Punct" Text="," />
</eTree>
      [...]
```

Simple			Perfect		
They	write	a letter	They	have written	a letter.
(NP)	VBP	(NP)	(NP)	VBP VBN	(NP)
He	writes	a letter.	He	has written	a letter.
(NP)	VBZ	(NP)	(NP)	VBZ VBN	(NP)
Progressive			**Perfect Progressive**		
They	are writing	a letter.	They	have been writing	a letter.
(NP)	VBP VBG	(NP)	(NP)	VBP VBN VBG	(NP)
He	is writing	a letter.	He	has been writing	a letter.
(NP)	VBZ VBG	(NP)	(NP)	VBZ VBN VBG	(NP)

Table 22: Present Active Indicative tense and aspect definitions with indicated Penn Treebank Tags.

Grammatical Definitions

Tense, Aspect, Mode and Voice Definitions

7.2 X-query Definitions

7.2.1 The Present Active

Source Code 6: Present Simple (Non-3rd Person Singular)

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”VBP”
and not($d/eTree/@Label=”VP”)
and not($d/eTree/eTree/@Label=”VP”)
and not($d/eTree/eTree/eTree/@Label=”VBG”)
and not($d/eTree/eTree/eTree/@Label=”VBN”)
return base−uri($d)
```

Source Code 7: Present Simple (3rd Person Singular)

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”VBZ”
and not($d/eTree/@Label=”VP”)
and not($d/eTree/eTree/@Label=”VP”)
and not($d/eTree/eTree/eTree/@Label=”VBG”)
and not($d/eTree/eTree/eTree/@Label=”VBN”)
return base−uri($d)
```

Source Code 8: Present Perfect (have)

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”VBP”
and $d/eTree/eLeaf/@Text=”have”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VBN”
and not($d/eTree/eTree/eLeaf/@Text=”been”)
return base−uri($d)
```

Source Code 9: Present Perfect (has)

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”VBZ”
and $d/eTree/eLeaf/@Text=”has”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VBN”
and not($d/eTree/eTree/eLeaf/@Text=”been”)
return base−uri($d)
```

Source Code 10: Present Progressive (are)

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”VBP”
and $d/eTree/eLeaf/@Text=”are”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VBG”
and not($d/eTree/eTree/@Label=”VBN”)
and not($d/eTree/eTree/@Label=”VP”)
and not($d/eTree/eTree/eTree/@Label=”VBG”)
return base−uri($d)
```

Source Code 11: Present Progressive (is)

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”VBZ”
and $d/eTree/eLeaf/@Text=”is”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VBG”
and not($d/eTree/eTree/@Label=”VBN”)
and not($d/eTree/eTree/@Label=”VP”)
and not($d/eTree/eTree/eTree/@Label=”VBG”)
```

```
return base-uri($d)
```

Source Code 12: Present Perfect Progressive (have)

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="VBP"
and $d/eTree/eLeaf/@Text="have"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VBN"
and $d/eTree/eTree/eLeaf/@Text="been"
and $d/eTree/eTree/@Label="VP"
and $d/eTree/eTree/eTree/@Label="VBG"
return base-uri($d)
```

Source Code 13: Present Perfect Progressive Indicative (has)

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="VBZ"
and $d/eTree/eLeaf/@Text="has"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VBN"
and $d/eTree/eTree/eLeaf/@Text="been"
and $d/eTree/eTree/@Label="VP"
and $d/eTree/eTree/eTree/@Label="VBG"
return base-uri($d)
```

7.2.2 The Past Active

Source Code 14: Past Simple

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="VBD"
and not($d/eTree/@Label="VP")
and not($d/eTree/eTree/@Label="VP")
and not($d/eTree/eTree/eTree/@Label="VBG")
and not($d/eTree/eTree/eTree/@Label="VBN")
return $d
```

Source Code 15: Past Perfect

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="VBD"
and $d/eTree/eLeaf/@Text="had"
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VP”
and $d/eTree/eTree/eTree/@Label=”VBN”
and not($d/eTree/eTree/eLeaf/@Text=”been”)
return $d
```

Source Code 16: Past Progressive

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”VBD”
and not($d/eTree/eLeaf/@Text=”had”)
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VP”
and $d/eTree/eTree/eTree/@Label=”VBG”
and not($d/eTree/eTree/eLeaf/@Text=”been”)
return $d
```

Source Code 17: Past Perfect Progressive

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”VBD”
and $d/eTree/eLeaf/@Text=”had”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VP”
and $d/eTree/eTree/eTree/@Label=”VBN”
and $d/eTree/eTree/eLeaf/@Text=”been”
and $d/eTree/eTree/eTree/eTree/@Label=”VBG”
return $d
```

7.2.3 The Future Active

Source Code 18: Going-to Future (is)

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”VBZ”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VBG”
and $d/eTree/eTree/eLeaf/@Text=”going”
return $d
```

Source Code 19: Going-to Future (are)

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
```

```
where $d/eTree/@Label=”VBP”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VBG”
and $d/eTree/eTree/eLeaf/@Text=”going”
return $d
```

Source Code 20: Will Future Simple

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”MD”
and $d/eTree/eLeaf/@Text=”will”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VB”
and not($d/eTree/eTree/eTree/@Label=”VBG”)
return $d
```

Source Code 21: Will Future Perfect

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”MD”
and $d/eTree/eLeaf/@Text=”will”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VB”
and $d/eTree/eTree/eLeaf/@Text=”have”
and $d/eTree/eTree/eTree/@Label=”VBN”
return $d
```

Source Code 22: Will Future Progressive

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”MD”
and $d/eTree/eLeaf/@Text=”will”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VB”
and $d/eTree/eTree/@Label=”VP”
and $d/eTree/eTree/eTree/@Label=”VBG”
return $d
```

Source Code 23: Will Future Perfect Progressive

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”MD”
```

```
and $d/eTree/eLeaf/@Text="will"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VB"
and $d/eTree/eTree/eLeaf/@Text="have"
and $d/eTree/eTree/eTree/@Label="VBN"
and $d/eTree/eTree/eTree/eLeaf/@Text="been"
and $d/eTree/eTree/eTree/eTree/@Label="VBG"
return $d
```

7.2.4 The Present Passive

Source Code 24: Present Simple (Non-3rd Person Singular)

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="VBP"
and $d/eTree/eLeaf/@Text="are"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VBN"
and not($d/eTree/eTree/eLeaf/@Text="been")
return $d
```

Source Code 25: Present Simple (3rd Person Singular)

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="VBZ"
and $d/eTree/eLeaf/@Text="is"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VBN"
and not($d/eTree/eTree/eLeaf/@Text="been")
return $d
```

Source Code 26: Present Perfect (have)

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="VBP"
and $d/eTree/eLeaf/@Text="have"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VBN"
and $d/eTree/eTree/eLeaf/@Text="been"
and $d/eTree/eTree/eTree/@Label="VBN"
return $d
```

Source Code 27: Present Perfect (has)

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”VBZ”
and $d/eTree/eLeaf/@Text=”has”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VBN”
and $d/eTree/eTree/eLeaf/@Text=”been”
and $d/eTree/eTree/eTree/@Label=”VBN”
return $d
```

Source Code 28: Present Progressive (are)

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”VBP”
and $d/eTree/eLeaf/@Text=”are”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VBG”
and $d/eTree/eTree/eLeaf/@Text=”being”
and $d/eTree/eTree/@Label=”VP”
and $d/eTree/eTree/eTree/@Label=”VBN”
return $d
```

Source Code 29: Present Progressive (is)

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”VBZ”
and $d/eTree/eLeaf/@Text=”is”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VBG”
and $d/eTree/eTree/eLeaf/@Text=”being”
and $d/eTree/eTree/@Label=”VP”
and $d/eTree/eTree/eTree/@Label=”VBN”
return $d
```

Source Code 30: Present Perfect Progressive (have)

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”VBP”
and $d/eTree/eLeaf/@Text=”have”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VBN”
and $d/eTree/eTree/eLeaf/@Text="been"
and $d/eTree/eTree/@Label="VP"
and $d/eTree/eTree/eTree/@Label="VBG"
and $d/eTree/eTree/eTree/eLeaf/@Text="being"
and $d/eTree/eTree/eTree/eTree/@Label="VBN"
return $d
```

Source Code 31: Present Perfect Progressive (has)

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="VBZ"
and $d/eTree/eLeaf/@Text="has"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VBN"
and $d/eTree/eTree/eLeaf/@Text="been"
and $d/eTree/eTree/@Label="VP"
and $d/eTree/eTree/eTree/@Label="VBG"
and $d/eTree/eTree/eTree/eLeaf/@Text="being"
and $d/eTree/eTree/eTree/eTree/@Label="VBN"
return $d
```

7.2.5 The Past Passive

Source Code 32: Past Simple

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="VBD"
and not($d/eTree/eLeaf/@Text="had")
and $d/eTree/@Label="VP"
and not($d/eTree/eTree/@Label="VP")
and not($d/eTree/eTree/eTree/@Label="VBG")
and $d/eTree/eTree/@Label="VBN"
return $d
```

Source Code 33: Past Perfect

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="VBD"
and $d/eTree/eLeaf/@Text="had"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VP"
and not($d/eTree/eTree/eTree/@Label="VBG")
and $d/eTree/eTree/@Label="VBN"
and $d/eTree/eTree/eLeaf/@Text="been"
and $d/eTree/eTree/eTree/@Label="VBN"
return $d
```

Source Code 34: Past Perfect Progressive

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="VBD"
and $d/eTree/eLeaf/@Text="had"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VP"
and not($d/eTree/eTree/eTree/@Label="VBG")
and $d/eTree/eTree/@Label="VBN"
and $d/eTree/eTree/eLeaf/@Text="been"
and $d/eTree/eTree/eTree/@Label="VBG"
and $d/eTree/eTree/eTree/eLeaf/@Text="being"
and $d/eTree/eTree/eTree/eTree/@Label="VBN"
return $d
```

7.2.6 The Future Passive

Source Code 35: Going-to Future (are)

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="VBP"
and $d/eTree/eLeaf/@Text="are"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VBG"
and $d/eTree/eTree/eLeaf/@Text="going"
and $d/eTree/eTree/@Label="S"
and $d/eTree/eTree/eTree/eTree/eTree/@Label="VB"
and $d/eTree/eTree/eTree/eTree/eTree/eTree/@Label="VBN"
return $d
```

Source Code 36: Going-to Future (is)

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="VBZ"
and $d/eTree/eLeaf/@Text="is"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VBG"
and $d/eTree/eTree/eLeaf/@Text="going"
and $d/eTree/eTree/@Label="S"
and $d/eTree/eTree/eTree/eTree/eTree/@Label="VB"
and $d/eTree/eTree/eTree/eTree/eTree/eTree/@Label="VBN"
return $d
```

Source Code 37: Will Future Simple

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="MD"
and $d/eTree/eLeaf/@Text="will"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VB"
and $d/eTree/eTree/eLeaf/@Text="be"
and $d/eTree/eTree/eTree/@Label="VBN"
return $d
```

Source Code 38: Will Future Perfect

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="MD"
and $d/eTree/eLeaf/@Text="will"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VB"
and $d/eTree/eTree/eLeaf/@Text="have"
and $d/eTree/eTree/eTree/@Label="VBN"
and $d/eTree/eTree/eTree/eLeaf/@Text="been"
return $d
```

Source Code 39: Will Future Progressive

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="MD"
and $d/eTree/eLeaf/@Text="will"
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VB"
and $d/eTree/eTree/eLeaf/@Text="be"
and $d/eTree/eTree/eTree/@Label="VBG"
and $d/eTree/eTree/eTree/eLeaf/@Text="being"
and $d/eTree/eTree/eTree/eTree/@Label="VBN"
return $d
```

Source Code 40: Will Future Perfect Progressive

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”MD”
and $d/eTree/eLeaf/@Text=”will”
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VB”
and $d/eTree/eTree/eLeaf/@Text=”have”
and $d/eTree/eTree/eTree/@Label=”VBN”
and $d/eTree/eTree/eTree/eLeaf/@Text=”been”
and $d/eTree/eTree/eTree/eTree/@Label=”VBG”
and $d/eTree/eTree/eTree/eTree/eLeaf/@Text=”being”
and $d/eTree/eTree/eTree/eTree/eTree/@Label=”VBN”
return $d
```

7.2.7 The Conditional Active

Source Code 41: Conditional Simple

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”MD”
and not($d/eTree/eLeaf/@Text=”will”)
and not($d/eTree/eLeaf/@Text=”can”)
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VB”
and not ($d/eTree/eTree/eLeaf/@Text=”have”)
and not($d/eTree/eTree/eTree/@Label=”VBG”)
return $d
```

Source Code 42: Conditional Perfect

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”MD”
and not($d/eTree/eLeaf/@Text=”will”)
and not($d/eTree/eLeaf/@Text=”can”)
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VB”
and $d/eTree/eTree/eLeaf/@Text=”have”
and $d/eTree/eTree/eTree/@Label=”VBN”
return $d
```

Source Code 43: Conditional Progressive

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”MD”
and not($d/eTree/eLeaf/@Text=”will”)
and not($d/eTree/eLeaf/@Text=”can”)
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VB”
and $d/eTree/eTree/eLeaf/@Text=”be”
and $d/eTree/eTree/eTree/@Label=”VBG”
return $d
```

Source Code 44: Conditional Perfect Progressive

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”MD”
and not($d/eTree/eLeaf/@Text=”will”)
and not($d/eTree/eLeaf/@Text=”can”)
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VB”
and $d/eTree/eTree/eLeaf/@Text=”have”
and $d/eTree/eTree/eTree/@Label=”VBN”
and $d/eTree/eTree/eTree/eTree/@Label=”VBG”
return $d
```

7.2.8 The Conditional Passive

Source Code 45: Conditional Simple

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
where $d/eTree/@Label=”MD”
and not($d/eTree/eLeaf/@Text=”will”)
and not($d/eTree/eLeaf/@Text=”can”)
and $d/eTree/@Label=”VP”
and $d/eTree/eTree/@Label=”VB”
and $d/eTree/eTree/eLeaf/@Label=”be”)
and not ($d/eTree/eTree/eLeaf/@Text=”have”)
and not($d/eTree/eTree/eTree/@Label=”VBG”)
and $d/eTree/eTree/eTree/@Label=”VBN”
return $d
```

Source Code 46: Conditional Perfect

```
for $d in collection(”ICLE”)//eTree[@Label=”S”]/eTree[@Label=”VP”]
```

```
where $d/eTree/@Label="MD"
and not($d/eTree/eLeaf/@Text="will")
and not($d/eTree/eLeaf/@Text="can")
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VB"
and $d/eTree/eTree/eLeaf/@Label="have")
and not($d/eTree/eTree/eTree/@Label="VBG")
and $d/eTree/eTree/eTree/@Label="VBN"
and $d/eTree/eTree/eTree/eLeaf/@Text="been"
and $d/eTree/eTree/eTree/eTree/@Label="VBN"
return $d
```

Source Code 47: Conditional Perfect Progressive

```
for $d in collection("ICLE")//eTree[@Label="S"]/eTree[@Label="VP"]
where $d/eTree/@Label="MD"
and not($d/eTree/eLeaf/@Text="will")
and not($d/eTree/eLeaf/@Text="can")
and $d/eTree/@Label="VP"
and $d/eTree/eTree/@Label="VB"
and $d/eTree/eTree/eLeaf/@Label="have")
and not($d/eTree/eTree/eTree/@Label="VBG")
and $d/eTree/eTree/eTree/@Label="VBN"
and $d/eTree/eTree/eTree/eLeaf/@Text="been"
and $d/eTree/eTree/eTree/eTree/@Label="VBG"
and $d/eTree/eTree/eTree/eTree/eLeaf/@Text="being"
and $d/eTree/eTree/eTree/eTree/eTree/@Label="VBN"
return $d
```

7.3 Additional Probability Density and Mass Functions for Empirical Distributions

Cauchy distribution:

$$p(x; \lambda, \mu) = \frac{1}{\pi} \frac{\lambda}{\lambda^2 + (x - \mu)^2} \tag{34}$$

Exponential Distribution:

$$p(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \\ 0 & x \leq 0 \end{cases} \tag{35}$$

Logarithmic Normal Distribution:

$$p(x) = \begin{cases} \frac{log\ \ e}{\sigma x \sqrt{2\pi}} e^{\frac{(log\ \ (x) - \mu)^2}{2\sigma^2}} & x > 0 \\ 0 & x \leq 0 \end{cases} \tag{36}$$

Poisson Distribution for non-Negative Integer:

$$p(x = k) = e^{-\lambda} \frac{\lambda^k}{k!} \tag{37}$$

Negative Binomial Distribution:

$$p(x = k) = \binom{r + k - 1}{k} p^r (1 - p)^k \tag{38}$$

7.4 Additional Plots, Tables and Matrices

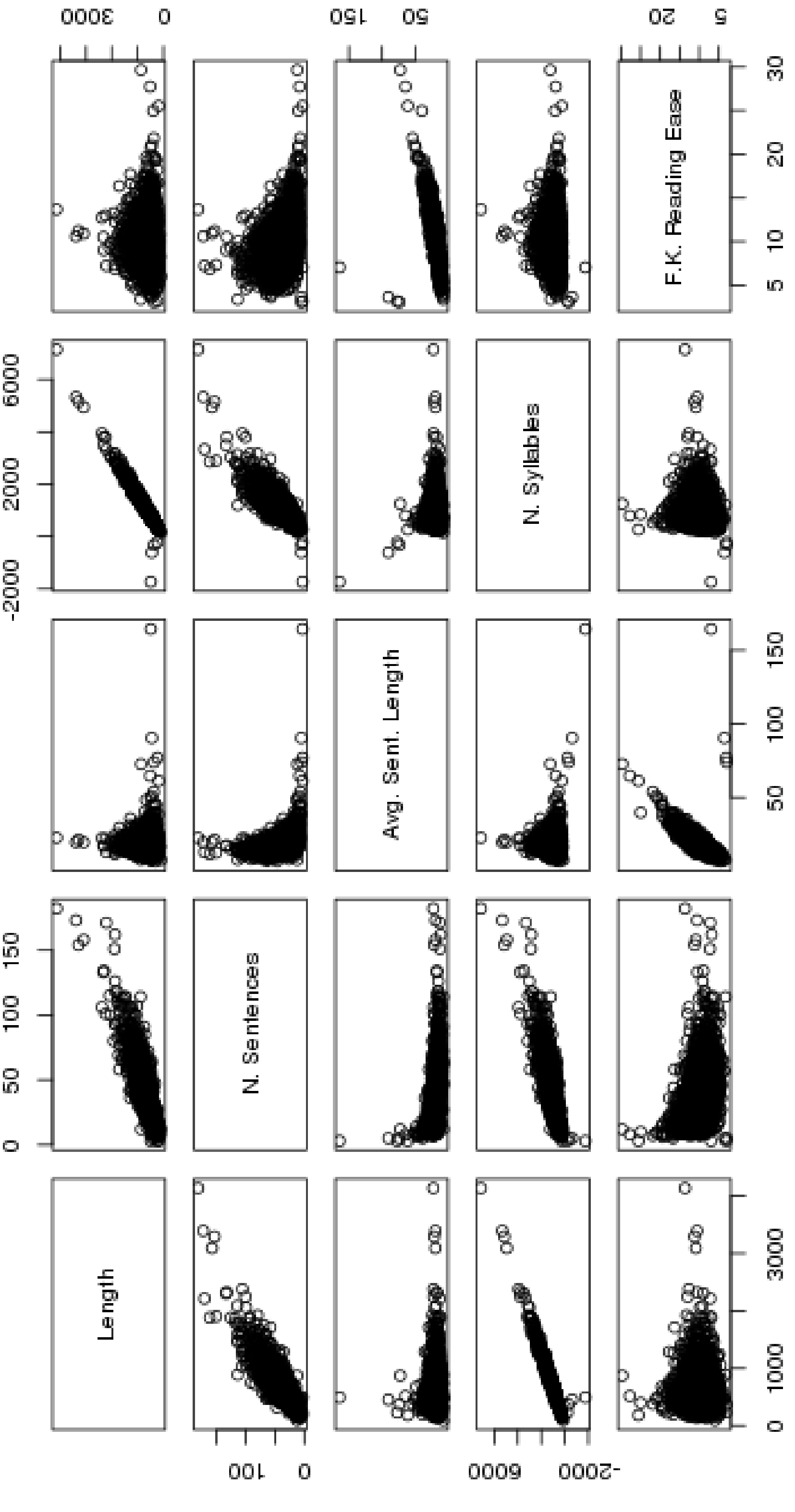


Figure 16: Scatterplot Matrices of Metric Independent Linguistics Variables.

	Length	N. Sentences	Avg. Sent. Length	N. Syllables	F.K. Reading Ease
Length	1	0.857	−0.011	0.982	0.028
N. Sentences	0.857	1	-0.399	0.826	-0.365
Avg. Sent. Length	−0.011	−0.399	1	−0.046	0.707
N. Syllables	0.982	0.826	−0.046	1	0.123
F.K. Reading Ease	0.028	−0.365	0.707	0.123	1

Table 23: Correlation Matrix of Metric Independent Linguistic Variables.

7.5 Additional Pairwise L1 Differences

Unfortunately not all plots can be shown at this point. To avoid redundancy combined constructions have been omitted. In some cases the standard error are that large that they look like as if the programme did not work correctly. Yet, to print the plots that large that the standard errors fit into their boxes would go beyond the margins of the paper.1

7.5.1 The Present Active

Out of 13 possible Tense and Aspect combinations the pairwise differences in L1 parameter values for 13 combinations could have been plotted. To assure readability the plots of the combined constructions have been omitted.

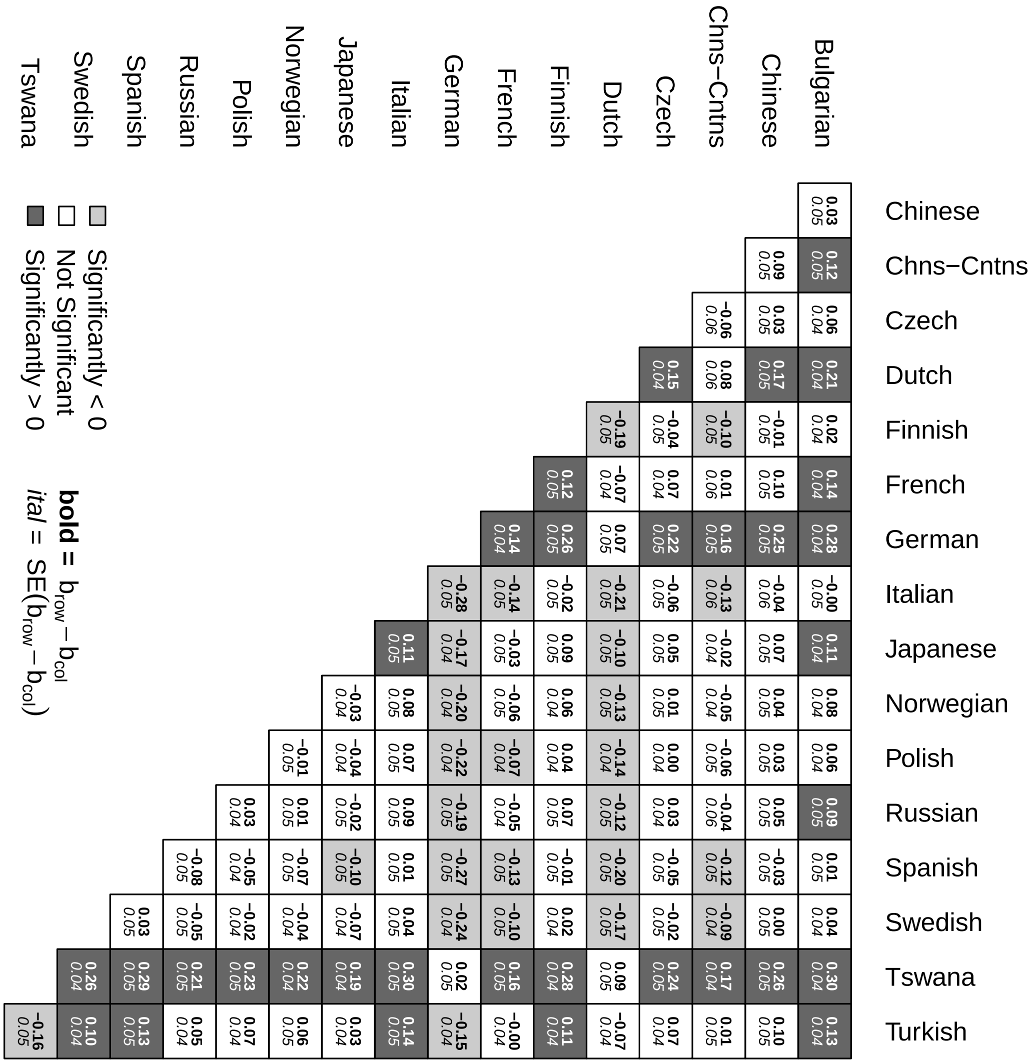
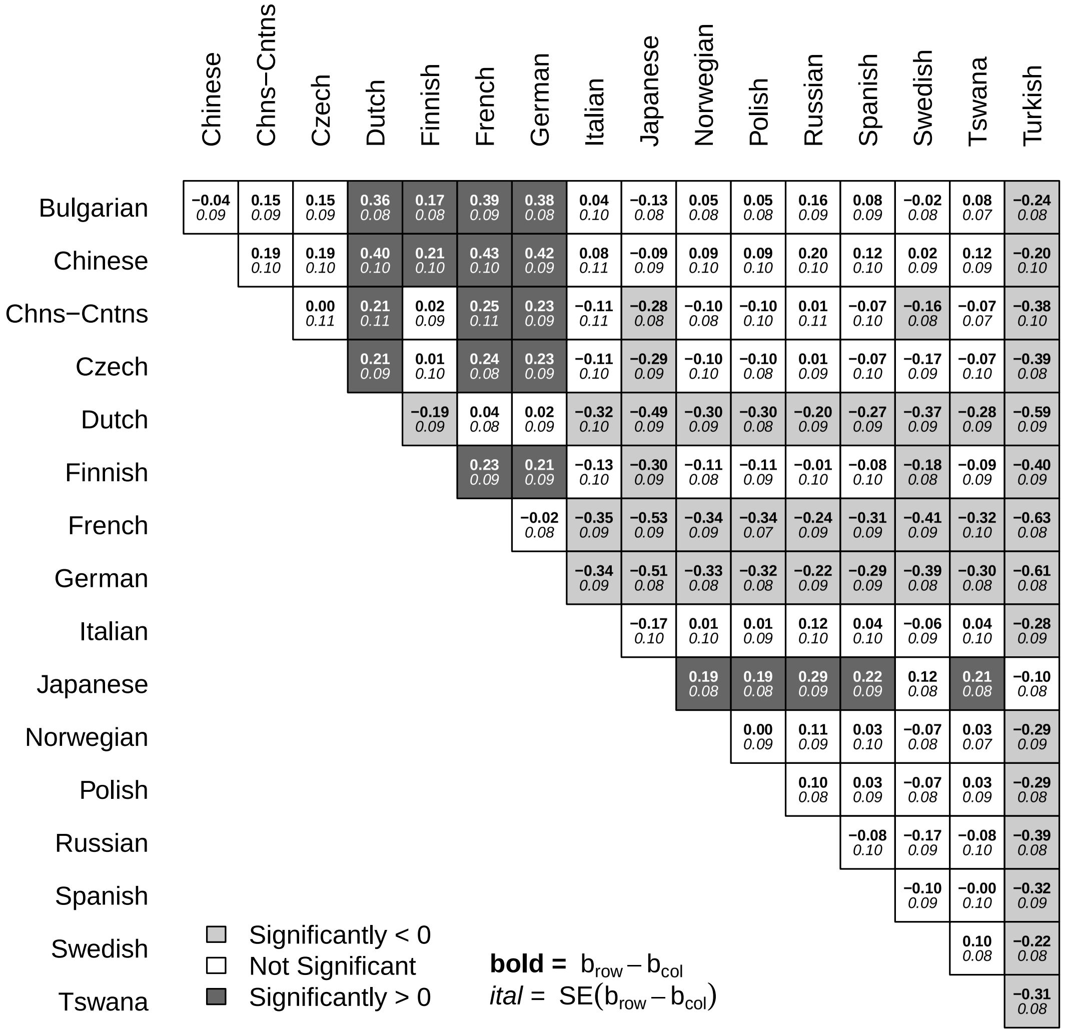


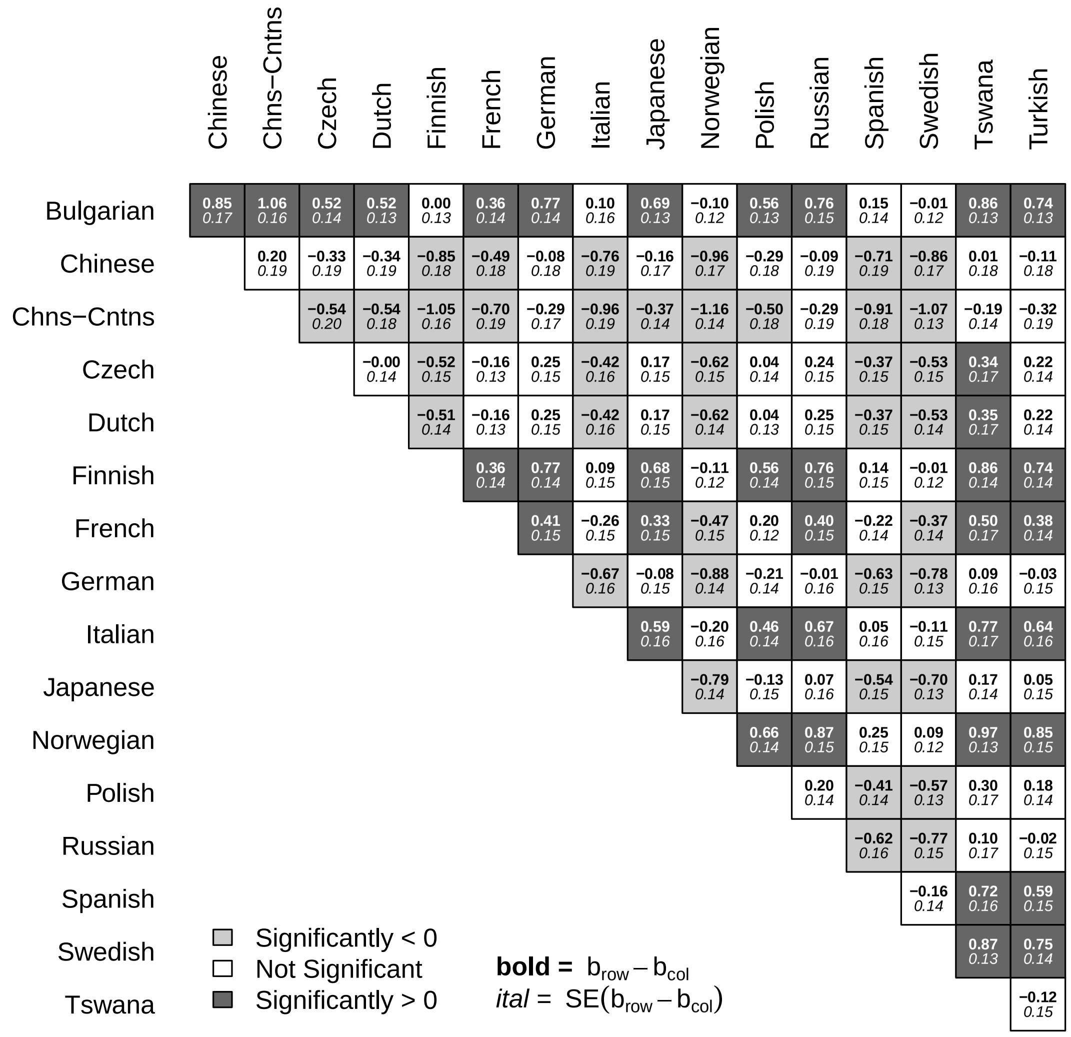
Figure 17: Pairwise Differences in L1 Parameter Values for Present Simple 3ps Neg. Binomial Type I Regression.



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**−0.04** *0.09*	**0.15** *0.09*	**0.15** *0.09*	**0.36** *0.08*	**0.17** *0.08*	**0.39** *0.09*	**0.38** *0.08*	**0.04** *0.10*	**−0.13** *0.08*	**0.05** *0.08*	**0.05** *0.08*	**0.16** *0.09*	**0.08** *0.09*	**−0.02** *0.08*	**0.08** *0.07*	**−0.24** *0.08*
Chinese		**0.19** *0.10*	**0.19** *0.10*	**0.40** *0.10*	**0.21** *0.10*	**0.43** *0.10*	**0.42** *0.09*	**0.08** *0.11*	**−0.09** *0.09*	**0.09** *0.10*	**0.09** *0.10*	**0.20** *0.10*	**0.12** *0.10*	**0.02** *0.09*	**0.12** *0.09*	**−0.20** *0.10*
Chns−Cntns			**0.00** *0.11*	**0.21** *0.11*	**0.02** *0.09*	**0.25** *0.11*	**0.23** *0.09*	**−0.11** *0.11*	**−0.28** *0.08*	**−0.10** *0.08*	**−0.10** *0.10*	**0.01** *0.11*	**−0.07** *0.10*	**−0.16** *0.08*	**−0.07** *0.07*	**−0.38** *0.10*
Czech				**0.21** *0.09*	**0.01** *0.10*	**0.24** *0.08*	**0.23** *0.09*	**−0.11** *0.10*	**−0.29** *0.09*	**−0.10** *0.10*	**−0.10** *0.08*	**0.01** *0.09*	**−0.07** *0.10*	**−0.17** *0.09*	**−0.07** *0.10*	**−0.39** *0.08*
Dutch					**−0.19** *0.09*	**0.04** *0.08*	**0.02** *0.09*	**−0.32** *0.10*	**−0.49** *0.09*	**−0.30** *0.09*	**−0.30** *0.08*	**−0.20** *0.09*	**−0.27** *0.09*	**−0.37** *0.09*	**−0.28** *0.09*	**−0.59** *0.09*
Finnish						**0.23** *0.09*	**0.21** *0.09*	**−0.13** *0.10*	**−0.30** *0.09*	**−0.11** *0.08*	**−0.11** *0.09*	**−0.01** *0.10*	**−0.08** *0.10*	**−0.18** *0.08*	**−0.09** *0.09*	**−0.40** *0.09*
French							**−0.02** *0.08*	**−0.35** *0.09*	**−0.53** *0.09*	**−0.34** *0.09*	**−0.34** *0.07*	**−0.24** *0.09*	**−0.31** *0.09*	**−0.41** *0.09*	**−0.32** *0.10*	**−0.63** *0.08*
German								**−0.34** *0.09*	**−0.51** *0.08*	**−0.33** *0.08*	**−0.32** *0.08*	**−0.22** *0.09*	**−0.29** *0.09*	**−0.39** *0.08*	**−0.30** *0.08*	**−0.61** *0.08*
Italian									**−0.17** *0.10*	**0.01** *0.10*	**0.01** *0.09*	**0.12** *0.10*	**0.04** *0.10*	**−0.06** *0.09*	**0.04** *0.10*	**−0.28** *0.09*
Japanese										**0.19** *0.08*	**0.19** *0.08*	**0.29** *0.09*	**0.22** *0.09*	**0.12** *0.08*	**0.21** *0.08*	**−0.10** *0.08*
Norwegian											**0.00** *0.09*	**0.11** *0.09*	**0.03** *0.10*	**−0.07** *0.08*	**0.03** *0.07*	**−0.29** *0.09*
Polish												**0.10** *0.08*	**0.03** *0.09*	**−0.07** *0.08*	**0.03** *0.09*	**−0.29** *0.08*
Russian													**−0.08** *0.10*	**−0.17** *0.09*	**−0.08** *0.10*	**−0.39** *0.08*
Spanish														**−0.10** *0.09*	**−0.00** *0.10*	**−0.32** *0.09*
Swedish															**0.10** *0.08*	**−0.22** *0.08*
Tswana																**−0.31** *0.08*

▒ Significantly < 0
□ Not Significant
■ Significantly > 0
bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 18: Pairwise Differences in L1 Parameter Values for Present Simple non 3ps Neg. Binomial Type I Regression.

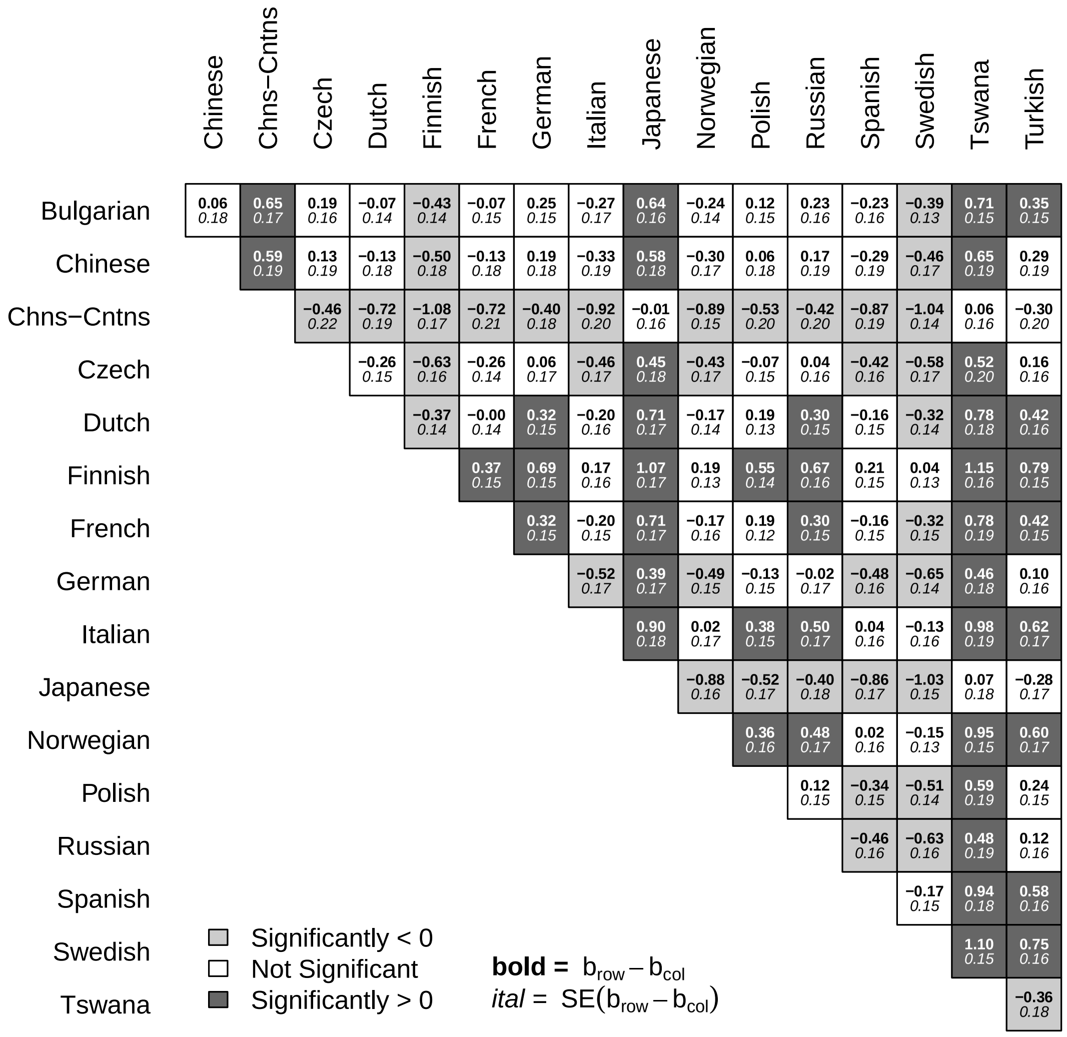


	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**0.85** *0.17*	**1.06** *0.16*	**0.52** *0.14*	**0.52** *0.13*	**0.00** *0.13*	**0.36** *0.14*	**0.77** *0.14*	**0.10** *0.16*	**0.69** *0.13*	**−0.10** *0.12*	**0.56** *0.13*	**0.76** *0.15*	**0.15** *0.14*	**−0.01** *0.12*	**0.86** *0.13*	**0.74** *0.13*
Chinese		**0.20** *0.19*	**−0.33** *0.19*	**−0.34** *0.19*	**−0.85** *0.18*	**−0.49** *0.18*	**−0.08** *0.18*	**−0.76** *0.19*	**−0.16** *0.17*	**−0.96** *0.17*	**−0.29** *0.18*	**−0.09** *0.19*	**−0.71** *0.19*	**−0.86** *0.17*	**0.01** *0.18*	**−0.11** *0.18*
Chns−Cntns			**−0.54** *0.20*	**−0.54** *0.18*	**−1.05** *0.16*	**−0.70** *0.19*	**−0.29** *0.17*	**−0.96** *0.19*	**−0.37** *0.14*	**−1.16** *0.14*	**−0.50** *0.18*	**−0.29** *0.19*	**−0.91** *0.18*	**−1.07** *0.13*	**−0.19** *0.14*	**−0.32** *0.19*
Czech				**−0.00** *0.14*	**−0.52** *0.15*	**−0.16** *0.13*	**0.25** *0.15*	**−0.42** *0.16*	**0.17** *0.15*	**−0.62** *0.15*	**0.04** *0.14*	**0.24** *0.15*	**−0.37** *0.15*	**−0.53** *0.15*	**0.34** *0.17*	**0.22** *0.14*
Dutch					**−0.51** *0.14*	**−0.16** *0.13*	**0.25** *0.15*	**−0.42** *0.16*	**0.17** *0.15*	**−0.62** *0.14*	**0.04** *0.13*	**0.25** *0.15*	**−0.37** *0.15*	**−0.53** *0.14*	**0.35** *0.17*	**0.22** *0.14*
Finnish						**0.36** *0.14*	**0.77** *0.14*	**0.09** *0.15*	**0.68** *0.15*	**−0.11** *0.12*	**0.56** *0.14*	**0.76** *0.15*	**0.14** *0.15*	**−0.01** *0.12*	**0.86** *0.14*	**0.74** *0.14*
French							**0.41** *0.15*	**−0.26** *0.15*	**0.33** *0.15*	**−0.47** *0.15*	**0.20** *0.12*	**0.40** *0.15*	**−0.22** *0.14*	**−0.37** *0.14*	**0.50** *0.17*	**0.38** *0.14*
German								**−0.67** *0.16*	**−0.08** *0.15*	**−0.88** *0.14*	**−0.21** *0.14*	**−0.01** *0.16*	**−0.63** *0.15*	**−0.78** *0.13*	**0.09** *0.16*	**−0.03** *0.15*
Italian									**0.59** *0.16*	**−0.20** *0.16*	**0.46** *0.14*	**0.67** *0.16*	**0.05** *0.16*	**−0.11** *0.15*	**0.77** *0.17*	**0.64** *0.16*
Japanese										**−0.79** *0.14*	**−0.13** *0.15*	**0.07** *0.16*	**−0.54** *0.15*	**−0.70** *0.13*	**0.17** *0.14*	**0.05** *0.15*
Norwegian											**0.66** *0.14*	**0.87** *0.15*	**0.25** *0.15*	**0.09** *0.12*	**0.97** *0.13*	**0.85** *0.15*
Polish												**0.20** *0.14*	**−0.41** *0.14*	**−0.57** *0.13*	**0.30** *0.17*	**0.18** *0.14*
Russian													**−0.62** *0.16*	**−0.77** *0.15*	**0.10** *0.17*	**−0.02** *0.15*
Spanish														**−0.16** *0.14*	**0.72** *0.16*	**0.59** *0.15*
Swedish															**0.87** *0.13*	**0.75** *0.14*
Tswana																**−0.12** *0.15*

□ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 19: Pairwise Differences in L1 Parameter Values for Present Perfect (have) Neg. Binomial Type I Regression.

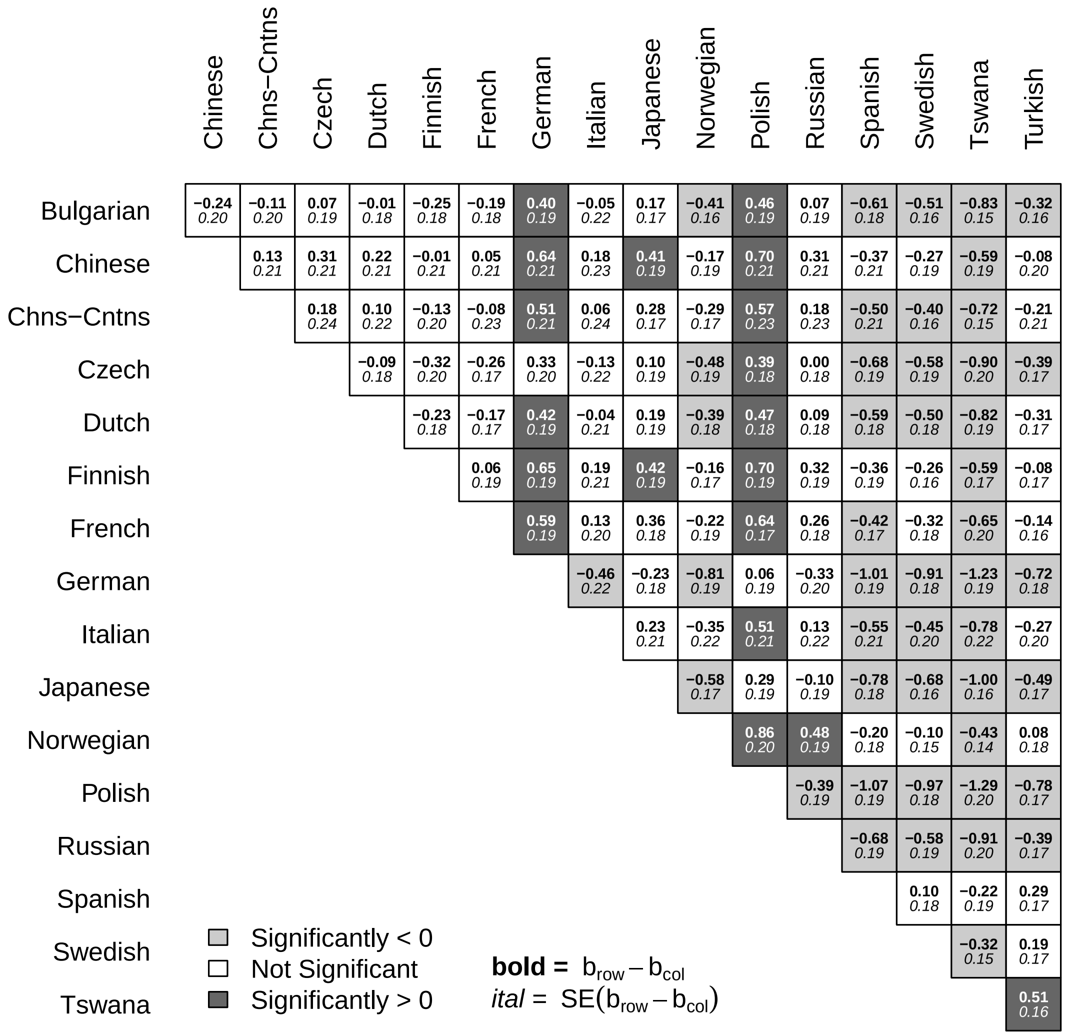


	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**0.06** *0.18*	**0.65** *0.17*	**0.19** *0.16*	**−0.07** *0.14*	**−0.43** *0.14*	**−0.07** *0.15*	**0.25** *0.15*	**−0.27** *0.17*	**0.64** *0.16*	**−0.24** *0.14*	**0.12** *0.15*	**0.23** *0.16*	**−0.23** *0.16*	**−0.39** *0.13*	**0.71** *0.15*	**0.35** *0.15*
Chinese		**0.59** *0.19*	**0.13** *0.19*	**−0.13** *0.18*	**−0.50** *0.18*	**−0.13** *0.18*	**0.19** *0.18*	**−0.33** *0.19*	**0.58** *0.18*	**−0.30** *0.17*	**0.06** *0.18*	**0.17** *0.19*	**−0.29** *0.19*	**−0.46** *0.17*	**0.65** *0.19*	**0.29** *0.19*
Chns−Cntns			**−0.46** *0.22*	**−0.72** *0.19*	**−1.08** *0.17*	**−0.72** *0.21*	**−0.40** *0.18*	**−0.92** *0.20*	**−0.01** *0.16*	**−0.89** *0.15*	**−0.53** *0.20*	**−0.42** *0.20*	**−0.87** *0.19*	**−1.04** *0.14*	**0.06** *0.16*	**−0.30** *0.20*
Czech				**−0.26** *0.15*	**−0.63** *0.16*	**−0.26** *0.14*	**0.06** *0.17*	**−0.46** *0.17*	**0.45** *0.18*	**−0.43** *0.17*	**−0.07** *0.15*	**0.04** *0.16*	**−0.42** *0.16*	**−0.58** *0.17*	**0.52** *0.20*	**0.16** *0.16*
Dutch					**−0.37** *0.14*	**−0.00** *0.14*	**0.32** *0.15*	**−0.20** *0.16*	**0.71** *0.17*	**−0.17** *0.14*	**0.19** *0.13*	**0.30** *0.15*	**−0.16** *0.15*	**−0.32** *0.14*	**0.78** *0.18*	**0.42** *0.16*
Finnish						**0.37** *0.15*	**0.69** *0.15*	**0.17** *0.16*	**1.07** *0.17*	**0.19** *0.13*	**0.55** *0.14*	**0.67** *0.16*	**0.21** *0.15*	**0.04** *0.13*	**1.15** *0.16*	**0.79** *0.15*
French							**0.32** *0.15*	**−0.20** *0.15*	**0.71** *0.17*	**−0.17** *0.16*	**0.19** *0.12*	**0.30** *0.15*	**−0.16** *0.15*	**−0.32** *0.15*	**0.78** *0.19*	**0.42** *0.15*
German								**−0.52** *0.17*	**0.39** *0.17*	**−0.49** *0.15*	**−0.13** *0.15*	**−0.02** *0.17*	**−0.48** *0.16*	**−0.65** *0.14*	**0.46** *0.18*	**0.10** *0.16*
Italian									**0.90** *0.18*	**0.02** *0.17*	**0.38** *0.15*	**0.50** *0.17*	**0.04** *0.16*	**−0.13** *0.16*	**0.98** *0.19*	**0.62** *0.17*
Japanese										**−0.88** *0.16*	**−0.52** *0.17*	**−0.40** *0.18*	**−0.86** *0.17*	**−1.03** *0.15*	**0.07** *0.18*	**−0.28** *0.17*
Norwegian											**0.36** *0.16*	**0.48** *0.17*	**0.02** *0.16*	**−0.15** *0.13*	**0.95** *0.15*	**0.60** *0.17*
Polish												**0.12** *0.15*	**−0.34** *0.15*	**−0.51** *0.14*	**0.59** *0.19*	**0.24** *0.15*
Russian													**−0.46** *0.16*	**−0.63** *0.16*	**0.48** *0.19*	**0.12** *0.16*
Spanish														**−0.17** *0.15*	**0.94** *0.18*	**0.58** *0.16*
Swedish															**1.10** *0.15*	**0.75** *0.16*
Tswana																**−0.36** *0.18*

▒ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 20: Pairwise Differences in L1 Parameter Values for Present Perfect (has) Neg. Binomial Type I Regression.



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**−0.24** *0.20*	**−0.11** *0.20*	**0.07** *0.19*	**−0.01** *0.18*	**−0.25** *0.18*	**−0.19** *0.18*	**0.40** *0.19*	**−0.05** *0.22*	**0.17** *0.17*	**−0.41** *0.16*	**0.46** *0.19*	**0.07** *0.19*	**−0.61** *0.18*	**−0.51** *0.16*	**−0.83** *0.15*	**−0.32** *0.16*
Chinese		**0.13** *0.21*	**0.31** *0.21*	**0.22** *0.21*	**−0.01** *0.21*	**0.05** *0.21*	**0.64** *0.21*	**0.18** *0.23*	**0.41** *0.19*	**−0.17** *0.19*	**0.70** *0.21*	**0.31** *0.21*	**−0.37** *0.21*	**−0.27** *0.19*	**−0.59** *0.19*	**−0.08** *0.20*
Chns−Cntns			**0.18** *0.24*	**0.10** *0.22*	**−0.13** *0.20*	**−0.08** *0.23*	**0.51** *0.21*	**0.06** *0.24*	**0.28** *0.17*	**−0.29** *0.17*	**0.57** *0.23*	**0.18** *0.23*	**−0.50** *0.21*	**−0.40** *0.16*	**−0.72** *0.15*	**−0.21** *0.21*
Czech				**−0.09** *0.18*	**−0.32** *0.20*	**−0.26** *0.17*	**0.33** *0.20*	**−0.13** *0.22*	**0.10** *0.19*	**−0.48** *0.19*	**0.39** *0.18*	**0.00** *0.18*	**−0.68** *0.19*	**−0.58** *0.19*	**−0.90** *0.20*	**−0.39** *0.17*
Dutch					**−0.23** *0.18*	**−0.17** *0.17*	**0.42** *0.19*	**−0.04** *0.21*	**0.19** *0.19*	**−0.39** *0.18*	**0.47** *0.18*	**0.09** *0.18*	**−0.59** *0.18*	**−0.50** *0.18*	**−0.82** *0.19*	**−0.31** *0.17*
Finnish						**0.06** *0.19*	**0.65** *0.19*	**0.19** *0.21*	**0.42** *0.19*	**−0.16** *0.17*	**0.70** *0.19*	**0.32** *0.19*	**−0.36** *0.19*	**−0.26** *0.16*	**−0.59** *0.17*	**−0.08** *0.17*
French							**0.59** *0.19*	**0.13** *0.20*	**0.36** *0.18*	**−0.22** *0.19*	**0.64** *0.17*	**0.26** *0.18*	**−0.42** *0.17*	**−0.32** *0.18*	**−0.65** *0.20*	**−0.14** *0.16*
German								**−0.46** *0.22*	**−0.23** *0.18*	**−0.81** *0.19*	**0.06** *0.19*	**−0.33** *0.20*	**−1.01** *0.19*	**−0.91** *0.18*	**−1.23** *0.19*	**−0.72** *0.18*
Italian									**0.23** *0.21*	**−0.35** *0.22*	**0.51** *0.21*	**0.13** *0.22*	**−0.55** *0.21*	**−0.45** *0.20*	**−0.78** *0.22*	**−0.27** *0.20*
Japanese										**−0.58** *0.17*	**0.29** *0.19*	**−0.10** *0.19*	**−0.78** *0.18*	**−0.68** *0.16*	**−1.00** *0.16*	**−0.49** *0.17*
Norwegian											**0.86** *0.20*	**0.48** *0.19*	**−0.20** *0.18*	**−0.10** *0.15*	**−0.43** *0.14*	**0.08** *0.18*
Polish												**−0.39** *0.19*	**−1.07** *0.19*	**−0.97** *0.18*	**−1.29** *0.20*	**−0.78** *0.17*
Russian													**−0.68** *0.19*	**−0.58** *0.19*	**−0.91** *0.20*	**−0.39** *0.17*
Spanish														**0.10** *0.18*	**−0.22** *0.19*	**0.29** *0.17*
Swedish															**−0.32** *0.15*	**0.19** *0.17*
Tswana																**0.51** *0.16*

▒ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

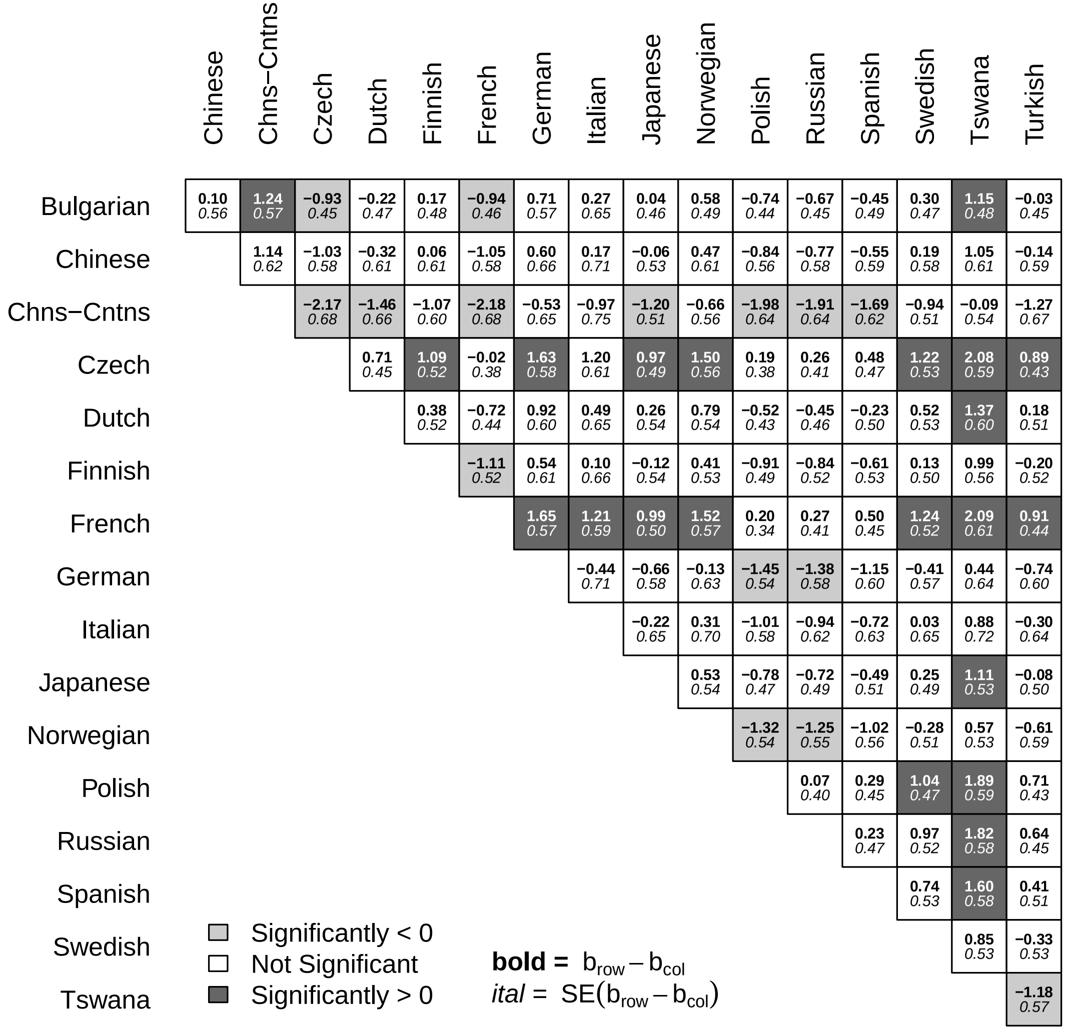
Figure 21: Pairwise Differences in L1 Parameter Values for Present Progressive (are) Neg. Binomial Type I Regression.

	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	▨ **−0.51** *0.20*	▨ **−0.44** *0.19*	**−0.25** *0.18*	**−0.10** *0.18*	▨ **−0.59** *0.17*	▨ **−0.38** *0.18*	**−0.05** *0.18*	▨ **−0.40** *0.20*	▨ **−0.63** *0.16*	▨ **−0.37** *0.17*	**−0.14** *0.18*	▨ **−0.37** *0.18*	▨ **−1.01** *0.17*	▨ **−0.84** *0.16*	▨ **−0.74** *0.15*	▨ **−0.35** *0.16*
Chinese		**0.07** *0.20*	**0.26** *0.21*	**0.41** *0.21*	**−0.08** *0.20*	**0.13** *0.20*	■ **0.45** *0.20*	**0.10** *0.22*	**−0.12** *0.18*	**0.14** *0.20*	**0.37** *0.20*	**0.14** *0.20*	▨ **−0.51** *0.20*	**−0.33** *0.19*	**−0.23** *0.19*	**0.16** *0.20*
Chns−Cntns			**0.19** *0.23*	**0.34** *0.21*	**−0.15** *0.18*	**0.06** *0.22*	■ **0.39** *0.20*	**0.04** *0.22*	**−0.19** *0.15*	**0.07** *0.17*	**0.30** *0.21*	**0.07** *0.21*	▨ **−0.57** *0.19*	▨ **−0.40** *0.15*	▨ **−0.30** *0.15*	**0.09** *0.21*
Czech				**0.15** *0.17*	**−0.34** *0.18*	**−0.13** *0.15*	**0.19** *0.18*	**−0.16** *0.19*	▨ **−0.38** *0.17*	**−0.12** *0.19*	**0.11** *0.16*	**−0.13** *0.17*	▨ **−0.77** *0.17*	▨ **−0.59** *0.18*	▨ **−0.49** *0.19*	**−0.10** *0.16*
Dutch					▨ **−0.49** *0.17*	**−0.28** *0.16*	**0.05** *0.18*	**−0.30** *0.20*	▨ **−0.53** *0.18*	**−0.27** *0.18*	**−0.04** *0.17*	**−0.27** *0.17*	▨ **−0.91** *0.17*	▨ **−0.74** *0.17*	▨ **−0.64** *0.19*	**−0.25** *0.17*
Finnish						**0.21** *0.17*	■ **0.54** *0.18*	**0.19** *0.19*	**−0.04** *0.17*	**0.22** *0.16*	■ **0.45** *0.17*	**0.22** *0.18*	▨ **−0.42** *0.17*	**−0.25** *0.15*	**−0.15** *0.16*	**0.24** *0.16*
French							**0.32** *0.17*	**−0.03** *0.18*	**−0.25** *0.17*	**0.01** *0.19*	**0.24** *0.15*	**0.00** *0.16*	▨ **−0.64** *0.15*	▨ **−0.46** *0.17*	**−0.36** *0.19*	**0.03** *0.16*
German								**−0.35** *0.20*	▨ **−0.57** *0.17*	**−0.32** *0.18*	**−0.09** *0.17*	**−0.32** *0.18*	▨ **−0.96** *0.17*	▨ **−0.79** *0.16*	▨ **−0.69** *0.18*	**−0.29** *0.17*
Italian									**−0.22** *0.19*	**0.04** *0.20*	**0.26** *0.18*	**0.03** *0.19*	▨ **−0.61** *0.18*	▨ **−0.43** *0.18*	**−0.34** *0.20*	**0.06** *0.19*
Japanese										**0.26** *0.16*	■ **0.49** *0.17*	**0.26** *0.17*	▨ **−0.39** *0.16*	**−0.21** *0.15*	**−0.11** *0.15*	**0.28** *0.16*
Norwegian											**0.23** *0.18*	**−0.00** *0.18*	▨ **−0.64** *0.17*	▨ **−0.47** *0.15*	▨ **−0.37** *0.15*	**0.02** *0.18*
Polish												**−0.23** *0.17*	▨ **−0.87** *0.16*	▨ **−0.70** *0.16*	▨ **−0.60** *0.19*	**−0.21** *0.16*
Russian													▨ **−0.64** *0.17*	▨ **−0.47** *0.18*	▨ **−0.37** *0.19*	**0.02** *0.16*
Spanish														**0.17** *0.16*	**0.27** *0.18*	■ **0.67** *0.16*
Swedish															**0.10** *0.14*	■ **0.49** *0.16*
Tswana																■ **0.39** *0.16*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 22: Pairwise Differences in L1 Parameter Values for Present Progressive (is) Neg. Binomial Type I Regression.

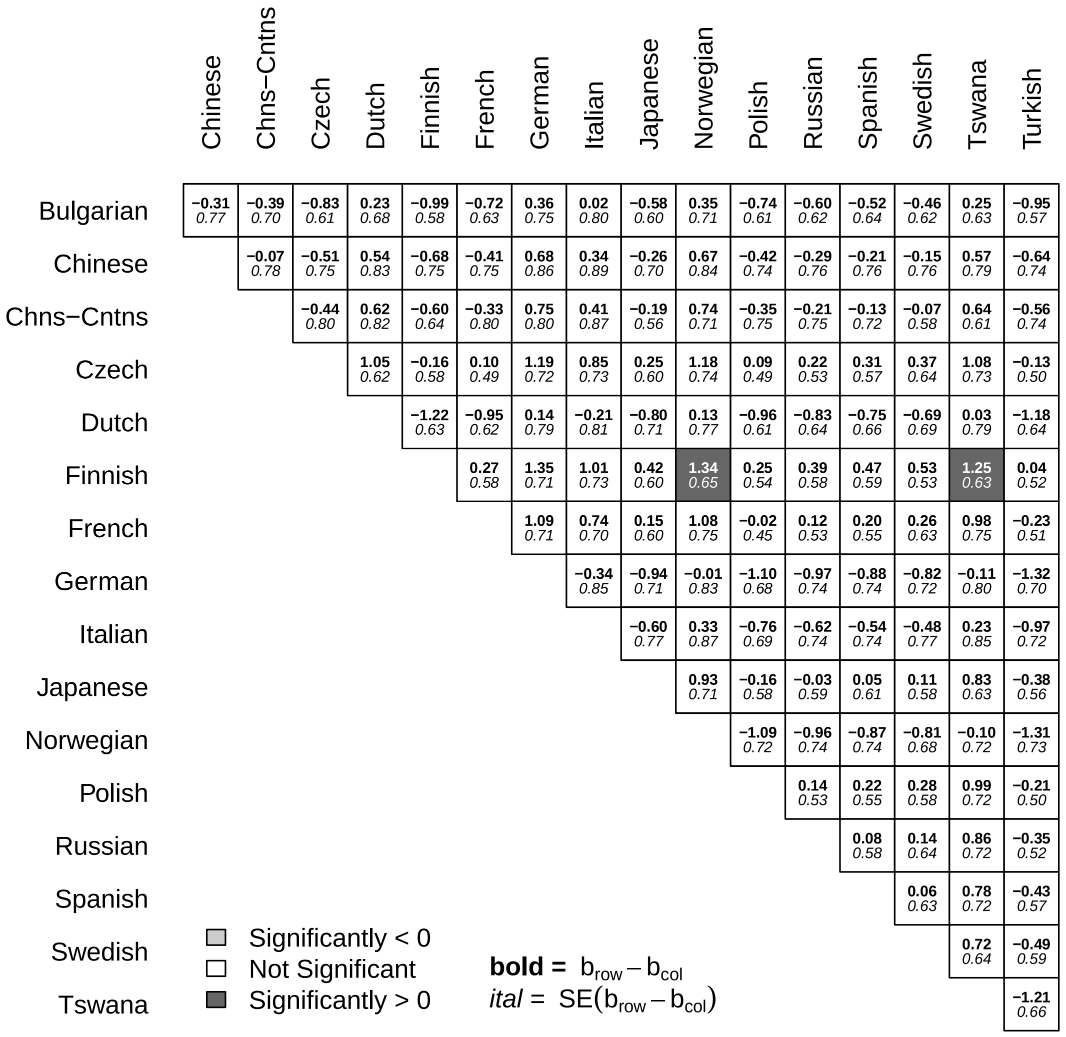


	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**0.10** *0.56*	**1.24** *0.57*	**−0.93** *0.45*	**−0.22** *0.47*	**0.17** *0.48*	**−0.94** *0.46*	**0.71** *0.57*	**0.27** *0.65*	**0.04** *0.46*	**0.58** *0.49*	**−0.74** *0.44*	**−0.67** *0.45*	**−0.45** *0.49*	**0.30** *0.47*	**1.15** *0.48*	**−0.03** *0.45*
Chinese		**1.14** *0.62*	**−1.03** *0.58*	**−0.32** *0.61*	**0.06** *0.61*	**−1.05** *0.58*	**0.60** *0.66*	**0.17** *0.71*	**−0.06** *0.53*	**0.47** *0.61*	**−0.84** *0.56*	**−0.77** *0.58*	**−0.55** *0.59*	**0.19** *0.58*	**1.05** *0.61*	**−0.14** *0.59*
Chns−Cntns			**−2.17** *0.68*	**−1.46** *0.66*	**−1.07** *0.60*	**−2.18** *0.68*	**−0.53** *0.65*	**−0.97** *0.75*	**−1.20** *0.51*	**−0.66** *0.56*	**−1.98** *0.64*	**−1.91** *0.64*	**−1.69** *0.62*	**−0.94** *0.51*	**−0.09** *0.54*	**−1.27** *0.67*
Czech				**0.71** *0.45*	**1.09** *0.52*	**−0.02** *0.38*	**1.63** *0.58*	**1.20** *0.61*	**0.97** *0.49*	**1.50** *0.56*	**0.19** *0.38*	**0.26** *0.41*	**0.48** *0.47*	**1.22** *0.53*	**2.08** *0.59*	**0.89** *0.43*
Dutch					**0.38** *0.52*	**−0.72** *0.44*	**0.92** *0.60*	**0.49** *0.65*	**0.26** *0.54*	**0.79** *0.54*	**−0.52** *0.43*	**−0.45** *0.46*	**−0.23** *0.50*	**0.52** *0.53*	**1.37** *0.60*	**0.18** *0.51*
Finnish						**−1.11** *0.52*	**0.54** *0.61*	**0.10** *0.66*	**−0.12** *0.54*	**0.41** *0.53*	**−0.91** *0.49*	**−0.84** *0.52*	**−0.61** *0.53*	**0.13** *0.50*	**0.99** *0.56*	**−0.20** *0.52*
French							**1.65** *0.57*	**1.21** *0.59*	**0.99** *0.50*	**1.52** *0.57*	**0.20** *0.34*	**0.27** *0.41*	**0.50** *0.45*	**1.24** *0.52*	**2.09** *0.61*	**0.91** *0.44*
German								**−0.44** *0.71*	**−0.66** *0.58*	**−0.13** *0.63*	**−1.45** *0.54*	**−1.38** *0.58*	**−1.15** *0.60*	**−0.41** *0.57*	**0.44** *0.64*	**−0.74** *0.60*
Italian									**−0.22** *0.65*	**0.31** *0.70*	**−1.01** *0.58*	**−0.94** *0.62*	**−0.72** *0.63*	**0.03** *0.65*	**0.88** *0.72*	**−0.30** *0.64*
Japanese										**0.53** *0.54*	**−0.78** *0.47*	**−0.72** *0.49*	**−0.49** *0.51*	**0.25** *0.49*	**1.11** *0.53*	**−0.08** *0.50*
Norwegian											**−1.32** *0.54*	**−1.25** *0.55*	**−1.02** *0.56*	**−0.28** *0.51*	**0.57** *0.53*	**−0.61** *0.59*
Polish												**0.07** *0.40*	**0.29** *0.45*	**1.04** *0.47*	**1.89** *0.59*	**0.71** *0.43*
Russian													**0.23** *0.47*	**0.97** *0.52*	**1.82** *0.58*	**0.64** *0.45*
Spanish														**0.74** *0.53*	**1.60** *0.58*	**0.41** *0.51*
Swedish															**0.85** *0.53*	**−0.33** *0.53*
Tswana																**−1.18** *0.57*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 23: Pairwise Differences in L1 Parameter Values for Present Perfect Progressive (have) Neg. Binomial Type I Regression.



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**−0.31** *0.77*	**−0.39** *0.70*	**−0.83** *0.61*	**0.23** *0.68*	**−0.99** *0.58*	**−0.72** *0.63*	**0.36** *0.75*	**0.02** *0.80*	**−0.58** *0.60*	**0.35** *0.71*	**−0.74** *0.61*	**−0.60** *0.62*	**−0.52** *0.64*	**−0.46** *0.62*	**0.25** *0.63*	**−0.95** *0.57*
Chinese		**−0.07** *0.78*	**−0.51** *0.75*	**0.54** *0.83*	**−0.68** *0.75*	**−0.41** *0.75*	**0.68** *0.86*	**0.34** *0.89*	**−0.26** *0.70*	**0.67** *0.84*	**−0.42** *0.74*	**−0.29** *0.76*	**−0.21** *0.76*	**−0.15** *0.76*	**0.57** *0.79*	**−0.64** *0.74*
Chns−Cntns			**−0.44** *0.80*	**0.62** *0.82*	**−0.60** *0.64*	**−0.33** *0.80*	**0.75** *0.80*	**0.41** *0.87*	**−0.19** *0.56*	**0.74** *0.71*	**−0.35** *0.75*	**−0.21** *0.75*	**−0.13** *0.72*	**−0.07** *0.58*	**0.64** *0.61*	**−0.56** *0.74*
Czech				**1.05** *0.62*	**−0.16** *0.58*	**0.10** *0.49*	**1.19** *0.72*	**0.85** *0.73*	**0.25** *0.60*	**1.18** *0.74*	**0.09** *0.49*	**0.22** *0.53*	**0.31** *0.57*	**0.37** *0.64*	**1.08** *0.73*	**−0.13** *0.50*
Dutch					**−1.22** *0.63*	**−0.95** *0.62*	**0.14** *0.79*	**−0.21** *0.81*	**−0.80** *0.71*	**0.13** *0.77*	**−0.96** *0.61*	**−0.83** *0.64*	**−0.75** *0.66*	**−0.69** *0.69*	**0.03** *0.79*	**−1.18** *0.64*
Finnish						**0.27** *0.58*	**1.35** *0.71*	**1.01** *0.73*	**0.42** *0.60*	**1.34** *0.65*	**0.25** *0.54*	**0.39** *0.58*	**0.47** *0.59*	**0.53** *0.53*	**1.25** *0.63*	**0.04** *0.52*
French							**1.09** *0.71*	**0.74** *0.70*	**0.15** *0.60*	**1.08** *0.75*	**−0.02** *0.45*	**0.12** *0.53*	**0.20** *0.55*	**0.26** *0.63*	**0.98** *0.75*	**−0.23** *0.51*
German								**−0.34** *0.85*	**−0.94** *0.71*	**−0.01** *0.83*	**−1.10** *0.68*	**−0.97** *0.74*	**−0.88** *0.74*	**−0.82** *0.72*	**−0.11** *0.80*	**−1.32** *0.70*
Italian									**−0.60** *0.77*	**0.33** *0.87*	**−0.76** *0.69*	**−0.62** *0.74*	**−0.54** *0.74*	**−0.48** *0.77*	**0.23** *0.85*	**−0.97** *0.72*
Japanese										**0.93** *0.71*	**−0.16** *0.58*	**−0.03** *0.59*	**0.05** *0.61*	**0.11** *0.58*	**0.83** *0.63*	**−0.38** *0.56*
Norwegian											**−1.09** *0.72*	**−0.96** *0.74*	**−0.87** *0.74*	**−0.81** *0.68*	**−0.10** *0.72*	**−1.31** *0.73*
Polish												**0.14** *0.53*	**0.22** *0.55*	**0.28** *0.58*	**0.99** *0.72*	**−0.21** *0.50*
Russian													**0.08** *0.58*	**0.14** *0.64*	**0.86** *0.72*	**−0.35** *0.52*
Spanish														**0.06** *0.63*	**0.78** *0.72*	**−0.43** *0.57*
Swedish															**0.72** *0.64*	**−0.49** *0.59*
Tswana																**−1.21** *0.66*

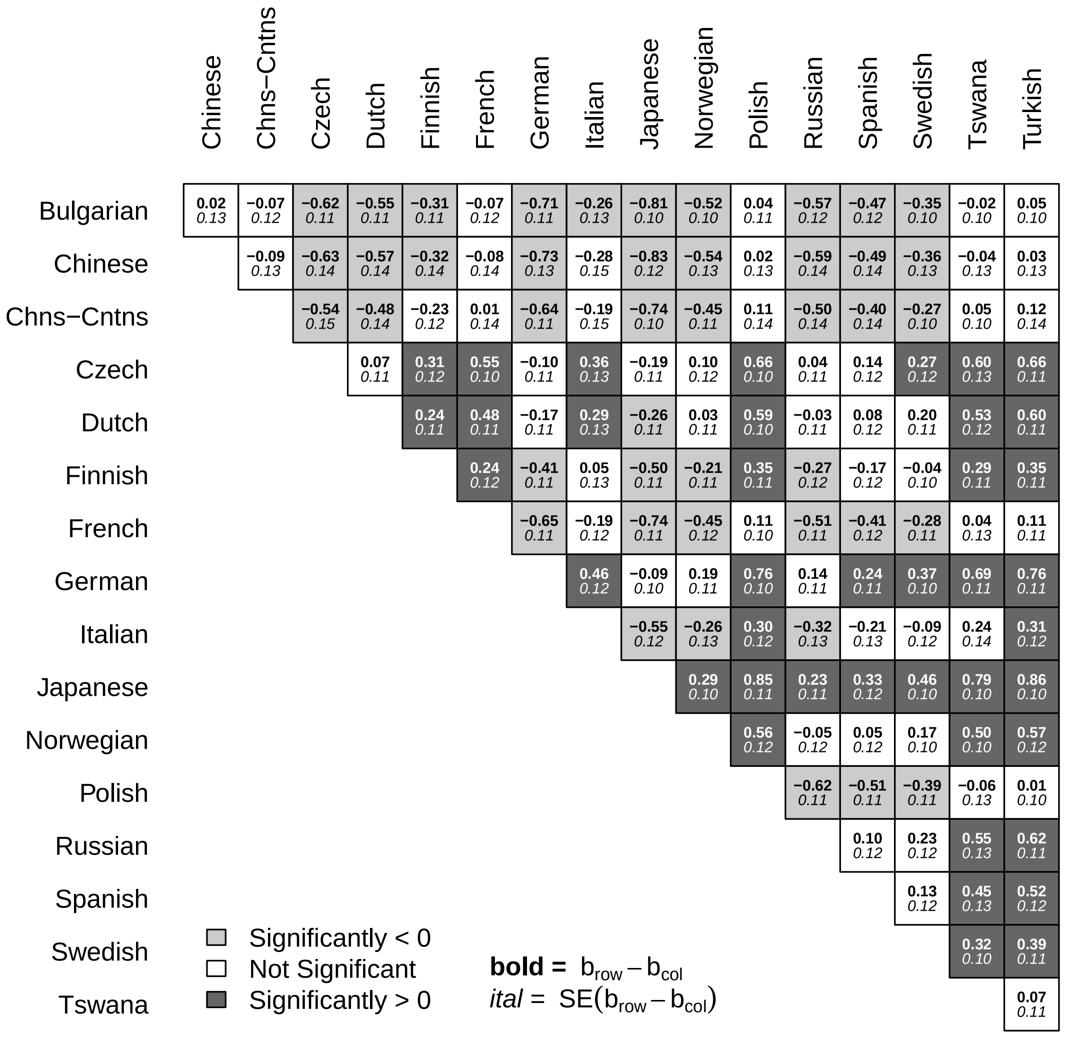
░ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 24: Pairwise Differences in L1 Parameter Values for Present Perfect Progressive (has) Neg. Binomial Type I Regression.

7.5.2 The Past Active

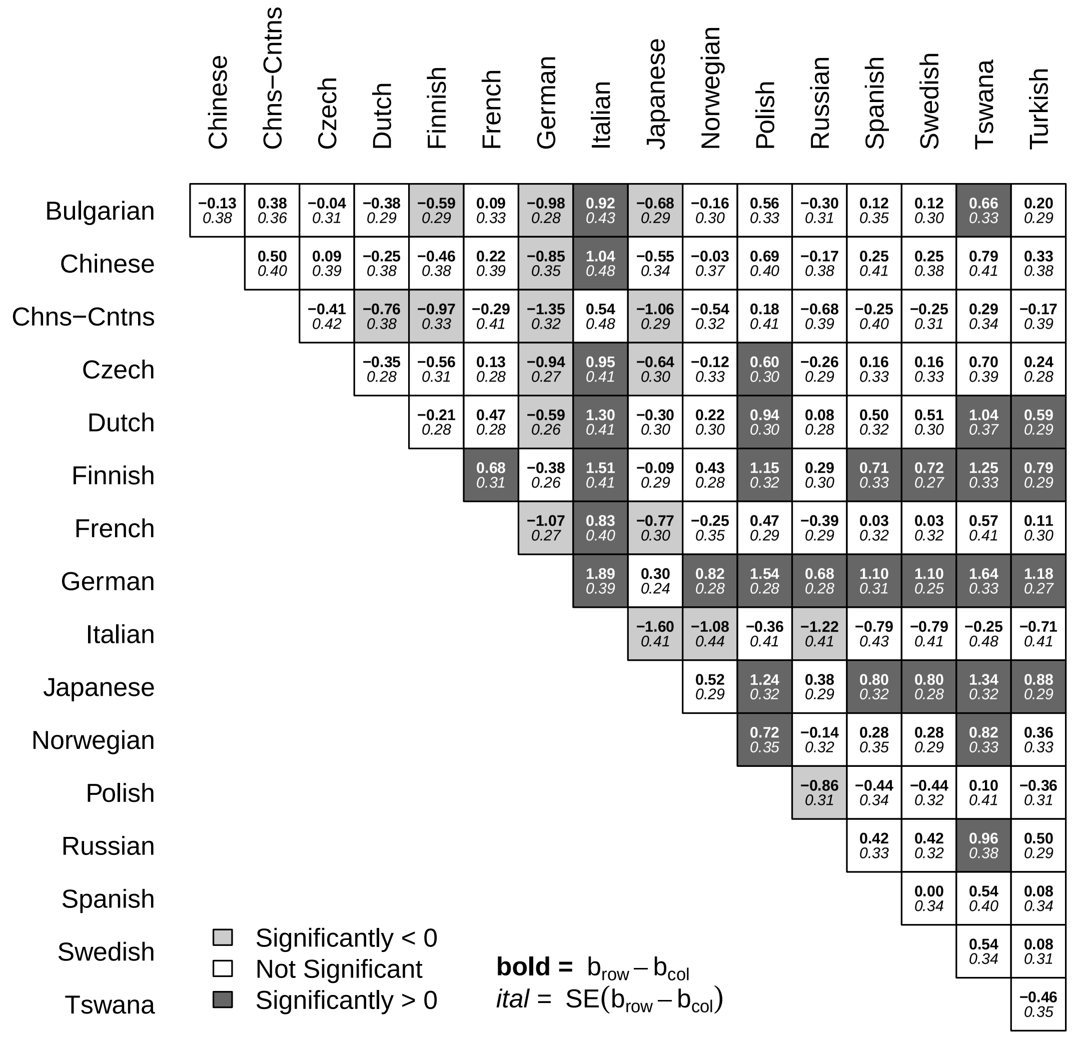
Out of 5 possible Tense and Aspect combinations the pairwise differences in L1 parameter values for 5 could have been plotted.



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**0.02** *0.13*	**−0.07** *0.12*	**−0.62** *0.11*	**−0.55** *0.11*	**−0.31** *0.11*	**−0.07** *0.12*	**−0.71** *0.11*	**−0.26** *0.13*	**−0.81** *0.10*	**−0.52** *0.10*	**0.04** *0.11*	**−0.57** *0.12*	**−0.47** *0.12*	**−0.35** *0.10*	**−0.02** *0.10*	**0.05** *0.10*
Chinese		**−0.09** *0.13*	**−0.63** *0.14*	**−0.57** *0.14*	**−0.32** *0.14*	**−0.08** *0.14*	**−0.73** *0.13*	**−0.28** *0.15*	**−0.83** *0.12*	**−0.54** *0.13*	**0.02** *0.13*	**−0.59** *0.14*	**−0.49** *0.14*	**−0.36** *0.13*	**−0.04** *0.13*	**0.03** *0.13*
Chns−Cntns			**−0.54** *0.15*	**−0.48** *0.14*	**−0.23** *0.12*	**0.01** *0.14*	**−0.64** *0.11*	**−0.19** *0.15*	**−0.74** *0.10*	**−0.45** *0.11*	**0.11** *0.14*	**−0.50** *0.14*	**−0.40** *0.14*	**−0.27** *0.10*	**0.05** *0.10*	**0.12** *0.14*
Czech				**0.07** *0.11*	**0.31** *0.12*	**0.55** *0.10*	**−0.10** *0.11*	**0.36** *0.13*	**−0.19** *0.11*	**0.10** *0.12*	**0.66** *0.10*	**0.04** *0.11*	**0.14** *0.12*	**0.27** *0.12*	**0.60** *0.13*	**0.66** *0.11*
Dutch					**0.24** *0.11*	**0.48** *0.11*	**−0.17** *0.11*	**0.29** *0.13*	**−0.26** *0.11*	**0.03** *0.11*	**0.59** *0.10*	**−0.03** *0.11*	**0.08** *0.12*	**0.20** *0.11*	**0.53** *0.12*	**0.60** *0.11*
Finnish						**0.24** *0.12*	**−0.41** *0.11*	**0.05** *0.13*	**−0.50** *0.11*	**−0.21** *0.11*	**0.35** *0.11*	**−0.27** *0.12*	**−0.17** *0.12*	**−0.04** *0.10*	**0.29** *0.11*	**0.35** *0.11*
French							**−0.65** *0.11*	**−0.19** *0.12*	**−0.74** *0.11*	**−0.45** *0.12*	**0.11** *0.10*	**−0.51** *0.11*	**−0.41** *0.12*	**−0.28** *0.11*	**0.04** *0.13*	**0.11** *0.11*
German								**0.46** *0.12*	**−0.09** *0.10*	**0.19** *0.11*	**0.76** *0.10*	**0.14** *0.11*	**0.24** *0.11*	**0.37** *0.10*	**0.69** *0.11*	**0.76** *0.11*
Italian									**−0.55** *0.12*	**−0.26** *0.13*	**0.30** *0.12*	**−0.32** *0.13*	**−0.21** *0.13*	**−0.09** *0.12*	**0.24** *0.14*	**0.31** *0.12*
Japanese										**0.29** *0.10*	**0.85** *0.11*	**0.23** *0.11*	**0.33** *0.12*	**0.46** *0.10*	**0.79** *0.10*	**0.86** *0.10*
Norwegian											**0.56** *0.12*	**−0.05** *0.12*	**0.05** *0.12*	**0.17** *0.10*	**0.50** *0.10*	**0.57** *0.12*
Polish												**−0.62** *0.11*	**−0.51** *0.11*	**−0.39** *0.11*	**−0.06** *0.13*	**0.01** *0.10*
Russian													**0.10** *0.12*	**0.23** *0.12*	**0.55** *0.13*	**0.62** *0.11*
Spanish														**0.13** *0.12*	**0.45** *0.13*	**0.52** *0.12*
Swedish															**0.32** *0.10*	**0.39** *0.11*
Tswana																**0.07** *0.11*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0
bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

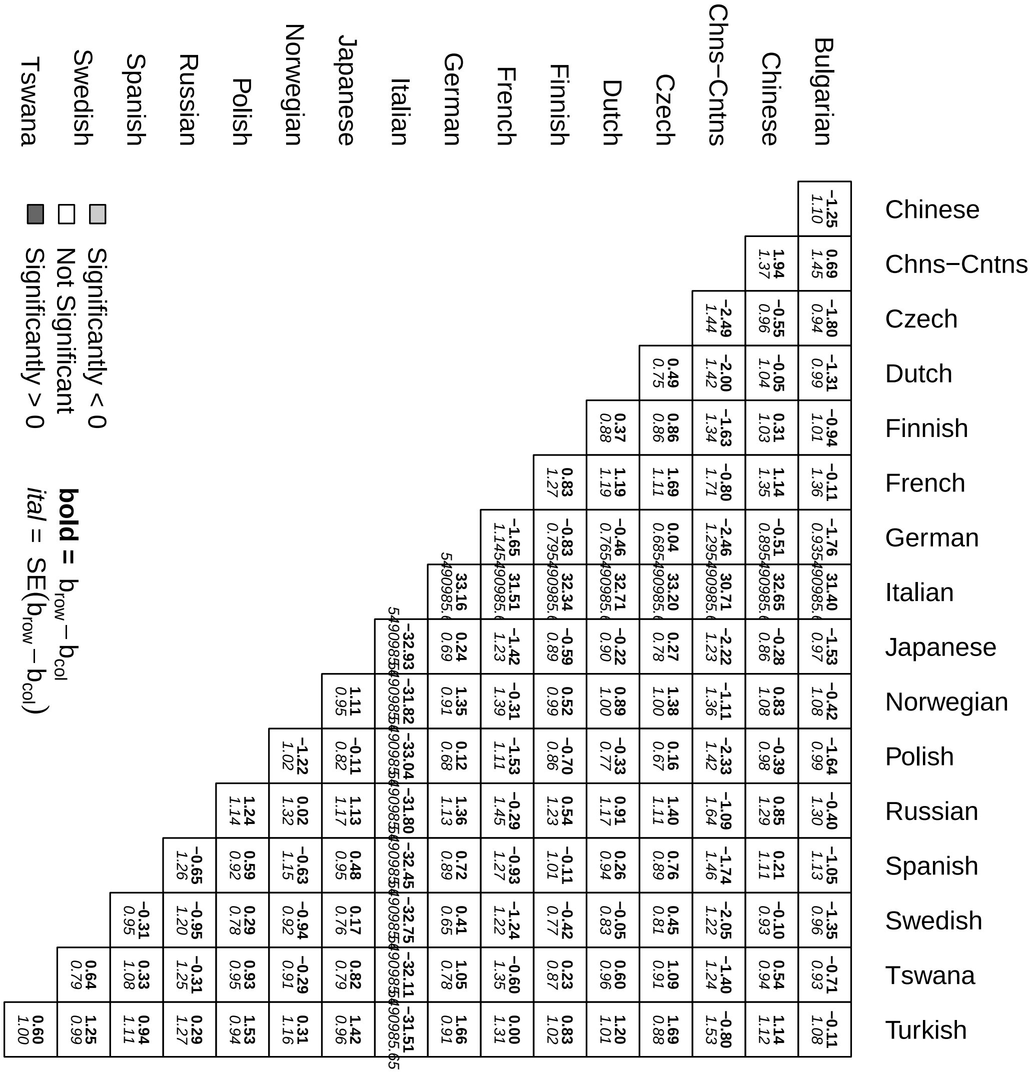
Figure 25: Pairwise Differences in L1 Parameter Values for Past Simple Neg. Binomial Type I Regression.



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**−0.13** *0.38*	**0.38** *0.36*	**−0.04** *0.31*	**−0.38** *0.29*	**−0.59** *0.29*	**0.09** *0.33*	**−0.98** *0.28*	**0.92** *0.43*	**−0.68** *0.29*	**−0.16** *0.30*	**0.56** *0.33*	**−0.30** *0.31*	**0.12** *0.35*	**0.12** *0.30*	**0.66** *0.33*	**0.20** *0.29*
Chinese		**0.50** *0.40*	**0.09** *0.39*	**−0.25** *0.38*	**−0.46** *0.38*	**0.22** *0.39*	**−0.85** *0.35*	**1.04** *0.48*	**−0.55** *0.34*	**−0.03** *0.37*	**0.69** *0.40*	**−0.17** *0.38*	**0.25** *0.41*	**0.25** *0.38*	**0.79** *0.41*	**0.33** *0.38*
Chns−Cntns			**−0.41** *0.42*	**−0.76** *0.38*	**−0.97** *0.33*	**−0.29** *0.41*	**−1.35** *0.32*	**0.54** *0.48*	**−1.06** *0.29*	**−0.54** *0.32*	**0.18** *0.41*	**−0.68** *0.39*	**−0.25** *0.40*	**−0.25** *0.31*	**0.29** *0.34*	**−0.17** *0.39*
Czech				**−0.35** *0.28*	**−0.56** *0.31*	**0.13** *0.28*	**−0.94** *0.27*	**0.95** *0.41*	**−0.64** *0.30*	**−0.12** *0.33*	**0.60** *0.30*	**−0.26** *0.29*	**0.16** *0.33*	**0.16** *0.33*	**0.70** *0.39*	**0.24** *0.28*
Dutch					**−0.21** *0.28*	**0.47** *0.28*	**−0.59** *0.26*	**1.30** *0.41*	**−0.30** *0.30*	**0.22** *0.30*	**0.94** *0.30*	**0.08** *0.28*	**0.50** *0.32*	**0.51** *0.30*	**1.04** *0.37*	**0.59** *0.29*
Finnish						**0.68** *0.31*	**−0.38** *0.26*	**1.51** *0.41*	**−0.09** *0.29*	**0.43** *0.28*	**1.15** *0.32*	**0.29** *0.30*	**0.71** *0.33*	**0.72** *0.27*	**1.25** *0.33*	**0.79** *0.29*
French							**−1.07** *0.27*	**0.83** *0.40*	**−0.77** *0.30*	**−0.25** *0.35*	**0.47** *0.29*	**−0.39** *0.29*	**0.03** *0.32*	**0.03** *0.32*	**0.57** *0.41*	**0.11** *0.30*
German								**1.89** *0.39*	**0.30** *0.24*	**0.82** *0.28*	**1.54** *0.28*	**0.68** *0.28*	**1.10** *0.31*	**1.10** *0.25*	**1.64** *0.33*	**1.18** *0.27*
Italian									**−1.60** *0.41*	**−1.08** *0.44*	**−0.36** *0.41*	**−1.22** *0.41*	**−0.79** *0.43*	**−0.79** *0.41*	**−0.25** *0.48*	**−0.71** *0.41*
Japanese										**0.52** *0.29*	**1.24** *0.32*	**0.38** *0.29*	**0.80** *0.32*	**0.80** *0.28*	**1.34** *0.32*	**0.88** *0.29*
Norwegian											**0.72** *0.35*	**−0.14** *0.32*	**0.28** *0.35*	**0.28** *0.29*	**0.82** *0.33*	**0.36** *0.33*
Polish												**−0.86** *0.31*	**−0.44** *0.34*	**−0.44** *0.32*	**0.10** *0.41*	**−0.36** *0.31*
Russian													**0.42** *0.33*	**0.42** *0.32*	**0.96** *0.38*	**0.50** *0.29*
Spanish														**0.00** *0.34*	**0.54** *0.40*	**0.08** *0.34*
Swedish															**0.54** *0.34*	**0.08** *0.31*
Tswana																**−0.46** *0.35*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0
bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 26: Pairwise Differences in L1 Parameter Values for Past Perfect Neg. Binomial Type I Regression.

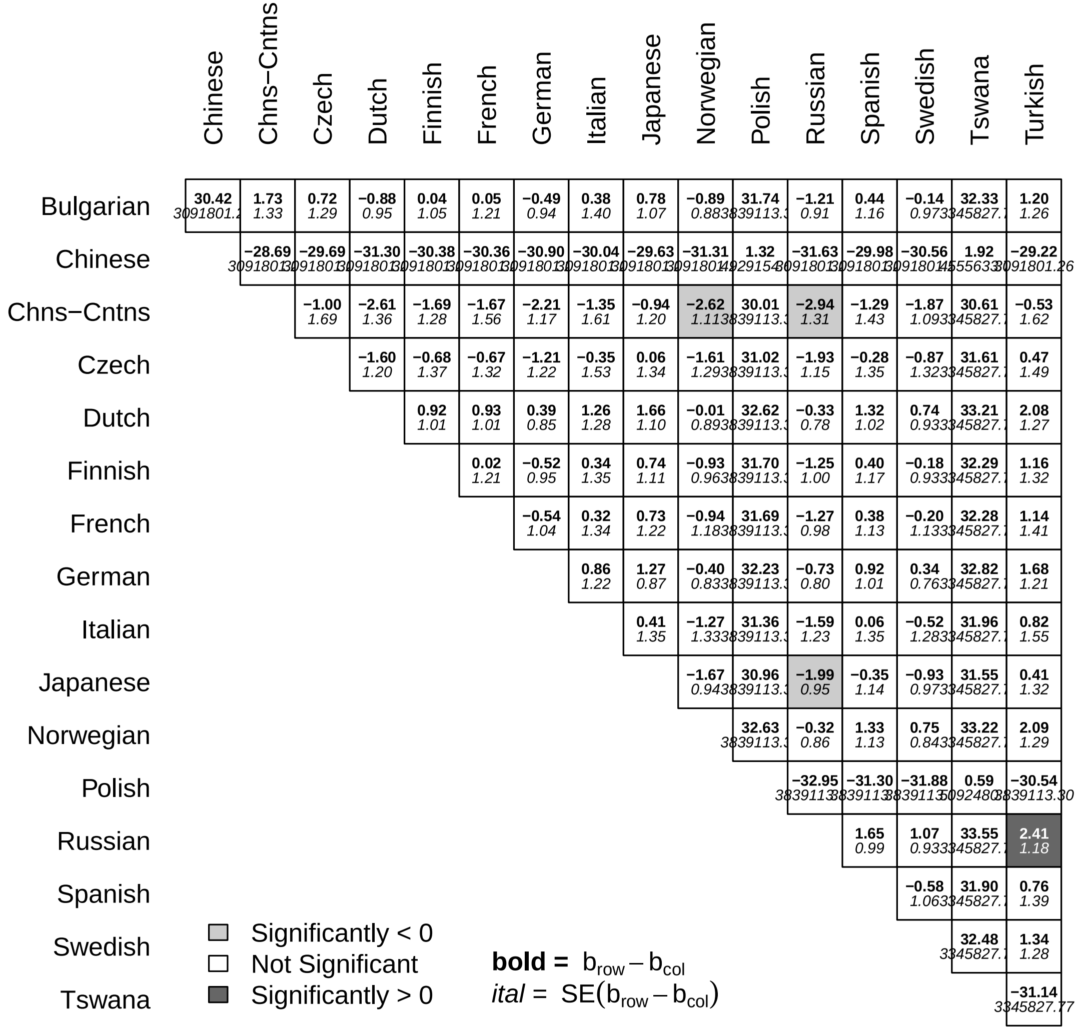


	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**−1.25** *1.10*	**0.69** *1.45*	**−1.80** *0.94*	**−1.31** *0.99*	**−0.94** *1.01*	**−0.11** *1.36*	**−1.76** *0.93*	**31.40** *5490985.6*	**−1.53** *0.97*	**−0.42** *1.08*	**−1.64** *0.99*	**−0.40** *1.30*	**−1.05** *1.13*	**−1.35** *0.96*	**−0.71** *0.93*	**−0.11** *1.08*
Chinese		**1.94** *1.37*	**−0.55** *0.96*	**−0.05** *1.04*	**0.31** *1.03*	**1.14** *1.35*	**−0.51** *0.89*	**32.65** *5490985.6*	**−0.28** *0.86*	**0.83** *1.08*	**−0.39** *0.98*	**0.85** *1.29*	**0.21** *1.11*	**−0.10** *0.93*	**0.54** *0.94*	**1.14** *1.12*
Chns−Cntns			**−2.49** *1.44*	**−2.00** *1.42*	**−1.63** *1.34*	**−0.80** *1.71*	**−2.46** *1.29*	**30.71** *5490985.6*	**−2.22** *1.23*	**−1.11** *1.36*	**−2.33** *1.42*	**−1.09** *1.64*	**−1.74** *1.46*	**−2.05** *1.22*	**−1.40** *1.24*	**−0.80** *1.53*
Czech				**0.49** *0.75*	**0.86** *0.86*	**1.69** *1.11*	**0.04** *0.68*	**33.20** *5490985.6*	**0.27** *0.78*	**1.38** *1.00*	**0.16** *0.67*	**1.40** *1.11*	**0.76** *0.89*	**0.45** *0.81*	**1.09** *0.91*	**1.69** *0.88*
Dutch					**0.37** *0.88*	**1.19** *1.19*	**−0.46** *0.76*	**32.71** *5490985.6*	**−0.22** *0.90*	**0.89** *1.00*	**−0.33** *0.77*	**0.91** *1.17*	**0.26** *0.94*	**−0.05** *0.83*	**0.60** *0.96*	**1.20** *1.01*
Finnish						**0.83** *1.27*	**−0.83** *0.79*	**32.34** *5490985.6*	**−0.59** *0.89*	**0.52** *0.99*	**−0.70** *0.86*	**0.54** *1.23*	**−0.11** *1.01*	**−0.42** *0.77*	**0.23** *0.87*	**0.83** *1.02*
French							**−1.65** *1.14*	**31.51** *5490985.6*	**−1.42** *1.23*	**−0.31** *1.39*	**−1.53** *1.11*	**−0.29** *1.45*	**−0.93** *1.27*	**−1.24** *1.22*	**−0.60** *1.35*	**0.00** *1.31*
German								**33.16** *5490985.6*	**0.24** *0.69*	**1.35** *0.91*	**0.12** *0.68*	**1.36** *1.13*	**0.72** *0.89*	**0.41** *0.65*	**1.05** *0.78*	**1.66** *0.91*
Italian									**−32.93** *5490985.65*	**−31.82** *5490985.65*	**−33.04** *5490985.65*	**−31.80** *5490985.65*	**−32.45** *5490985.65*	**−32.75** *5490985.65*	**−32.11** *5490985.65*	**−31.51** *5490985.65*
Japanese										**1.11** *0.95*	**−0.11** *0.82*	**1.13** *1.17*	**0.48** *0.95*	**0.17** *0.76*	**0.82** *0.79*	**1.42** *0.96*
Norwegian											**−1.22** *1.02*	**0.02** *1.32*	**−0.63** *1.15*	**−0.94** *0.92*	**−0.29** *0.91*	**0.31** *1.16*
Polish												**1.24** *1.14*	**0.59** *0.92*	**0.29** *0.78*	**0.93** *0.95*	**1.53** *0.94*
Russian													**−0.65** *1.26*	**−0.95** *1.20*	**−0.31** *1.25*	**0.29** *1.27*
Spanish														**−0.31** *0.95*	**0.33** *1.08*	**0.94** *1.11*
Swedish															**0.64** *0.79*	**1.25** *0.99*
Tswana																**0.60** *1.00*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 27: Pairwise Differences in L1 Parameter Values for Past Progressive Neg. Binomial Type I Regression.



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**30.42** *3091801.2*	**1.73** *1.33*	**0.72** *1.29*	**−0.88** *0.95*	**0.04** *1.05*	**0.05** *1.21*	**−0.49** *0.94*	**0.38** *1.40*	**0.78** *1.07*	**−0.89** *0.883*	**31.74** *3839113.3*	**−1.21** *0.91*	**0.44** *1.16*	**−0.14** *0.973*	**32.33** *3345827.7*	**1.20** *1.26*
Chinese		**−28.69** *3091801*	**−29.69** *3091801*	**−31.30** *3091801*	**−30.38** *3091801*	**−30.36** *3091801*	**−30.90** *3091801*	**−30.04** *3091801*	**−29.63** *3091801*	**−31.31** *3091801.4*	**1.32** *4929154*	**−31.63** *3091801*	**−29.98** *3091801*	**−30.56** *3091801.4*	**1.92** *4555633*	**−29.22** *3091801.26*
Chns−Cntns			**−1.00** *1.69*	**−2.61** *1.36*	**−1.69** *1.28*	**−1.67** *1.56*	**−2.21** *1.17*	**−1.35** *1.61*	**−0.94** *1.20*	░ **−2.62** *1.113*	**30.01** *3839113.3*	░ **−2.94** *1.31*	**−1.29** *1.43*	**−1.87** *1.093*	**30.61** *3345827.7*	**−0.53** *1.62*
Czech				**−1.60** *1.20*	**−0.68** *1.37*	**−0.67** *1.32*	**−1.21** *1.22*	**−0.35** *1.53*	**0.06** *1.34*	**−1.61** *1.293*	**31.02** *3839113.3*	**−1.93** *1.15*	**−0.28** *1.35*	**−0.87** *1.323*	**31.61** *3345827.7*	**0.47** *1.49*
Dutch					**0.92** *1.01*	**0.93** *1.01*	**0.39** *0.85*	**1.26** *1.28*	**1.66** *1.10*	**−0.01** *0.893*	**32.62** *3839113.3*	**−0.33** *0.78*	**1.32** *1.02*	**0.74** *0.933*	**33.21** *3345827.7*	**2.08** *1.27*
Finnish						**0.02** *1.21*	**−0.52** *0.95*	**0.34** *1.35*	**0.74** *1.11*	**−0.93** *0.963*	**31.70** *3839113.3*	**−1.25** *1.00*	**0.40** *1.17*	**−0.18** *0.933*	**32.29** *3345827.7*	**1.16** *1.32*
French							**−0.54** *1.04*	**0.32** *1.34*	**0.73** *1.22*	**−0.94** *1.183*	**31.69** *3839113.3*	**−1.27** *0.98*	**0.38** *1.13*	**−0.20** *1.133*	**32.28** *3345827.7*	**1.14** *1.41*
German								**0.86** *1.22*	**1.27** *0.87*	**−0.40** *0.833*	**32.23** *3839113.3*	**−0.73** *0.80*	**0.92** *1.01*	**0.34** *0.763*	**32.82** *3345827.7*	**1.68** *1.21*
Italian									**0.41** *1.35*	**−1.27** *1.333*	**31.36** *3839113.3*	**−1.59** *1.23*	**0.06** *1.35*	**−0.52** *1.283*	**31.96** *3345827.7*	**0.82** *1.55*
Japanese										**−1.67** *0.943*	**30.96** *3839113.3*	░ **−1.99** *0.95*	**−0.35** *1.14*	**−0.93** *0.973*	**31.55** *3345827.7*	**0.41** *1.32*
Norwegian											**32.63** *3839113.3*	**−0.32** *0.86*	**1.33** *1.13*	**0.75** *0.843*	**33.22** *3345827.7*	**2.09** *1.29*
Polish												**−32.95** *3839113*	**−31.30** *3839113*	**−31.88** *3839113*	**0.59** *5092480*	**−30.54** *3839113.30*
Russian													**1.65** *0.99*	**1.07** *0.933*	**33.55** *3345827.7*	■ **2.41** *1.18*
Spanish														**−0.58** *1.063*	**31.90** *3345827.7*	**0.76** *1.39*
Swedish															**32.48** *3345827.7*	**1.34** *1.28*
Tswana																**−31.14** *3345827.77*

░ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 28: Pairwise Differences in L1 Parameter Values for Past Perfect Progressive Neg. Binomial Type I Regression.

7.5.3 The Future Active

Out of 8 possible Tense and Aspect combinations the pairwise differences in L1 parameter values for 7 combinations could have been plotted. To assure readability the plots of the combined constructions have been omitted.

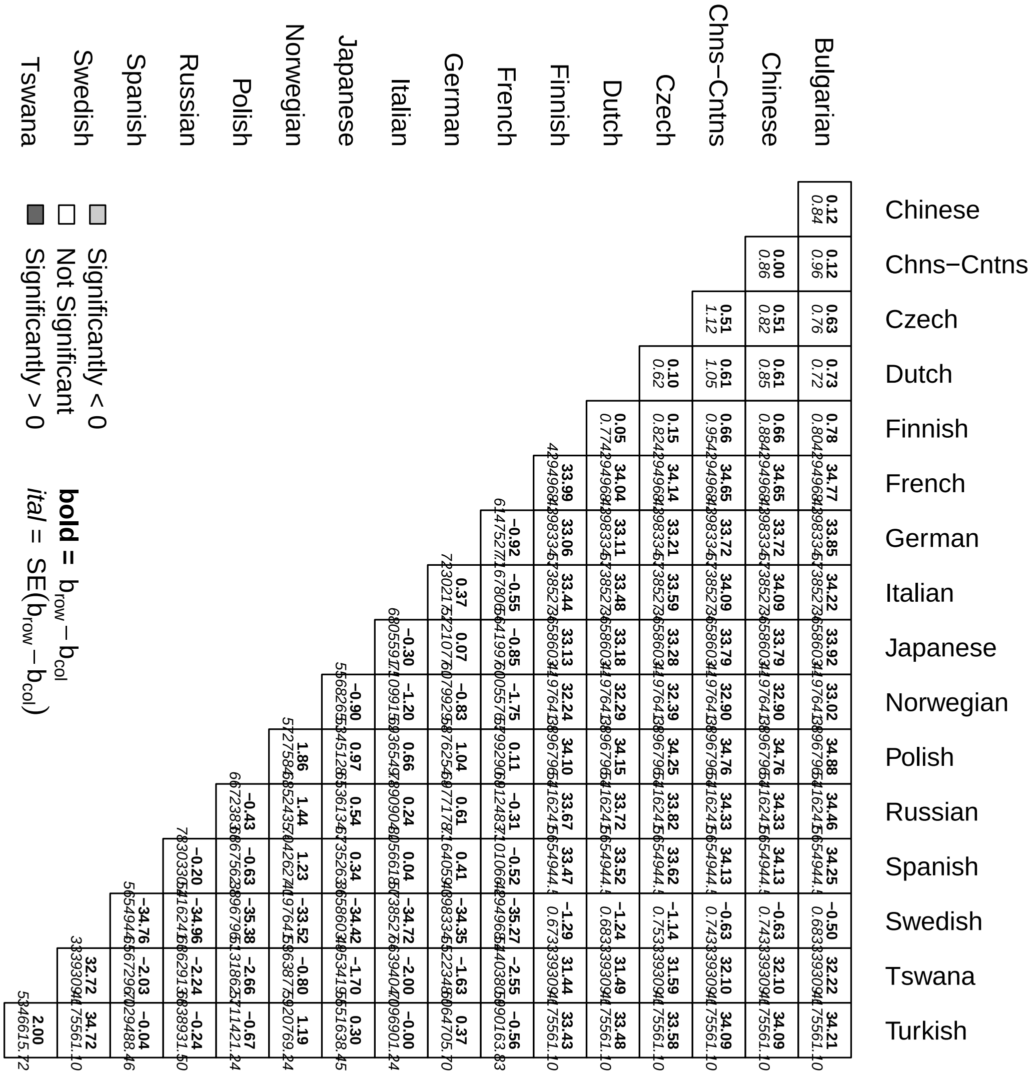
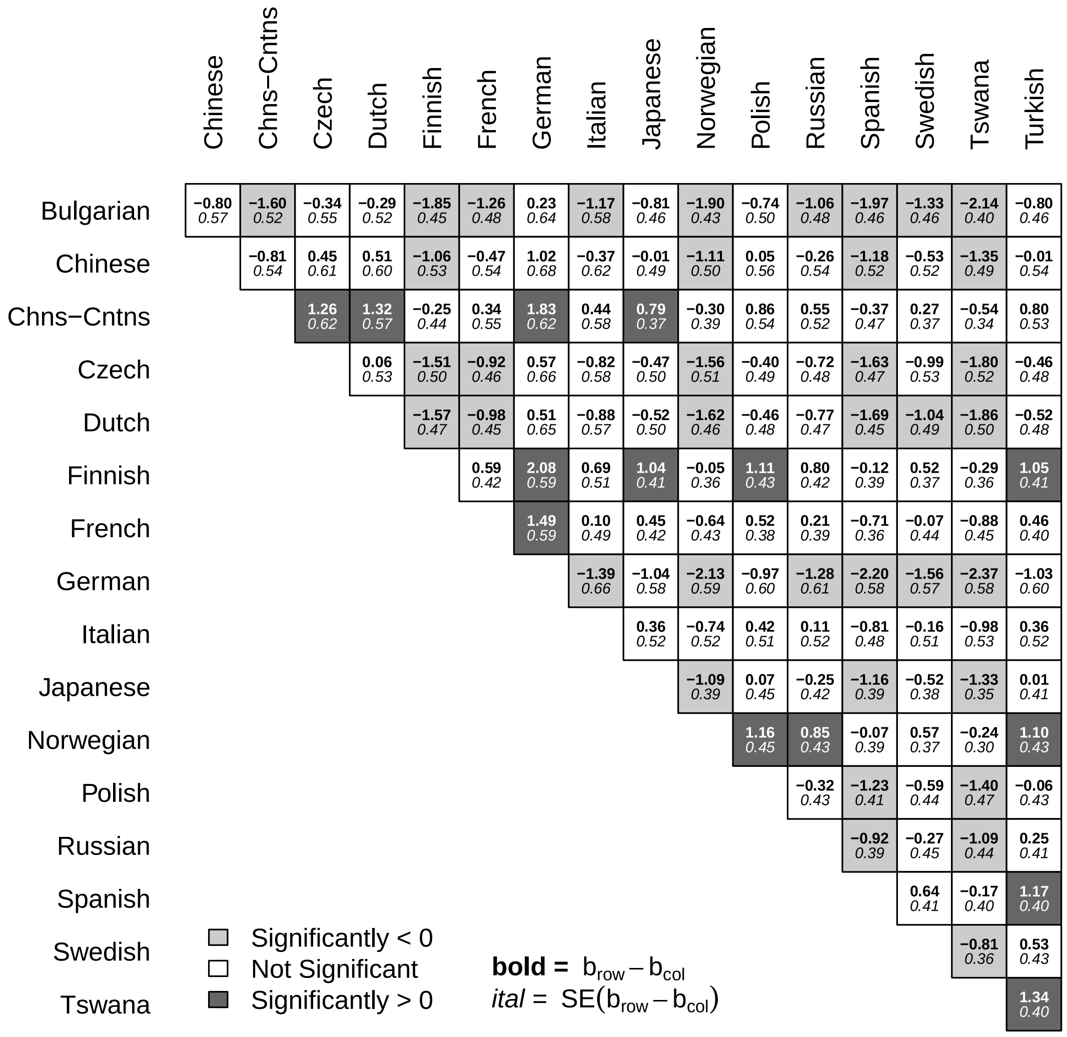


Figure 29: Pairwise Differences in L1 Parameter Values for Going-to (is) Neg. Binomial Type I Regression.



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**−0.80** *0.57*	**−1.60** *0.52*	**−0.34** *0.55*	**−0.29** *0.52*	**−1.85** *0.45*	**−1.26** *0.48*	**0.23** *0.64*	**−1.17** *0.58*	**−0.81** *0.46*	**−1.90** *0.43*	**−0.74** *0.50*	**−1.06** *0.48*	**−1.97** *0.46*	**−1.33** *0.46*	**−2.14** *0.40*	**−0.80** *0.46*
Chinese		**−0.81** *0.54*	**0.45** *0.61*	**0.51** *0.60*	**−1.06** *0.53*	**−0.47** *0.54*	**1.02** *0.68*	**−0.37** *0.62*	**−0.01** *0.49*	**−1.11** *0.50*	**0.05** *0.56*	**−0.26** *0.54*	**−1.18** *0.52*	**−0.53** *0.52*	**−1.35** *0.49*	**−0.01** *0.54*
Chns−Cntns			**1.26** *0.62*	**1.32** *0.57*	**−0.25** *0.44*	**0.34** *0.55*	**1.83** *0.62*	**0.44** *0.58*	**0.79** *0.37*	**−0.30** *0.39*	**0.86** *0.54*	**0.55** *0.52*	**−0.37** *0.47*	**0.27** *0.37*	**−0.54** *0.34*	**0.80** *0.53*
Czech				**0.06** *0.53*	**−1.51** *0.50*	**−0.92** *0.46*	**0.57** *0.66*	**−0.82** *0.58*	**−0.47** *0.50*	**−1.56** *0.51*	**−0.40** *0.49*	**−0.72** *0.48*	**−1.63** *0.47*	**−0.99** *0.53*	**−1.80** *0.52*	**−0.46** *0.48*
Dutch					**−1.57** *0.47*	**−0.98** *0.45*	**0.51** *0.65*	**−0.88** *0.57*	**−0.52** *0.50*	**−1.62** *0.46*	**−0.46** *0.48*	**−0.77** *0.47*	**−1.69** *0.45*	**−1.04** *0.49*	**−1.86** *0.50*	**−0.52** *0.48*
Finnish						**0.59** *0.42*	**2.08** *0.59*	**0.69** *0.51*	**1.04** *0.41*	**−0.05** *0.36*	**1.11** *0.43*	**0.80** *0.42*	**−0.12** *0.39*	**0.52** *0.37*	**−0.29** *0.36*	**1.05** *0.41*
French							**1.49** *0.59*	**0.10** *0.49*	**0.45** *0.42*	**−0.64** *0.43*	**0.52** *0.38*	**0.21** *0.39*	**−0.71** *0.36*	**−0.07** *0.44*	**−0.88** *0.45*	**0.46** *0.40*
German								**−1.39** *0.66*	**−1.04** *0.58*	**−2.13** *0.59*	**−0.97** *0.60*	**−1.28** *0.61*	**−2.20** *0.58*	**−1.56** *0.57*	**−2.37** *0.58*	**−1.03** *0.60*
Italian									**0.36** *0.52*	**−0.74** *0.52*	**0.42** *0.51*	**0.11** *0.52*	**−0.81** *0.48*	**−0.16** *0.51*	**−0.98** *0.53*	**0.36** *0.52*
Japanese										**−1.09** *0.39*	**0.07** *0.45*	**−0.25** *0.42*	**−1.16** *0.39*	**−0.52** *0.38*	**−1.33** *0.35*	**0.01** *0.41*
Norwegian											**1.16** *0.45*	**0.85** *0.43*	**−0.07** *0.39*	**0.57** *0.37*	**−0.24** *0.30*	**1.10** *0.43*
Polish												**−0.32** *0.43*	**−1.23** *0.41*	**−0.59** *0.44*	**−1.40** *0.47*	**−0.06** *0.43*
Russian													**−0.92** *0.39*	**−0.27** *0.45*	**−1.09** *0.44*	**0.25** *0.41*
Spanish														**0.64** *0.41*	**−0.17** *0.40*	**1.17** *0.40*
Swedish															**−0.81** *0.36*	**0.53** *0.43*
Tswana																**1.34** *0.40*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0
bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

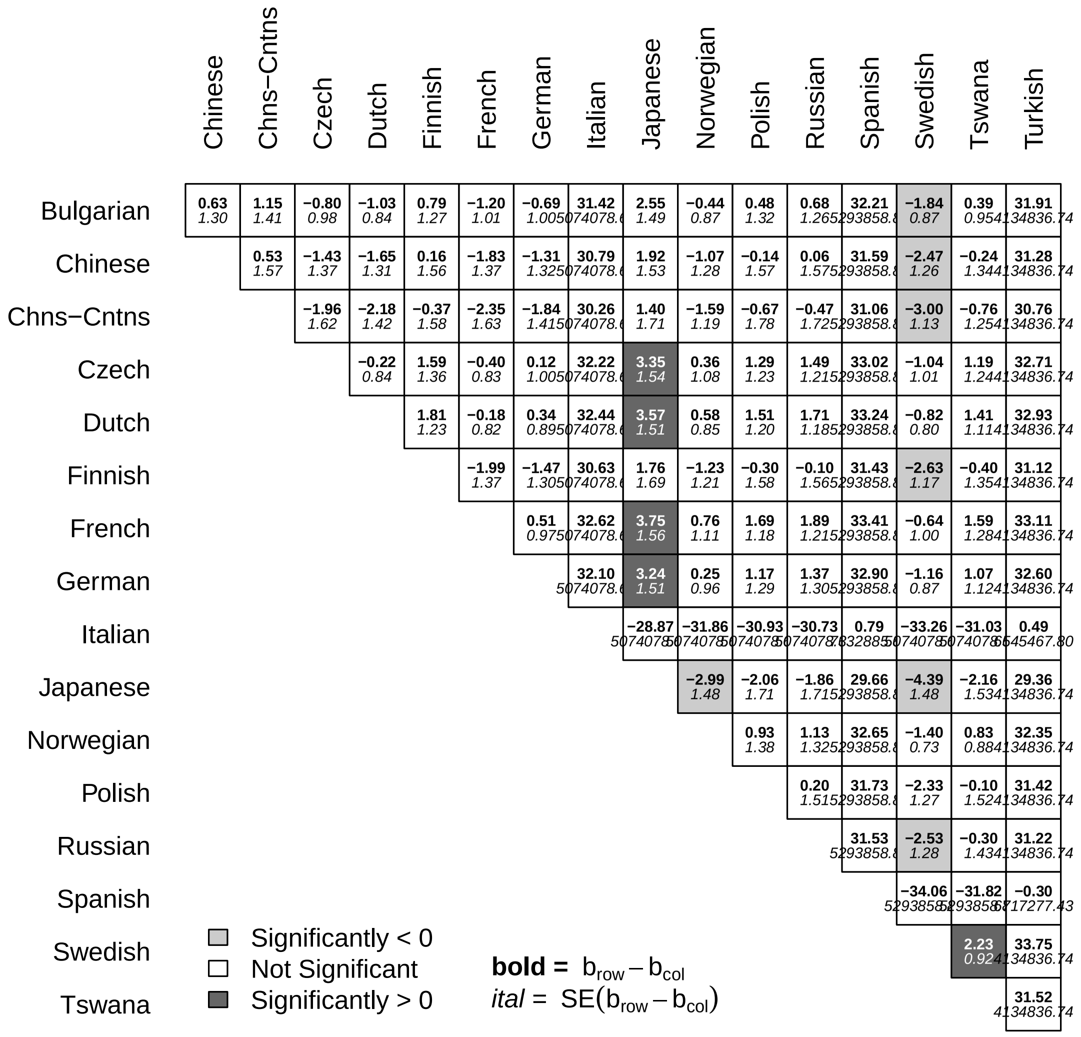
Figure 30: Pairwise Differences in L1 Parameter Values for Going-to (are) Neg. Binomial Type I Regression.

	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**−0.27** *0.14*	**−0.42** *0.14*	**0.48** *0.14*	**−0.21** *0.12*	**0.38** *0.13*	**−0.83** *0.12*	**0.61** *0.13*	**0.60** *0.16*	**0.20** *0.12*	**−0.08** *0.12*	**0.17** *0.13*	**0.44** *0.14*	**0.20** *0.14*	**−0.14** *0.12*	**−0.42** *0.11*	**0.05** *0.12*
Chinese		**−0.15** *0.14*	**0.76** *0.16*	**0.06** *0.15*	**0.66** *0.15*	**−0.55** *0.14*	**0.88** *0.15*	**0.87** *0.17*	**0.47** *0.13*	**0.20** *0.14*	**0.44** *0.14*	**0.71** *0.15*	**0.47** *0.16*	**0.13** *0.14*	**−0.15** *0.13*	**0.32** *0.14*
Chns−Cntns			**0.91** *0.17*	**0.21** *0.15*	**0.81** *0.14*	**−0.40** *0.16*	**1.03** *0.14*	**1.02** *0.18*	**0.62** *0.12*	**0.35** *0.12*	**0.59** *0.16*	**0.86** *0.16*	**0.62** *0.16*	**0.28** *0.11*	**0.00** *0.11*	**0.47** *0.15*
Czech				**−0.69** *0.13*	**−0.10** *0.15*	**−1.31** *0.12*	**0.13** *0.14*	**0.11** *0.17*	**−0.28** *0.14*	**−0.56** *0.15*	**−0.32** *0.13*	**−0.04** *0.14*	**−0.28** *0.15*	**−0.62** *0.14*	**−0.90** *0.15*	**−0.43** *0.13*
Dutch					**0.59** *0.13*	**−0.62** *0.11*	**0.82** *0.13*	**0.81** *0.16*	**0.41** *0.13*	**0.13** *0.13*	**0.38** *0.12*	**0.65** *0.13*	**0.41** *0.14*	**0.07** *0.13*	**−0.21** *0.14*	**0.26** *0.12*
Finnish						**−1.21** *0.14*	**0.23** *0.14*	**0.21** *0.17*	**−0.18** *0.14*	**−0.46** *0.13*	**−0.22** *0.13*	**0.06** *0.15*	**−0.18** *0.15*	**−0.52** *0.12*	**−0.80** *0.13*	**−0.34** *0.13*
French							**1.44** *0.13*	**1.43** *0.15*	**1.03** *0.13*	**0.75** *0.14*	**1.00** *0.10*	**1.27** *0.13*	**1.03** *0.13*	**0.69** *0.13*	**0.41** *0.14*	**0.88** *0.11*
German								**−0.01** *0.16*	**−0.41** *0.13*	**−0.69** *0.13*	**−0.44** *0.13*	**−0.17** *0.14*	**−0.41** *0.15*	**−0.75** *0.12*	**−1.03** *0.13*	**−0.56** *0.13*
Italian									**−0.40** *0.16*	**−0.67** *0.17*	**−0.43** *0.15*	**−0.16** *0.17*	**−0.40** *0.17*	**−0.74** *0.16*	**−1.02** *0.17*	**−0.55** *0.16*
Japanese										**−0.28** *0.12*	**−0.03** *0.13*	**0.24** *0.14*	**0.00** *0.14*	**−0.34** *0.11*	**−0.62** *0.11*	**−0.15** *0.12*
Norwegian											**0.25** *0.13*	**0.52** *0.14*	**0.28** *0.14*	**−0.06** *0.11*	**−0.34** *0.11*	**0.13** *0.13*
Polish												**0.27** *0.13*	**0.03** *0.14*	**−0.31** *0.12*	**−0.59** *0.14*	**−0.12** *0.12*
Russian													**−0.24** *0.15*	**−0.58** *0.14*	**−0.86** *0.15*	**−0.39** *0.13*
Spanish														**−0.34** *0.14*	**−0.62** *0.15*	**−0.15** *0.14*
Swedish															**−0.28** *0.11*	**0.19** *0.12*
Tswana																**0.47** *0.12*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 31: Pairwise Differences in L1 Parameter Values for Will Future Simple Neg. Binomial Type I Regression.

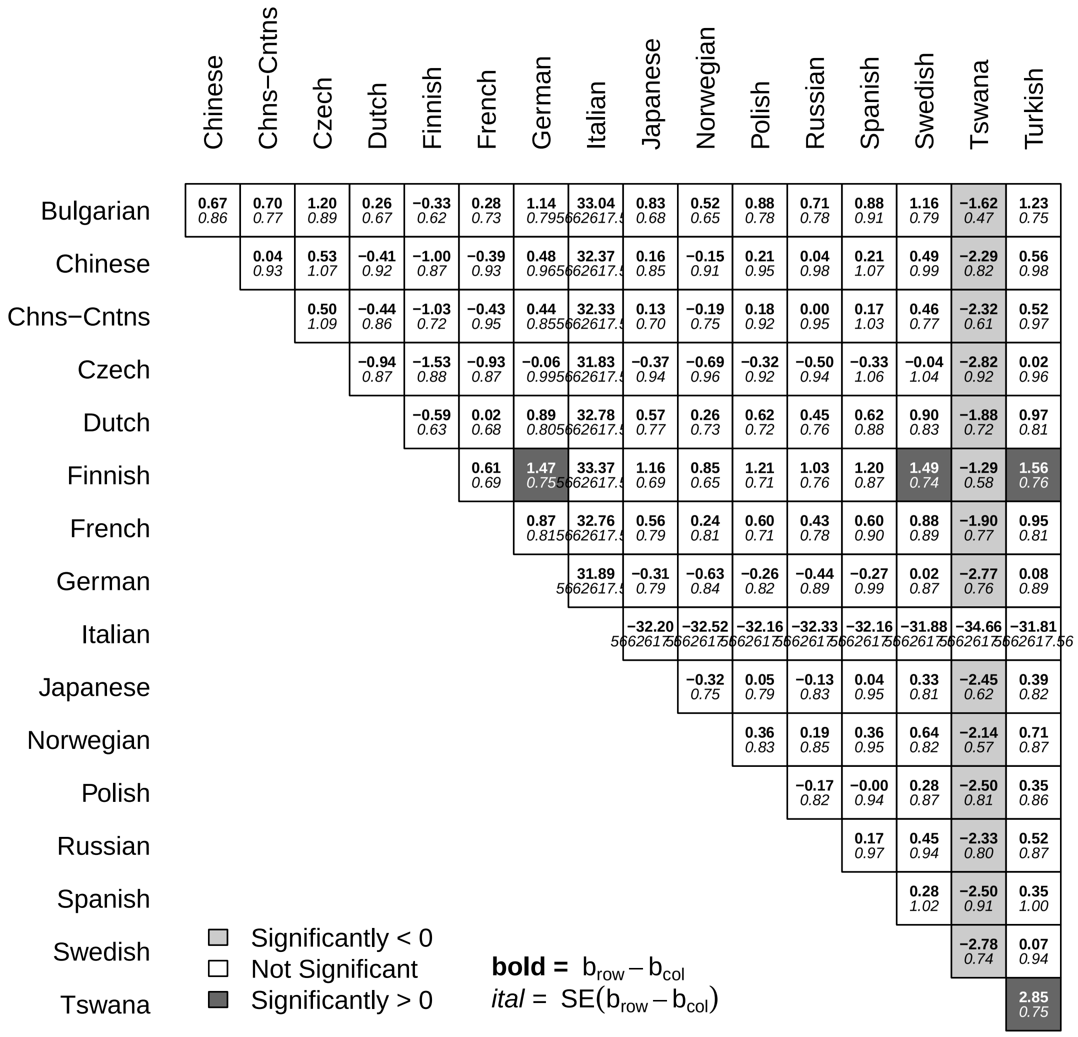


	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**0.63** *1.30*	**1.15** *1.41*	**−0.80** *0.98*	**−1.03** *0.84*	**0.79** *1.27*	**−1.20** *1.01*	**−0.69** *1.00*	**31.42** *5074078.6*	**2.55** *1.49*	**−0.44** *0.87*	**0.48** *1.32*	**0.68** *1.26*	**32.21** *5293858.8*	▨ **−1.84** *0.87*	**0.39** *0.95*	**31.91** *4134836.74*
Chinese		**0.53** *1.57*	**−1.43** *1.37*	**−1.65** *1.31*	**0.16** *1.56*	**−1.83** *1.37*	**−1.31** *1.32*	**30.79** *5074078.6*	**1.92** *1.53*	**−1.07** *1.28*	**−0.14** *1.57*	**0.06** *1.57*	**31.59** *5293858.8*	▨ **−2.47** *1.26*	**−0.24** *1.34*	**31.28** *4134836.74*
Chns−Cntns			**−1.96** *1.62*	**−2.18** *1.42*	**−0.37** *1.58*	**−2.35** *1.63*	**−1.84** *1.41*	**30.26** *5074078.6*	**1.40** *1.71*	**−1.59** *1.19*	**−0.67** *1.78*	**−0.47** *1.72*	**31.06** *5293858.8*	▨ **−3.00** *1.13*	**−0.76** *1.25*	**30.76** *4134836.74*
Czech				**−0.22** *0.84*	**1.59** *1.36*	**−0.40** *0.83*	**0.12** *1.00*	**32.22** *5074078.6*	■ **3.35** *1.54*	**0.36** *1.08*	**1.29** *1.23*	**1.49** *1.21*	**33.02** *5293858.8*	**−1.04** *1.01*	**1.19** *1.24*	**32.71** *4134836.74*
Dutch					**1.81** *1.23*	**−0.18** *0.82*	**0.34** *0.89*	**32.44** *5074078.6*	■ **3.57** *1.51*	**0.58** *0.85*	**1.51** *1.20*	**1.71** *1.18*	**33.24** *5293858.8*	**−0.82** *0.80*	**1.41** *1.11*	**32.93** *4134836.74*
Finnish						**−1.99** *1.37*	**−1.47** *1.30*	**30.63** *5074078.6*	**1.76** *1.69*	**−1.23** *1.21*	**−0.30** *1.58*	**−0.10** *1.56*	**31.43** *5293858.8*	▨ **−2.63** *1.17*	**−0.40** *1.35*	**31.12** *4134836.74*
French							**0.51** *0.97*	**32.62** *5074078.6*	■ **3.75** *1.56*	**0.76** *1.11*	**1.69** *1.18*	**1.89** *1.21*	**33.41** *5293858.8*	**−0.64** *1.00*	**1.59** *1.28*	**33.11** *4134836.74*
German								**32.10** *5074078.6*	■ **3.24** *1.51*	**0.25** *0.96*	**1.17** *1.29*	**1.37** *1.30*	**32.90** *5293858.8*	**−1.16** *0.87*	**1.07** *1.12*	**32.60** *4134836.74*
Italian									**−28.87** *5074078*	**−31.86** *5074078*	**−30.93** *5074078*	**−30.73** *5074078*	**0.79** *7832885*	**−33.26** *5074078*	**−31.03** *5074078*	**0.49** *6645467.80*
Japanese										▨ **−2.99** *1.48*	**−2.06** *1.71*	**−1.86** *1.71*	**29.66** *5293858.8*	▨ **−4.39** *1.48*	**−2.16** *1.53*	**29.36** *4134836.74*
Norwegian											**0.93** *1.38*	**1.13** *1.32*	**32.65** *5293858.8*	**−1.40** *0.73*	**0.83** *0.88*	**32.35** *4134836.74*
Polish												**0.20** *1.51*	**31.73** *5293858.8*	**−2.33** *1.27*	**−0.10** *1.52*	**31.42** *4134836.74*
Russian													**31.53** *5293858.8*	▨ **−2.53** *1.28*	**−0.30** *1.43*	**31.22** *4134836.74*
Spanish														**−34.06** *5293858*	**−31.82** *5293858*	**−0.30** *6717277.43*
Swedish															■ **2.23** *0.92*	**33.75** *4134836.74*
Tswana																**31.52** *4134836.74*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 32: Pairwise Differences in L1 Parameter Values for Will Future Perfect Neg. Binomial Type I Regression.



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**0.67** *0.86*	**0.70** *0.77*	**1.20** *0.89*	**0.26** *0.67*	**−0.33** *0.62*	**0.28** *0.73*	**1.14** *0.79*	**33.04** *5662617.5*	**0.83** *0.68*	**0.52** *0.65*	**0.88** *0.78*	**0.71** *0.78*	**0.88** *0.91*	**1.16** *0.79*	**−1.62** *0.47*	**1.23** *0.75*
Chinese		**0.04** *0.93*	**0.53** *1.07*	**−0.41** *0.92*	**−1.00** *0.87*	**−0.39** *0.93*	**0.48** *0.96*	**32.37** *5662617.5*	**0.16** *0.85*	**−0.15** *0.91*	**0.21** *0.95*	**0.04** *0.98*	**0.21** *1.07*	**0.49** *0.99*	**−2.29** *0.82*	**0.56** *0.98*
Chns−Cntns			**0.50** *1.09*	**−0.44** *0.86*	**−1.03** *0.72*	**−0.43** *0.95*	**0.44** *0.85*	**32.33** *5662617.5*	**0.13** *0.70*	**−0.19** *0.75*	**0.18** *0.92*	**0.00** *0.95*	**0.17** *1.03*	**0.46** *0.77*	**−2.32** *0.61*	**0.52** *0.97*
Czech				**−0.94** *0.87*	**−1.53** *0.88*	**−0.93** *0.87*	**−0.06** *0.99*	**31.83** *5662617.5*	**−0.37** *0.94*	**−0.69** *0.96*	**−0.32** *0.92*	**−0.50** *0.94*	**−0.33** *1.06*	**−0.04** *1.04*	**−2.82** *0.92*	**0.02** *0.96*
Dutch					**−0.59** *0.63*	**0.02** *0.68*	**0.89** *0.80*	**32.78** *5662617.5*	**0.57** *0.77*	**0.26** *0.73*	**0.62** *0.72*	**0.45** *0.76*	**0.62** *0.88*	**0.90** *0.83*	**−1.88** *0.72*	**0.97** *0.81*
Finnish						**0.61** *0.69*	**1.47** *0.75*	**33.37** *5662617.5*	**1.16** *0.69*	**0.85** *0.65*	**1.21** *0.71*	**1.03** *0.76*	**1.20** *0.87*	**1.49** *0.74*	**−1.29** *0.58*	**1.56** *0.76*
French							**0.87** *0.81*	**32.76** *5662617.5*	**0.56** *0.79*	**0.24** *0.81*	**0.60** *0.71*	**0.43** *0.78*	**0.60** *0.90*	**0.88** *0.89*	**−1.90** *0.77*	**0.95** *0.81*
German								**31.89** *5662617.5*	**−0.31** *0.79*	**−0.63** *0.84*	**−0.26** *0.82*	**−0.44** *0.89*	**−0.27** *0.99*	**0.02** *0.87*	**−2.77** *0.76*	**0.08** *0.89*
Italian									**−32.20** *5662617*	**−32.52** *5662617*	**−32.16** *5662617*	**−32.33** *5662617*	**−32.16** *5662617*	**−31.88** *5662617*	**−34.66** *5662617*	**−31.81** *5662617.56*
Japanese										**−0.32** *0.75*	**0.05** *0.79*	**−0.13** *0.83*	**0.04** *0.95*	**0.33** *0.81*	**−2.45** *0.62*	**0.39** *0.82*
Norwegian											**0.36** *0.83*	**0.19** *0.85*	**0.36** *0.95*	**0.64** *0.82*	**−2.14** *0.57*	**0.71** *0.87*
Polish												**−0.17** *0.82*	**−0.00** *0.94*	**0.28** *0.87*	**−2.50** *0.81*	**0.35** *0.86*
Russian													**0.17** *0.97*	**0.45** *0.94*	**−2.33** *0.80*	**0.52** *0.87*
Spanish														**0.28** *1.02*	**−2.50** *0.91*	**0.35** *1.00*
Swedish															**−2.78** *0.74*	**0.07** *0.94*
Tswana																**2.85** *0.75*

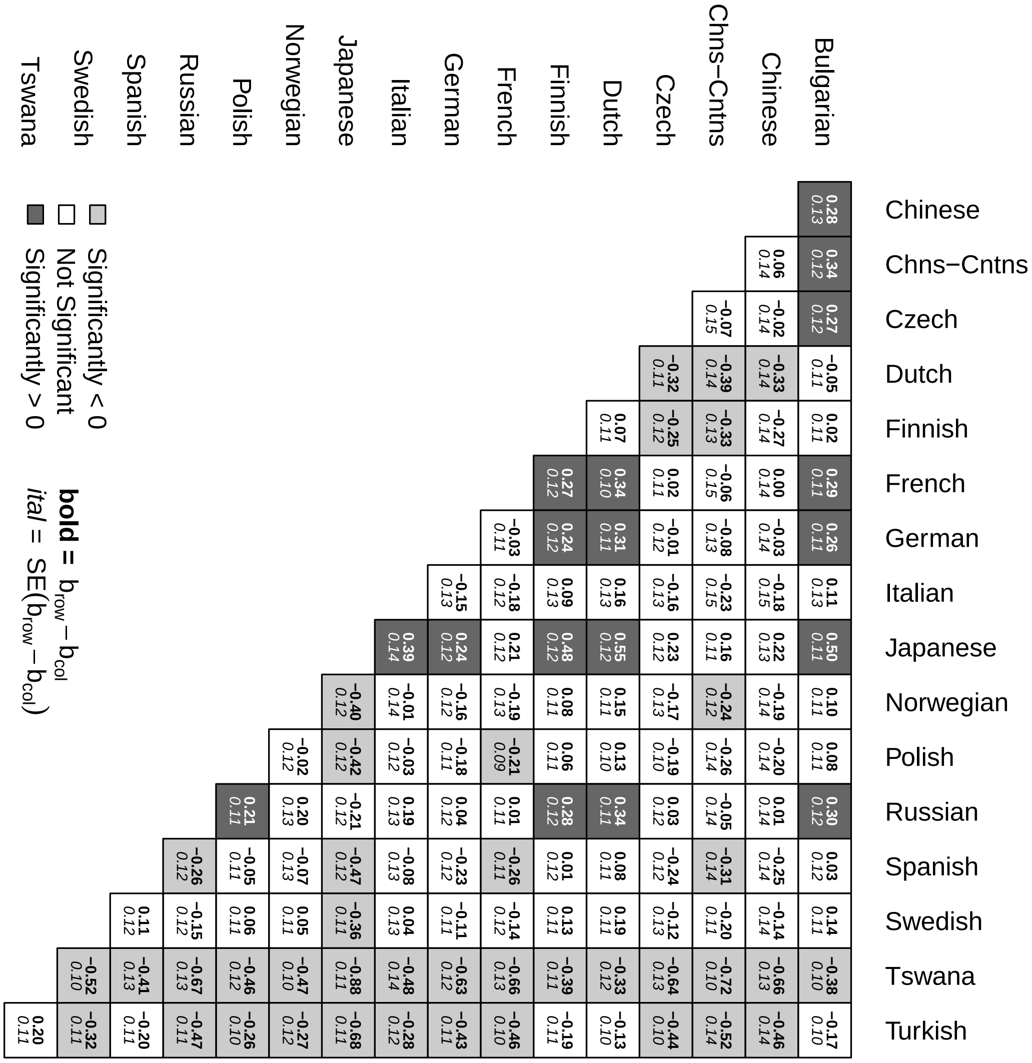
▒ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 33: Pairwise Differences in L1 Parameter Values for Will Future Progressive Neg. Binomial Type I Regression.

7.5.4 The Present Passive

Out of 11 possible Tense and Aspect combinations the pairwise differences in L1 parameter values for 11 combinations could have been plotted. To assure readability the plots of the combined constructions have been omitted.



	Bulgarian	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana
Chinese	**0.28** *0.13*															
Chns−Cntns	**0.34** *0.12*	**0.06** *0.14*														
Czech	**0.27** *0.12*	**−0.02** *0.14*	**−0.07** *0.15*													
Dutch	**−0.05** *0.11*	**−0.33** *0.14*	**−0.39** *0.14*	**−0.32** *0.11*												
Finnish	**0.02** *0.11*	**−0.27** *0.14*	**−0.33** *0.13*	**−0.25** *0.12*	**0.07** *0.11*											
French	**0.29** *0.11*	**0.00** *0.14*	**−0.06** *0.15*	**0.02** *0.11*	**0.34** *0.10*	**0.27** *0.12*										
German	**0.26** *0.11*	**−0.03** *0.14*	**−0.08** *0.13*	**−0.01** *0.12*	**0.31** *0.11*	**0.24** *0.12*	**−0.03** *0.11*									
Italian	**0.11** *0.13*	**−0.18** *0.15*	**−0.23** *0.15*	**−0.16** *0.13*	**0.16** *0.13*	**0.09** *0.13*	**−0.18** *0.12*	**−0.15** *0.13*								
Japanese	**0.50** *0.11*	**0.22** *0.13*	**0.16** *0.11*	**0.23** *0.12*	**0.55** *0.12*	**0.48** *0.12*	**0.21** *0.12*	**0.24** *0.12*	**0.39** *0.14*							
Norwegian	**0.10** *0.11*	**−0.19** *0.14*	**−0.24** *0.12*	**−0.17** *0.13*	**0.15** *0.11*	**0.08** *0.11*	**−0.19** *0.13*	**−0.16** *0.12*	**−0.01** *0.14*	**−0.40** *0.12*						
Polish	**0.08** *0.11*	**−0.20** *0.14*	**−0.26** *0.14*	**−0.19** *0.10*	**0.13** *0.10*	**0.06** *0.11*	**−0.21** *0.09*	**−0.18** *0.11*	**−0.03** *0.12*	**−0.42** *0.12*	**−0.02** *0.12*					
Russian	**0.30** *0.12*	**0.01** *0.14*	**−0.05** *0.14*	**0.03** *0.12*	**0.34** *0.11*	**0.28** *0.12*	**0.01** *0.11*	**0.04** *0.12*	**0.19** *0.13*	**−0.21** *0.12*	**0.20** *0.13*	**0.21** *0.11*				
Spanish	**0.03** *0.12*	**−0.25** *0.14*	**−0.31** *0.14*	**−0.24** *0.12*	**0.08** *0.11*	**0.01** *0.12*	**−0.26** *0.11*	**−0.23** *0.12*	**−0.08** *0.13*	**−0.47** *0.12*	**−0.07** *0.13*	**−0.05** *0.11*	**−0.26** *0.12*			
Swedish	**0.14** *0.11*	**−0.14** *0.14*	**−0.20** *0.11*	**−0.12** *0.13*	**0.19** *0.11*	**0.13** *0.11*	**−0.14** *0.12*	**−0.11** *0.11*	**0.04** *0.13*	**−0.36** *0.11*	**0.05** *0.11*	**0.06** *0.11*	**−0.15** *0.12*	**0.11** *0.12*		
Tswana	**−0.38** *0.10*	**−0.66** *0.13*	**−0.72** *0.10*	**−0.64** *0.13*	**−0.33** *0.12*	**−0.39** *0.11*	**−0.66** *0.13*	**−0.63** *0.12*	**−0.48** *0.14*	**−0.88** *0.11*	**−0.47** *0.10*	**−0.46** *0.12*	**−0.67** *0.13*	**−0.41** *0.13*	**−0.52** *0.10*	
Turkish	**−0.17** *0.10*	**−0.46** *0.14*	**−0.52** *0.14*	**−0.44** *0.10*	**−0.13** *0.10*	**−0.19** *0.11*	**−0.46** *0.10*	**−0.43** *0.11*	**−0.28** *0.12*	**−0.68** *0.11*	**−0.27** *0.12*	**−0.26** *0.10*	**−0.47** *0.11*	**−0.20** *0.11*	**−0.32** *0.11*	**0.20** *0.11*

Figure 34: Pairwise Differences in L1 Parameter Values for Present Simple Passive (are) Neg. Binomial Type I Regression.

	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**0.38** *0.14*	**−0.07** *0.12*	**0.12** *0.11*	**0.06** *0.11*	**−0.15** *0.10*	**−0.05** *0.11*	**0.31** *0.11*	**−0.05** *0.13*	**−0.26** *0.10*	**−0.00** *0.10*	**0.01** *0.11*	**−0.04** *0.11*	**−0.23** *0.11*	**−0.00** *0.10*	**−0.08** *0.10*	**−0.30** *0.10*
Chinese		**−0.45** *0.15*	**−0.26** *0.15*	**−0.32** *0.15*	**−0.53** *0.15*	**−0.43** *0.15*	**−0.07** *0.15*	**−0.43** *0.15*	**−0.64** *0.14*	**−0.38** *0.14*	**−0.37** *0.14*	**−0.42** *0.15*	**−0.61** *0.15*	**−0.38** *0.14*	**−0.47** *0.14*	**−0.68** *0.14*
Chns−Cntns			**0.18** *0.15*	**0.12** *0.14*	**−0.09** *0.12*	**0.02** *0.14*	**0.38** *0.13*	**0.02** *0.14*	**−0.19** *0.10*	**0.06** *0.11*	**0.08** *0.14*	**0.03** *0.14*	**−0.17** *0.13*	**0.06** *0.10*	**−0.02** *0.10*	**−0.23** *0.13*
Czech				**−0.06** *0.11*	**−0.27** *0.12*	**−0.16** *0.10*	**0.19** *0.12*	**−0.17** *0.12*	**−0.38** *0.12*	**−0.12** *0.12*	**−0.11** *0.10*	**−0.16** *0.11*	**−0.35** *0.11*	**−0.12** *0.12*	**−0.20** *0.13*	**−0.41** *0.10*
Dutch					**−0.21** *0.11*	**−0.10** *0.10*	**0.26** *0.12*	**−0.11** *0.12*	**−0.31** *0.12*	**−0.06** *0.11*	**−0.04** *0.10*	**−0.10** *0.11*	**−0.29** *0.11*	**−0.06** *0.11*	**−0.14** *0.12*	**−0.35** *0.11*
Finnish						**0.11** *0.11*	**0.46** *0.11*	**0.10** *0.12*	**−0.11** *0.11*	**0.15** *0.10*	**0.16** *0.11*	**0.11** *0.11*	**−0.08** *0.11*	**0.15** *0.10*	**0.07** *0.11*	**−0.15** *0.10*
French							**0.36** *0.11*	**−0.00** *0.11*	**−0.21** *0.11*	**0.05** *0.12*	**0.06** *0.09*	**0.01** *0.10*	**−0.19** *0.10*	**0.04** *0.11*	**−0.04** *0.13*	**−0.25** *0.10*
German								**−0.36** *0.12*	**−0.57** *0.11*	**−0.31** *0.12*	**−0.30** *0.11*	**−0.35** *0.12*	**−0.54** *0.12*	**−0.32** *0.11*	**−0.40** *0.12*	**−0.61** *0.11*
Italian									**−0.21** *0.12*	**0.05** *0.13*	**0.06** *0.11*	**0.01** *0.12*	**−0.18** *0.12*	**0.05** *0.12*	**−0.04** *0.13*	**−0.25** *0.12*
Japanese										**0.26** *0.11*	**0.27** *0.11*	**0.22** *0.11*	**0.03** *0.11*	**0.25** *0.10*	**0.17** *0.11*	**−0.04** *0.10*
Norwegian											**0.01** *0.12*	**−0.04** *0.12*	**−0.23** *0.12*	**−0.00** *0.10*	**−0.08** *0.10*	**−0.30** *0.12*
Polish												**−0.05** *0.10*	**−0.25** *0.10*	**−0.02** *0.10*	**−0.10** *0.13*	**−0.31** *0.10*
Russian													**−0.19** *0.11*	**0.04** *0.12*	**−0.05** *0.13*	**−0.26** *0.10*
Spanish														**0.23** *0.11*	**0.15** *0.12*	**−0.06** *0.11*
Swedish															**−0.08** *0.10*	**−0.29** *0.11*
Tswana																**−0.21** *0.11*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 35: Pairwise Differences in L1 Parameter Values for Present Simple Passive (is) Neg. Binomial Type I Regression.

	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**0.33** *0.10*	**0.14** *0.09*	**0.21** *0.09*	**0.00** *0.08*	**−0.06** *0.08*	**0.10** *0.08*	**0.29** *0.08*	**0.03** *0.10*	**0.08** *0.08*	**0.05** *0.08*	**0.05** *0.08*	**0.13** *0.09*	**−0.09** *0.09*	**0.07** *0.08*	**−0.26** *0.07*	**−0.24** *0.07*
Chinese		**−0.19** *0.11*	**−0.12** *0.11*	**−0.32** *0.11*	**−0.39** *0.11*	**−0.22** *0.11*	**−0.04** *0.11*	**−0.30** *0.11*	**−0.24** *0.10*	**−0.27** *0.10*	**−0.28** *0.10*	**−0.20** *0.11*	**−0.42** *0.11*	**−0.25** *0.10*	**−0.58** *0.10*	**−0.56** *0.10*
Chns−Cntns			**0.07** *0.11*	**−0.13** *0.10*	**−0.20** *0.09*	**−0.03** *0.11*	**0.15** *0.10*	**−0.11** *0.11*	**−0.05** *0.08*	**−0.08** *0.09*	**−0.08** *0.10*	**−0.01** *0.11*	**−0.23** *0.10*	**−0.06** *0.08*	**−0.39** *0.08*	**−0.37** *0.10*
Czech				**−0.20** *0.08*	**−0.27** *0.09*	**−0.10** *0.08*	**0.08** *0.09*	**−0.18** *0.09*	**−0.12** *0.09*	**−0.15** *0.09*	**−0.16** *0.08*	**−0.08** *0.09*	**−0.30** *0.09*	**−0.13** *0.09*	**−0.46** *0.10*	**−0.44** *0.08*
Dutch					**−0.06** *0.08*	**0.10** *0.08*	**0.29** *0.09*	**0.03** *0.09*	**0.08** *0.09*	**0.05** *0.08*	**0.05** *0.07*	**0.13** *0.08*	**−0.09** *0.09*	**0.07** *0.08*	**−0.26** *0.09*	**−0.24** *0.08*
Finnish						**0.16** *0.09*	**0.35** *0.09*	**0.09** *0.09*	**0.15** *0.09*	**0.12** *0.08*	**0.11** *0.08*	**0.19** *0.09*	**−0.03** *0.09*	**0.14** *0.08*	**−0.19** *0.08*	**−0.18** *0.08*
French							**0.19** *0.08*	**−0.07** *0.09*	**−0.02** *0.09*	**−0.05** *0.09*	**−0.05** *0.07*	**0.03** *0.08*	**−0.19** *0.08*	**−0.03** *0.09*	**−0.36** *0.10*	**−0.34** *0.08*
German								**−0.26** *0.09*	**−0.20** *0.08*	**−0.24** *0.09*	**−0.24** *0.08*	**−0.16** *0.09*	**−0.38** *0.09*	**−0.21** *0.08*	**−0.55** *0.09*	**−0.53** *0.08*
Italian									**0.05** *0.10*	**0.02** *0.10*	**0.02** *0.08*	**0.10** *0.10*	**−0.12** *0.09*	**0.04** *0.09*	**−0.29** *0.10*	**−0.27** *0.09*
Japanese										**−0.03** *0.09*	**−0.03** *0.08*	**0.04** *0.09*	**−0.18** *0.09*	**−0.01** *0.08*	**−0.34** *0.08*	**−0.32** *0.08*
Norwegian											**−0.00** *0.09*	**0.08** *0.09*	**−0.14** *0.09*	**0.02** *0.08*	**−0.31** *0.08*	**−0.29** *0.09*
Polish												**0.08** *0.08*	**−0.14** *0.08*	**0.02** *0.08*	**−0.31** *0.09*	**−0.29** *0.07*
Russian													**−0.22** *0.09*	**−0.05** *0.09*	**−0.39** *0.10*	**−0.37** *0.08*
Spanish														**0.17** *0.09*	**−0.17** *0.09*	**−0.15** *0.08*
Swedish															**−0.33** *0.08*	**−0.31** *0.08*
Tswana																**0.02** *0.08*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 36: Pairwise Differences in L1 Parameter Values for Present Simple Passive Neg. Binomial Type I Regression.

	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**0.43** *0.36*	■ **0.77** *0.33*	**0.21** *0.33*	**−0.05** *0.29*	**−0.22** *0.27*	▨ **−0.67** *0.29*	**0.47** *0.32*	**0.32** *0.38*	**0.47** *0.30*	**0.16** *0.28*	**−0.27** *0.28*	**−0.15** *0.30*	**0.07** *0.33*	**−0.35** *0.26*	**0.46** *0.28*	**0.10** *0.29*
Chinese		**0.34** *0.37*	**−0.21** *0.40*	**−0.48** *0.38*	**−0.64** *0.36*	▨ **−1.10** *0.36*	**0.04** *0.39*	**−0.11** *0.43*	**0.04** *0.36*	**−0.27** *0.36*	**−0.70** *0.36*	**−0.58** *0.37*	**−0.36** *0.39*	▨ **−0.78** *0.34*	**0.03** *0.37*	**−0.33** *0.37*
Chns−Cntns			**−0.56** *0.41*	▨ **−0.82** *0.36*	▨ **−0.99** *0.31*	▨ **−1.44** *0.38*	**−0.31** *0.35*	**−0.45** *0.41*	**−0.30** *0.30*	▨ **−0.61** *0.29*	▨ **−1.05** *0.36*	▨ **−0.93** *0.36*	**−0.70** *0.37*	▨ **−1.12** *0.25*	**−0.31** *0.28*	**−0.67** *0.37*
Czech				**−0.26** *0.31*	**−0.43** *0.33*	▨ **−0.88** *0.29*	**0.25** *0.36*	**0.11** *0.40*	**0.26** *0.35*	**−0.05** *0.35*	**−0.49** *0.30*	**−0.37** *0.31*	**−0.15** *0.35*	**−0.57** *0.33*	**0.25** *0.38*	**−0.11** *0.32*
Dutch					**−0.17** *0.27*	▨ **−0.62** *0.26*	**0.52** *0.33*	**0.37** *0.37*	**0.52** *0.32*	**0.21** *0.29*	**−0.22** *0.25*	**−0.10** *0.28*	**0.12** *0.31*	**−0.30** *0.28*	**0.51** *0.35*	**0.15** *0.30*
Finnish						**−0.45** *0.28*	■ **0.68** *0.31*	**0.54** *0.36*	■ **0.69** *0.31*	**0.38** *0.26*	**−0.06** *0.26*	**0.06** *0.29*	**0.28** *0.31*	**−0.14** *0.23*	■ **0.68** *0.30*	**0.31** *0.28*
French							■ **1.13** *0.31*	■ **0.99** *0.35*	■ **1.14** *0.31*	■ **0.83** *0.31*	**0.39** *0.22*	■ **0.51** *0.26*	■ **0.74** *0.30*	**0.32** *0.28*	■ **1.13** *0.35*	■ **0.77** *0.27*
German								**−0.15** *0.39*	**0.01** *0.33*	**−0.30** *0.32*	▨ **−0.74** *0.30*	**−0.62** *0.33*	**−0.40** *0.35*	▨ **−0.82** *0.29*	**−0.00** *0.35*	**−0.37** *0.33*
Italian									**0.15** *0.39*	**−0.16** *0.38*	**−0.60** *0.34*	**−0.47** *0.36*	**−0.25** *0.38*	**−0.67** *0.35*	**0.14** *0.41*	**−0.22** *0.37*
Japanese										**−0.31** *0.30*	▨ **−0.75** *0.30*	▨ **−0.63** *0.31*	**−0.40** *0.34*	▨ **−0.82** *0.28*	**−0.01** *0.31*	**−0.37** *0.31*
Norwegian											**−0.44** *0.30*	**−0.32** *0.31*	**−0.09** *0.33*	▨ **−0.51** *0.25*	**0.30** *0.28*	**−0.06** *0.32*
Polish												**0.12** *0.26*	**0.34** *0.30*	**−0.08** *0.26*	■ **0.74** *0.34*	**0.37** *0.27*
Russian													**0.22** *0.32*	**−0.20** *0.29*	**0.62** *0.34*	**0.25** *0.29*
Spanish														**−0.42** *0.31*	**0.39** *0.36*	**0.03** *0.33*
Swedish															■ **0.82** *0.27*	**0.45** *0.28*
Tswana																**−0.36** *0.32*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

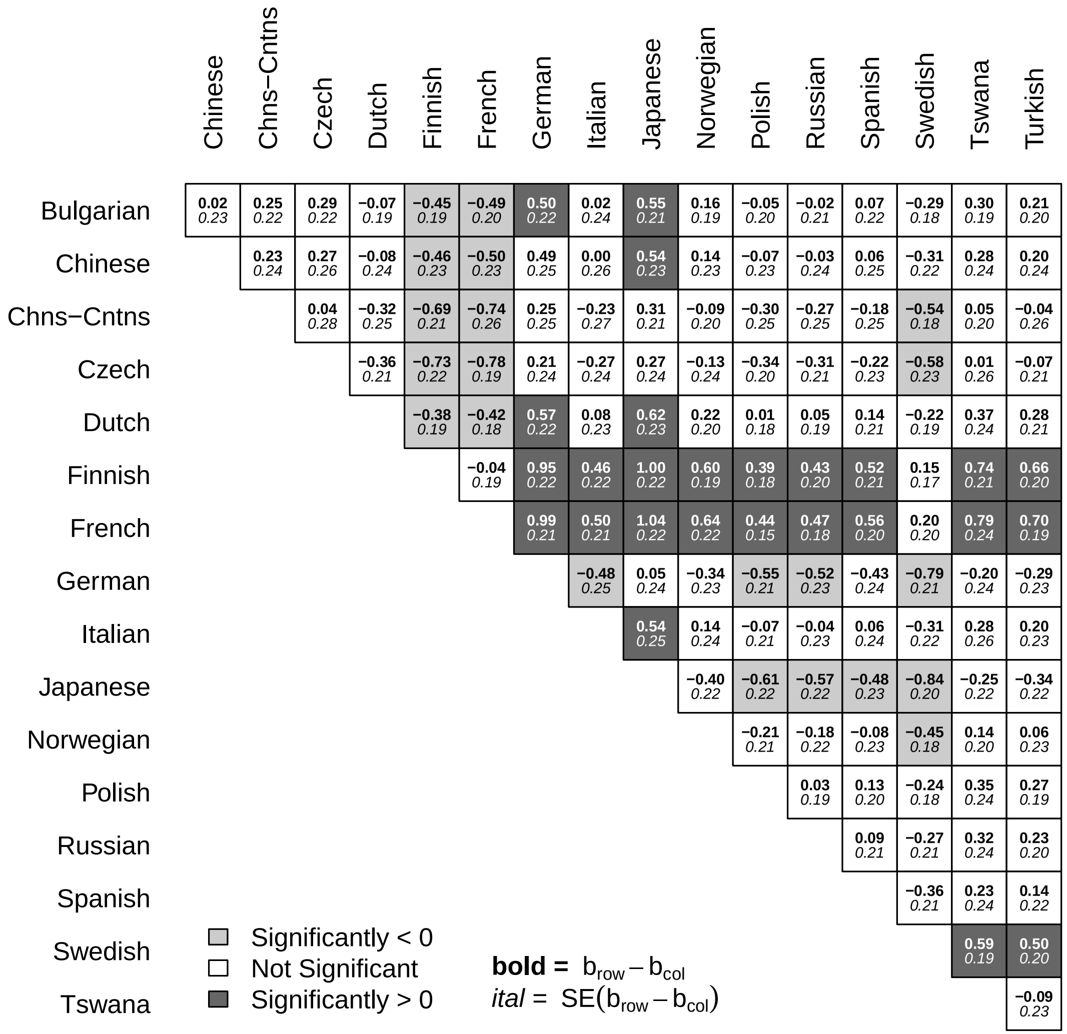
Figure 37: Pairwise Differences in L1 Parameter Values for Present Perfect Passive (have) Neg. Binomial Type I Regression.

	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**−0.25** *0.28*	**−0.10** *0.28*	**0.32** *0.27*	**−0.08** *0.24*	▒ **−0.63** *0.22*	**−0.36** *0.24*	**0.53** *0.28*	**−0.11** *0.29*	■ **0.67** *0.28*	**0.13** *0.25*	**0.10** *0.24*	**0.04** *0.26*	**0.07** *0.27*	**−0.23** *0.23*	**0.20** *0.25*	**0.29** *0.25*
Chinese		**0.15** *0.29*	**0.57** *0.31*	**0.17** *0.29*	**−0.38** *0.28*	**−0.11** *0.28*	■ **0.78** *0.31*	**0.15** *0.31*	■ **0.92** *0.30*	**0.38** *0.29*	**0.36** *0.28*	**0.29** *0.29*	**0.32** *0.30*	**0.02** *0.27*	**0.45** *0.30*	**0.54** *0.30*
Chns−Cntns			**0.43** *0.35*	**0.02** *0.31*	▒ **−0.52** *0.26*	**−0.25** *0.32*	■ **0.63** *0.32*	**−0.00** *0.33*	■ **0.77** *0.27*	**0.23** *0.26*	**0.21** *0.31*	**0.14** *0.31*	**0.17** *0.31*	**−0.12** *0.23*	**0.30** *0.25*	**0.40** *0.32*
Czech				**−0.40** *0.25*	▒ **−0.95** *0.27*	▒ **−0.68** *0.23*	**0.21** *0.30*	**−0.43** *0.29*	**0.34** *0.31*	**−0.19** *0.30*	**−0.22** *0.25*	**−0.28** *0.26*	**−0.25** *0.28*	**−0.55** *0.28*	**−0.12** *0.32*	**−0.03** *0.27*
Dutch					▒ **−0.55** *0.22*	**−0.28** *0.21*	■ **0.61** *0.28*	**−0.02** *0.27*	■ **0.75** *0.30*	**0.21** *0.26*	**0.19** *0.22*	**0.12** *0.24*	**0.15** *0.25*	**−0.14** *0.24*	**0.28** *0.30*	**0.38** *0.26*
Finnish						**0.27** *0.23*	■ **1.16** *0.27*	■ **0.52** *0.26*	■ **1.29** *0.28*	■ **0.75** *0.23*	■ **0.73** *0.23*	■ **0.67** *0.25*	■ **0.69** *0.25*	**0.40** *0.21*	■ **0.83** *0.26*	■ **0.92** *0.24*
French							■ **0.89** *0.27*	**0.25** *0.25*	■ **1.02** *0.29*	**0.48** *0.27*	■ **0.46** *0.19*	**0.40** *0.23*	**0.42** *0.24*	**0.13** *0.25*	**0.56** *0.31*	■ **0.65** *0.24*
German								▒ **−0.63** *0.30*	**0.14** *0.31*	**−0.40** *0.30*	**−0.42** *0.27*	**−0.49** *0.29*	**−0.46** *0.30*	▒ **−0.76** *0.27*	**−0.33** *0.31*	**−0.24** *0.29*
Italian									■ **0.77** *0.32*	**0.23** *0.30*	**0.21** *0.25*	**0.14** *0.28*	**0.17** *0.28*	**−0.12** *0.27*	**0.30** *0.32*	**0.40** *0.28*
Japanese										**−0.54** *0.30*	**−0.56** *0.29*	▒ **−0.63** *0.29*	▒ **−0.60** *0.30*	▒ **−0.89** *0.27*	**−0.47** *0.30*	**−0.37** *0.30*
Norwegian											**−0.02** *0.27*	**−0.09** *0.28*	**−0.06** *0.28*	**−0.35** *0.24*	**0.07** *0.26*	**0.17** *0.29*
Polish												**−0.07** *0.23*	**−0.04** *0.25*	**−0.33** *0.23*	**0.09** *0.30*	**0.19** *0.24*
Russian													**0.03** *0.26*	**−0.27** *0.26*	**0.16** *0.30*	**0.25** *0.26*
Spanish														**−0.29** *0.26*	**0.13** *0.31*	**0.23** *0.28*
Swedish															**0.43** *0.25*	■ **0.52** *0.26*
Tswana																**0.09** *0.29*

▒ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 38: Pairwise Differences in L1 Parameter Values for Present Perfect Passive (has) Neg. Binomial Type I Regression.



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**0.02** *0.23*	**0.25** *0.22*	**0.29** *0.22*	**−0.07** *0.19*	▒ **−0.45** *0.19*	▒ **−0.49** *0.20*	■ **0.50** *0.22*	**0.02** *0.24*	■ **0.55** *0.21*	**0.16** *0.19*	**−0.05** *0.20*	**−0.02** *0.21*	**0.07** *0.22*	**−0.29** *0.18*	**0.30** *0.19*	**0.21** *0.20*
Chinese		**0.23** *0.24*	**0.27** *0.26*	**−0.08** *0.24*	▒ **−0.46** *0.23*	▒ **−0.50** *0.23*	**0.49** *0.25*	**0.00** *0.26*	■ **0.54** *0.23*	**0.14** *0.23*	**−0.07** *0.23*	**−0.03** *0.24*	**0.06** *0.25*	**−0.31** *0.22*	**0.28** *0.24*	**0.20** *0.24*
Chns−Cntns			**0.04** *0.28*	**−0.32** *0.25*	▒ **−0.69** *0.21*	▒ **−0.74** *0.26*	**0.25** *0.25*	**−0.23** *0.27*	**0.31** *0.21*	**−0.09** *0.20*	**−0.30** *0.25*	**−0.27** *0.25*	**−0.18** *0.25*	▒ **−0.54** *0.18*	**0.05** *0.20*	**−0.04** *0.26*
Czech				**−0.36** *0.21*	▒ **−0.73** *0.22*	▒ **−0.78** *0.19*	**0.21** *0.24*	**−0.27** *0.24*	**0.27** *0.24*	**−0.13** *0.24*	**−0.34** *0.20*	**−0.31** *0.21*	**−0.22** *0.23*	▒ **−0.58** *0.23*	**0.01** *0.26*	**−0.07** *0.21*
Dutch					▒ **−0.38** *0.19*	▒ **−0.42** *0.18*	■ **0.57** *0.22*	**0.08** *0.23*	■ **0.62** *0.23*	**0.22** *0.20*	**0.01** *0.18*	**0.05** *0.19*	**0.14** *0.21*	**−0.22** *0.19*	**0.37** *0.24*	**0.28** *0.21*
Finnish						**−0.04** *0.19*	■ **0.95** *0.22*	■ **0.46** *0.22*	■ **1.00** *0.22*	■ **0.60** *0.19*	■ **0.39** *0.18*	■ **0.43** *0.20*	■ **0.52** *0.21*	**0.15** *0.17*	■ **0.74** *0.21*	■ **0.66** *0.20*
French							■ **0.99** *0.21*	■ **0.50** *0.21*	■ **1.04** *0.22*	■ **0.64** *0.22*	■ **0.44** *0.15*	■ **0.47** *0.18*	■ **0.56** *0.20*	**0.20** *0.20*	■ **0.79** *0.24*	■ **0.70** *0.19*
German								▒ **−0.48** *0.25*	**0.05** *0.24*	**−0.34** *0.23*	▒ **−0.55** *0.21*	▒ **−0.52** *0.23*	**−0.43** *0.24*	▒ **−0.79** *0.21*	**−0.20** *0.24*	**−0.29** *0.23*
Italian									■ **0.54** *0.25*	**0.14** *0.24*	**−0.07** *0.21*	**−0.04** *0.23*	**0.06** *0.24*	**−0.31** *0.22*	**0.28** *0.26*	**0.20** *0.23*
Japanese										**−0.40** *0.22*	▒ **−0.61** *0.22*	▒ **−0.57** *0.22*	▒ **−0.48** *0.23*	▒ **−0.84** *0.20*	**−0.25** *0.22*	**−0.34** *0.22*
Norwegian											**−0.21** *0.21*	**−0.18** *0.22*	**−0.08** *0.23*	▒ **−0.45** *0.18*	**0.14** *0.20*	**0.06** *0.23*
Polish												**0.03** *0.19*	**0.13** *0.20*	**−0.24** *0.18*	**0.35** *0.24*	**0.27** *0.19*
Russian													**0.09** *0.21*	**−0.27** *0.21*	**0.32** *0.24*	**0.23** *0.20*
Spanish														**−0.36** *0.21*	**0.23** *0.24*	**0.14** *0.22*
Swedish															■ **0.59** *0.19*	■ **0.50** *0.20*
Tswana																**−0.09** *0.23*

▒ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

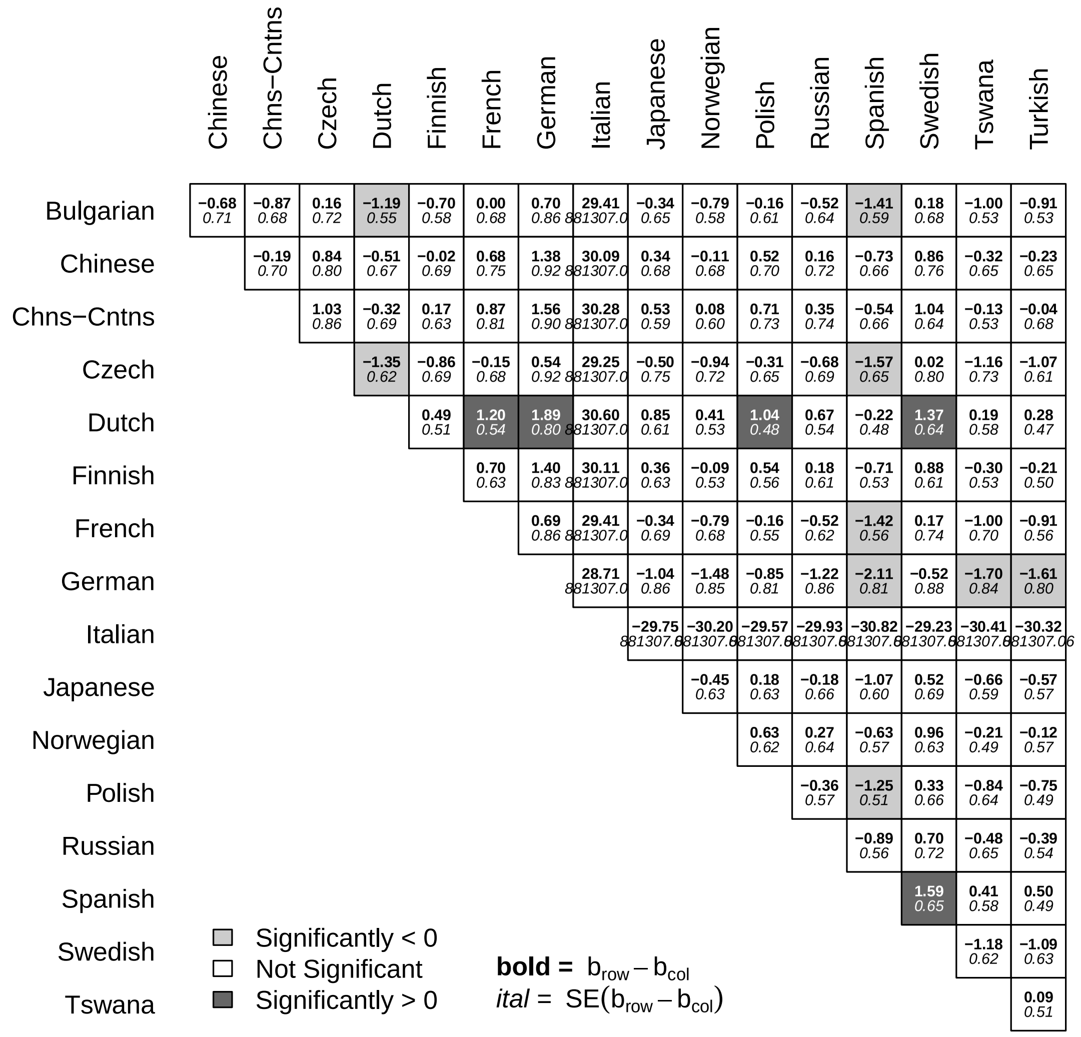
Figure 39: Pairwise Differences in L1 Parameter Values for Present Perfect Passive Neg. Binomial Type I Regression.

	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**0.53** *0.84*	**1.02** *0.70*	**0.57** *0.75*	**−0.98** *0.51*	**−0.69** *0.52*	**−0.03** *0.64*	**0.93** *0.83*	**1.04** *1.12*	**2.08** *1.10*	**−0.58** *0.50*	**−0.69** *0.54*	**0.63** *0.74*	**−0.51** *0.62*	**−0.92** *0.50*	**−0.67** *0.46*	**−0.69** *0.50*
Chinese		**0.50** *0.91*	**0.04** *0.99*	**−1.51** *0.84*	**−1.22** *0.84*	**−0.56** *0.91*	**0.40** *1.04*	**0.51** *1.28*	**1.55** *1.26*	**−1.11** *0.82*	**−1.22** *0.84*	**0.10** *0.98*	**−1.04** *0.88*	**−1.45** *0.81*	**−1.20** *0.81*	**−1.22** *0.83*
Chns−Cntns			**−0.45** *0.95*	▨ **−2.00** *0.73*	▨ **−1.72** *0.66*	**−1.06** *0.86*	**−0.09** *0.92*	**0.02** *1.20*	**1.05** *1.14*	▨ **−1.61** *0.60*	▨ **−1.72** *0.75*	**−0.40** *0.91*	▨ **−1.54** *0.76*	▨ **−1.94** *0.56*	▨ **−1.69** *0.57*	▨ **−1.72** *0.76*
Czech				▨ **−1.55** *0.69*	**−1.26** *0.74*	**−0.60** *0.75*	**0.36** *0.97*	**0.47** *1.20*	**1.51** *1.22*	**−1.15** *0.76*	**−1.26** *0.69*	**0.05** *0.86*	**−1.08** *0.78*	▨ **−1.49** *0.75*	**−1.24** *0.79*	**−1.26** *0.68*
Dutch					**0.29** *0.48*	**0.95** *0.56*	■ **1.91** *0.80*	**2.02** *1.08*	■ **3.06** *1.11*	**0.40** *0.48*	**0.29** *0.44*	■ **1.60** *0.69*	**0.47** *0.56*	**0.06** *0.48*	**0.31** *0.56*	**0.29** *0.49*
Finnish						**0.66** *0.62*	■ **1.62** *0.81*	**1.73** *1.09*	■ **2.77** *1.11*	**0.11** *0.46*	**−0.00** *0.50*	**1.32** *0.73*	**0.18** *0.58*	**−0.23** *0.44*	**0.02** *0.50*	**−0.00** *0.50*
French							**0.96** *0.87*	**1.07** *1.12*	**2.11** *1.16*	**−0.55** *0.66*	**−0.66** *0.52*	**0.66** *0.76*	**−0.48** *0.65*	**−0.89** *0.63*	**−0.64** *0.69*	**−0.66** *0.56*
German								**0.11** *1.25*	**1.15** *1.26*	**−1.51** *0.81*	▨ **−1.62** *0.79*	**−0.30** *0.96*	**−1.44** *0.86*	▨ **−1.85** *0.78*	**−1.60** *0.82*	▨ **−1.62** *0.81*
Italian									**1.04** *1.47*	**−1.62** *1.10*	**−1.73** *1.07*	**−0.42** *1.20*	**−1.55** *1.12*	**−1.96** *1.08*	**−1.71** *1.12*	**−1.73** *1.08*
Japanese										▨ **−2.66** *1.09*	▨ **−2.77** *1.10*	**−1.45** *1.21*	▨ **−2.59** *1.14*	▨ **−3.00** *1.08*	▨ **−2.74** *1.08*	▨ **−2.77** *1.09*
Norwegian											**−0.11** *0.55*	**1.21** *0.74*	**0.07** *0.59*	**−0.34** *0.43*	**−0.08** *0.42*	**−0.11** *0.55*
Polish												**1.32** *0.69*	**0.18** *0.56*	**−0.22** *0.49*	**0.03** *0.59*	**0.00** *0.45*
Russian													**−1.14** *0.77*	▨ **−1.54** *0.73*	**−1.29** *0.76*	**−1.32** *0.69*
Spanish														**−0.41** *0.58*	**−0.15** *0.62*	**−0.18** *0.58*
Swedish															**0.25** *0.45*	**0.23** *0.51*
Tswana																**−0.03** *0.52*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = SE($b_{row} - b_{col}$)

Figure 40: Pairwise Differences in L1 Parameter Values for Present Progressive Passive (are) Neg. Binomial Type I Regression.

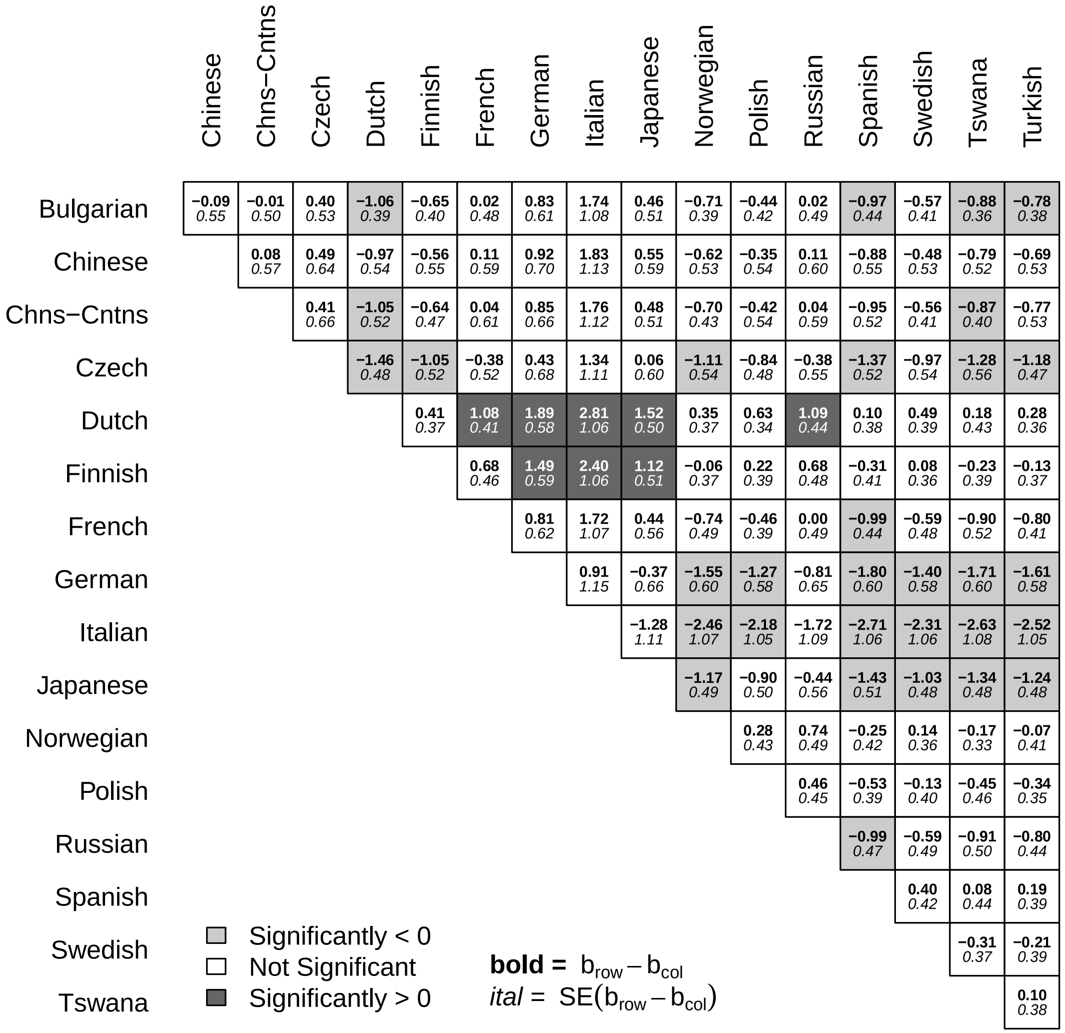


	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**−0.68** *0.71*	**−0.87** *0.68*	**0.16** *0.72*	▨ **−1.19** *0.55*	**−0.70** *0.58*	**0.00** *0.68*	**0.70** *0.86*	**29.41** *881307.0*	**−0.34** *0.65*	**−0.79** *0.58*	**−0.16** *0.61*	**−0.52** *0.64*	▨ **−1.41** *0.59*	**0.18** *0.68*	**−1.00** *0.53*	**−0.91** *0.53*
Chinese		**−0.19** *0.70*	**0.84** *0.80*	**−0.51** *0.67*	**−0.02** *0.69*	**0.68** *0.75*	**1.38** *0.92*	**30.09** *881307.0*	**0.34** *0.68*	**−0.11** *0.68*	**0.52** *0.70*	**0.16** *0.72*	**−0.73** *0.66*	**0.86** *0.76*	**−0.32** *0.65*	**−0.23** *0.65*
Chns−Cntns			**1.03** *0.86*	**−0.32** *0.69*	**0.17** *0.63*	**0.87** *0.81*	**1.56** *0.90*	**30.28** *881307.0*	**0.53** *0.59*	**0.08** *0.60*	**0.71** *0.73*	**0.35** *0.74*	**−0.54** *0.66*	**1.04** *0.64*	**−0.13** *0.53*	**−0.04** *0.68*
Czech				▨ **−1.35** *0.62*	**−0.86** *0.69*	**−0.15** *0.68*	**0.54** *0.92*	**29.25** *881307.0*	**−0.50** *0.75*	**−0.94** *0.72*	**−0.31** *0.65*	**−0.68** *0.69*	▨ **−1.57** *0.65*	**0.02** *0.80*	**−1.16** *0.73*	**−1.07** *0.61*
Dutch					**0.49** *0.51*	■ **1.20** *0.54*	■ **1.89** *0.80*	**30.60** *881307.0*	**0.85** *0.61*	**0.41** *0.53*	■ **1.04** *0.48*	**0.67** *0.54*	**−0.22** *0.48*	■ **1.37** *0.64*	**0.19** *0.58*	**0.28** *0.47*
Finnish						**0.70** *0.63*	**1.40** *0.83*	**30.11** *881307.0*	**0.36** *0.63*	**−0.09** *0.53*	**0.54** *0.56*	**0.18** *0.61*	**−0.71** *0.53*	**0.88** *0.61*	**−0.30** *0.53*	**−0.21** *0.50*
French							**0.69** *0.86*	**29.41** *881307.0*	**−0.34** *0.69*	**−0.79** *0.68*	**−0.16** *0.55*	**−0.52** *0.62*	▨ **−1.42** *0.56*	**0.17** *0.74*	**−1.00** *0.70*	**−0.91** *0.56*
German								**28.71** *881307.0*	**−1.04** *0.86*	**−1.48** *0.85*	**−0.85** *0.81*	**−1.22** *0.86*	▨ **−2.11** *0.81*	**−0.52** *0.88*	▨ **−1.70** *0.84*	▨ **−1.61** *0.80*
Italian									**−29.75** *881307.06*	**−30.20** *881307.06*	**−29.57** *881307.06*	**−29.93** *881307.06*	**−30.82** *881307.06*	**−29.23** *881307.06*	**−30.41** *881307.06*	**−30.32** *881307.06*
Japanese										**−0.45** *0.63*	**0.18** *0.63*	**−0.18** *0.66*	**−1.07** *0.60*	**0.52** *0.69*	**−0.66** *0.59*	**−0.57** *0.57*
Norwegian											**0.63** *0.62*	**0.27** *0.64*	**−0.63** *0.57*	**0.96** *0.63*	**−0.21** *0.49*	**−0.12** *0.57*
Polish												**−0.36** *0.57*	▨ **−1.25** *0.51*	**0.33** *0.66*	**−0.84** *0.64*	**−0.75** *0.49*
Russian													**−0.89** *0.56*	**0.70** *0.72*	**−0.48** *0.65*	**−0.39** *0.54*
Spanish														■ **1.59** *0.65*	**0.41** *0.58*	**0.50** *0.49*
Swedish															**−1.18** *0.62*	**−1.09** *0.63*
Tswana																**0.09** *0.51*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 41: Pairwise Differences in L1 Parameter Values for Present Progressive Passive (is) Neg. Binomial Type I Regression.

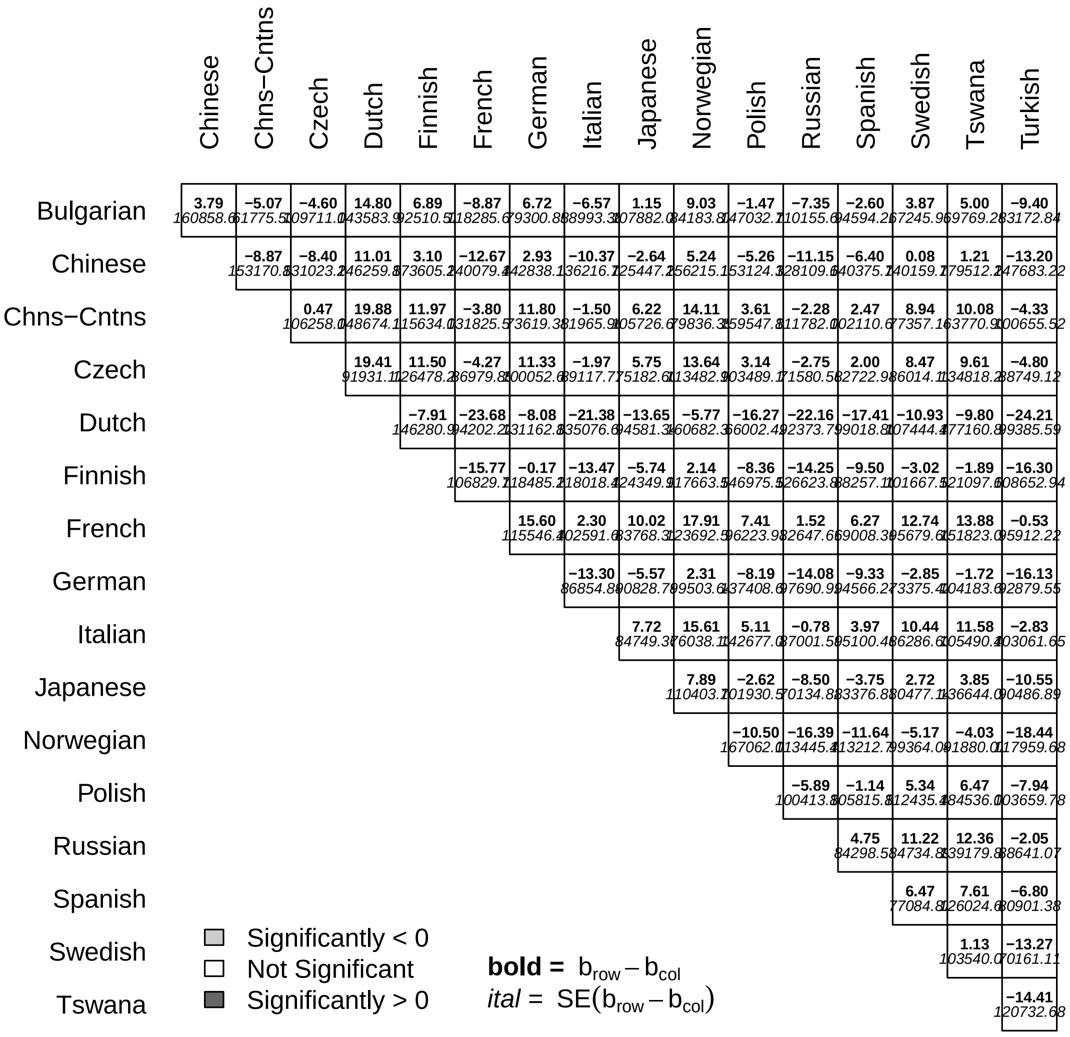


	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**−0.09** *0.55*	**−0.01** *0.50*	**0.40** *0.53*	**−1.06** *0.39*	**−0.65** *0.40*	**0.02** *0.48*	**0.83** *0.61*	**1.74** *1.08*	**0.46** *0.51*	**−0.71** *0.39*	**−0.44** *0.42*	**0.02** *0.49*	**−0.97** *0.44*	**−0.57** *0.41*	**−0.88** *0.36*	**−0.78** *0.38*
Chinese		**0.08** *0.57*	**0.49** *0.64*	**−0.97** *0.54*	**−0.56** *0.55*	**0.11** *0.59*	**0.92** *0.70*	**1.83** *1.13*	**0.55** *0.59*	**−0.62** *0.53*	**−0.35** *0.54*	**0.11** *0.60*	**−0.88** *0.55*	**−0.48** *0.53*	**−0.79** *0.52*	**−0.69** *0.53*
Chns−Cntns			**0.41** *0.66*	**−1.05** *0.52*	**−0.64** *0.47*	**0.04** *0.61*	**0.85** *0.66*	**1.76** *1.12*	**0.48** *0.51*	**−0.70** *0.43*	**−0.42** *0.54*	**0.04** *0.59*	**−0.95** *0.52*	**−0.56** *0.41*	**−0.87** *0.40*	**−0.77** *0.53*
Czech				**−1.46** *0.48*	**−1.05** *0.52*	**−0.38** *0.52*	**0.43** *0.68*	**1.34** *1.11*	**0.06** *0.60*	**−1.11** *0.54*	**−0.84** *0.48*	**−0.38** *0.55*	**−1.37** *0.52*	**−0.97** *0.54*	**−1.28** *0.56*	**−1.18** *0.47*
Dutch					**0.41** *0.37*	**1.08** *0.41*	**1.89** *0.58*	**2.81** *1.06*	**1.52** *0.50*	**0.35** *0.37*	**0.63** *0.34*	**1.09** *0.44*	**0.10** *0.38*	**0.49** *0.39*	**0.18** *0.43*	**0.28** *0.36*
Finnish						**0.68** *0.46*	**1.49** *0.59*	**2.40** *1.06*	**1.12** *0.51*	**−0.06** *0.37*	**0.22** *0.39*	**0.68** *0.48*	**−0.31** *0.41*	**0.08** *0.36*	**−0.23** *0.39*	**−0.13** *0.37*
French							**0.81** *0.62*	**1.72** *1.07*	**0.44** *0.56*	**−0.74** *0.49*	**−0.46** *0.39*	**0.00** *0.49*	**−0.99** *0.44*	**−0.59** *0.48*	**−0.90** *0.52*	**−0.80** *0.41*
German								**0.91** *1.15*	**−0.37** *0.66*	**−1.55** *0.60*	**−1.27** *0.58*	**−0.81** *0.65*	**−1.80** *0.60*	**−1.40** *0.58*	**−1.71** *0.60*	**−1.61** *0.58*
Italian									**−1.28** *1.11*	**−2.46** *1.07*	**−2.18** *1.05*	**−1.72** *1.09*	**−2.71** *1.06*	**−2.31** *1.06*	**−2.63** *1.08*	**−2.52** *1.05*
Japanese										**−1.17** *0.49*	**−0.90** *0.50*	**−0.44** *0.56*	**−1.43** *0.51*	**−1.03** *0.48*	**−1.34** *0.48*	**−1.24** *0.48*
Norwegian											**0.28** *0.43*	**0.74** *0.49*	**−0.25** *0.42*	**0.14** *0.36*	**−0.17** *0.33*	**−0.07** *0.41*
Polish												**0.46** *0.45*	**−0.53** *0.39*	**−0.13** *0.40*	**−0.45** *0.46*	**−0.34** *0.35*
Russian													**−0.99** *0.47*	**−0.59** *0.49*	**−0.91** *0.50*	**−0.80** *0.44*
Spanish														**0.40** *0.42*	**0.08** *0.44*	**0.19** *0.39*
Swedish															**−0.31** *0.37*	**−0.21** *0.39*
Tswana																**0.10** *0.38*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 42: Pairwise Differences in L1 Parameter Values for Present Progressive Passive Neg. Binomial Type I Regression.



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**3.79** *160858.6*	**−5.07** *61775.5*	**−4.60** *109711.0*	**14.80** *143583.9*	**6.89** *92510.5*	**−8.87** *118285.6*	**6.72** *79300.8*	**−6.57** *88993.3*	**1.15** *107882.0*	**9.03** *84183.8*	**−1.47** *147032.1*	**−7.35** *110155.6*	**−2.60** *94594.2*	**3.87** *67245.9*	**5.00** *69769.2*	**−9.40** *83172.84*
Chinese		**−8.87** *153170.8*	**−8.40** *131023.2*	**11.01** *146259.8*	**3.10** *173605.2*	**−12.67** *140079.4*	**2.93** *142838.1*	**−10.37** *136216.1*	**−2.64** *125447.2*	**5.24** *156215.1*	**−5.26** *153124.3*	**−11.15** *128109.6*	**−6.40** *140375.1*	**0.08** *140159.1*	**1.21** *179512.2*	**−13.20** *147683.22*
Chns−Cntns			**0.47** *106258.0*	**19.88** *148674.1*	**11.97** *115634.0*	**−3.80** *131825.5*	**11.80** *73619.3*	**−1.50** *81965.9*	**6.22** *105726.6*	**14.11** *79836.3*	**3.61** *159547.8*	**−2.28** *111782.0*	**2.47** *102110.6*	**8.94** *77357.1*	**10.08** *63770.9*	**−4.33** *100655.52*
Czech				**19.41** *91931.1*	**11.50** *126478.2*	**−4.27** *86979.8*	**11.33** *100052.6*	**−1.97** *89117.7*	**5.75** *75182.6*	**13.64** *113482.9*	**3.14** *103489.1*	**−2.75** *71580.5*	**2.00** *82722.9*	**8.47** *86014.1*	**9.61** *134818.2*	**−4.80** *88749.12*
Dutch					**−7.91** *146280.9*	**−23.68** *94202.2*	**−8.08** *131162.8*	**−21.38** *135076.6*	**−13.65** *94581.3*	**−5.77** *160682.3*	**−16.27** *66002.4*	**−22.16** *92373.7*	**−17.41** *99018.8*	**−10.93** *107444.4*	**−9.80** *177160.8*	**−24.21** *99385.59*
Finnish						**−15.77** *106829.1*	**−0.17** *118485.2*	**−13.47** *118018.4*	**−5.74** *124349.9*	**2.14** *117663.5*	**−8.36** *146975.5*	**−14.25** *126623.8*	**−9.50** *88257.1*	**−3.02** *101667.5*	**−1.89** *121097.6*	**−16.30** *108652.94*
French							**15.60** *115546.4*	**2.30** *102591.6*	**10.02** *83768.3*	**17.91** *123692.5*	**7.41** *96223.9*	**1.52** *82647.6*	**6.27** *69008.3*	**12.74** *95679.6*	**13.88** *151823.0*	**−0.53** *95912.22*
German								**−13.30** *86854.8*	**−5.57** *90828.7*	**2.31** *99503.6*	**−8.19** *137408.6*	**−14.08** *97690.9*	**−9.33** *94566.2*	**−2.85** *73375.4*	**−1.72** *104183.6*	**−16.13** *92879.55*
Italian									**7.72** *84749.3*	**15.61** *76038.1*	**5.11** *142677.0*	**−0.78** *87001.5*	**3.97** *95100.4*	**10.44** *86286.6*	**11.58** *105490.4*	**−2.83** *103061.65*
Japanese										**7.89** *110403.1*	**−2.62** *101930.5*	**−8.50** *70134.8*	**−3.75** *83376.8*	**2.72** *80477.1*	**3.85** *136644.0*	**−10.55** *90486.89*
Norwegian											**−10.50** *167062.0*	**−16.39** *113445.4*	**−11.64** *113212.7*	**−5.17** *99364.0*	**−4.03** *91880.0*	**−18.44** *117959.68*
Polish												**−5.89** *100413.8*	**−1.14** *105815.8*	**5.34** *112435.4*	**6.47** *184536.0*	**−7.94** *103659.78*
Russian													**4.75** *84298.5*	**11.22** *84734.8*	**12.36** *139179.8*	**−2.05** *88641.07*
Spanish														**6.47** *77084.8*	**7.61** *126024.6*	**−6.80** *80901.38*
Swedish															**1.13** *103540.0*	**−13.27** *70161.11*
Tswana																**−14.41** *120732.68*

□ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 43: Pairwise Differences in L1 Parameter Values for Present Perfect Progressive Passive (has) Neg. Binomial Type I Regression.

7.5.5 The Past Passive

Out of 5 possible Tense and Aspect combinations the pairwise differences in L1 parameter values for 1 combinations could have been plotted. To assure readability the plots of the combined constructions have been omitted.

	Bulgarian	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana
Chinese	**0.20** *0.21*															
Chns−Cntns	**0.19** *0.19*	**−0.01** *0.21*														
Czech	**−0.38** *0.17*	**−0.58** *0.22*	**−0.57** *0.22*													
Dutch	**−0.58** *0.16*	**−0.78** *0.21*	**−0.77** *0.21*	**−0.20** *0.16*												
Finnish	**−0.36** *0.17*	**−0.57** *0.22*	**−0.56** *0.18*	**0.01** *0.18*	**0.22** *0.16*											
French	**−0.03** *0.18*	**−0.24** *0.22*	**−0.23** *0.22*	**0.34** *0.15*	**0.55** *0.16*	**0.33** *0.18*										
German	**−0.66** *0.16*	**−0.86** *0.20*	**−0.85** *0.18*	**−0.28** *0.16*	**−0.08** *0.16*	**−0.30** *0.16*	**−0.63** *0.16*									
Italian	**0.11** *0.21*	**−0.10** *0.24*	**−0.09** *0.23*	**0.48** *0.19*	**0.69** *0.19*	**0.47** *0.20*	**0.14** *0.19*	**0.77** *0.18*								
Japanese	**−0.60** *0.16*	**−0.80** *0.19*	**−0.79** *0.15*	**−0.22** *0.17*	**−0.02** *0.17*	**−0.24** *0.17*	**−0.56** *0.17*	**0.06** *0.14*	**−0.71** *0.19*							
Norwegian	**−0.25** *0.16*	**−0.45** *0.21*	**−0.44** *0.17*	**0.13** *0.18*	**0.33** *0.16*	**0.11** *0.16*	**−0.22** *0.19*	**0.41** *0.16*	**−0.36** *0.21*	**0.35** *0.16*						
Polish	**0.00** *0.17*	**−0.20** *0.21*	**−0.19** *0.21*	**0.38** *0.15*	**0.58** *0.15*	**0.36** *0.17*	**0.04** *0.14*	**0.66** *0.15*	**−0.11** *0.18*	**0.60** *0.16*	**0.25** *0.18*					
Russian	**−0.41** *0.17*	**−0.61** *0.22*	**−0.60** *0.21*	**−0.03** *0.16*	**0.17** *0.16*	**−0.04** *0.18*	**−0.37** *0.16*	**0.25** *0.16*	**−0.51** *0.20*	**0.19** *0.16*	**−0.16** *0.18*	**−0.41** *0.16*				
Spanish	**−0.24** *0.18*	**−0.44** *0.22*	**−0.43** *0.21*	**0.14** *0.18*	**0.34** *0.17*	**0.13** *0.18*	**−0.20** *0.17*	**0.42** *0.17*	**−0.34** *0.20*	**0.36** *0.17*	**0.01** *0.19*	**−0.24** *0.17*	**0.17** *0.18*			
Swedish	**−0.23** *0.16*	**−0.43** *0.21*	**−0.42** *0.16*	**0.15** *0.18*	**0.35** *0.16*	**0.13** *0.15*	**−0.20** *0.17*	**0.43** *0.14*	**−0.34** *0.19*	**0.37** *0.15*	**0.02** *0.15*	**−0.23** *0.16*	**0.18** *0.18*	**0.01** *0.18*		
Tswana	**−0.06** *0.15*	**−0.27** *0.21*	**−0.26** *0.16*	**0.31** *0.19*	**0.52** *0.19*	**0.30** *0.17*	**−0.03** *0.20*	**0.60** *0.16*	**−0.17** *0.21*	**0.54** *0.15*	**0.19** *0.15*	**−0.07** *0.19*	**0.34** *0.19*	**0.17** *0.20*	**0.17** *0.16*	
Turkish	**0.11** *0.16*	**−0.09** *0.21*	**−0.08** *0.21*	**0.49** *0.16*	**0.69** *0.16*	**0.47** *0.17*	**0.14** *0.16*	**0.77** *0.16*	**0.00** *0.19*	**0.71** *0.16*	**0.36** *0.18*	**0.11** *0.16*	**0.51** *0.16*	**0.35** *0.18*	**0.34** *0.17*	**0.17** *0.17*

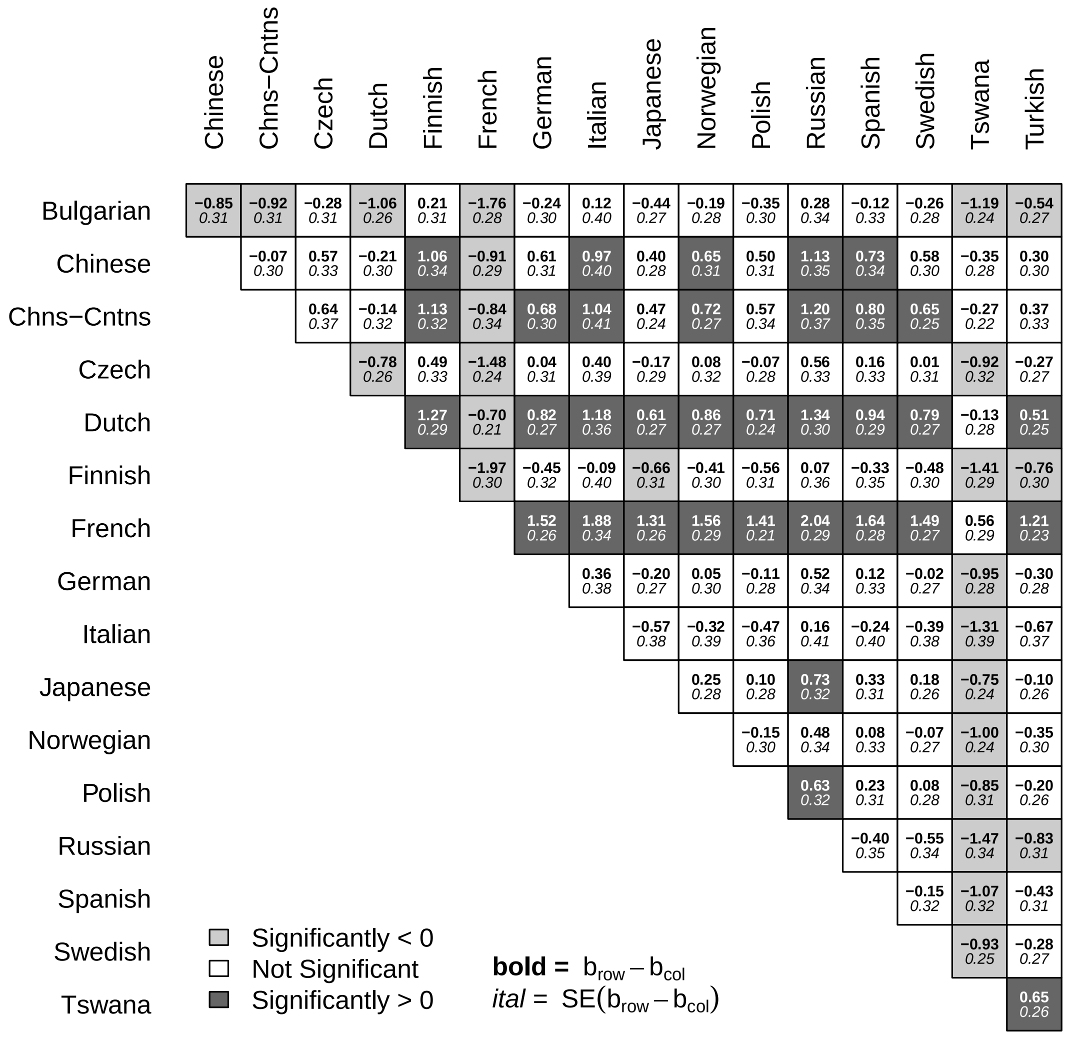
▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 44: Pairwise Differences in L1 Parameter Values for Past Simple Passive Neg. Binomial Type I Regression.

7.5.6 The Future Passive

Out of 5 possible Tense and Aspect combinations the pairwise differences in L1 parameter values for 3 combinations could have been plotted. To assure readability the plots of the combined constructions have been omitted.



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**−0.85** *0.31*	**−0.92** *0.31*	**−0.28** *0.31*	**−1.06** *0.26*	**0.21** *0.31*	**−1.76** *0.28*	**−0.24** *0.30*	**0.12** *0.40*	**−0.44** *0.27*	**−0.19** *0.28*	**−0.35** *0.30*	**0.28** *0.34*	**−0.12** *0.33*	**−0.26** *0.28*	**−1.19** *0.24*	**−0.54** *0.27*
Chinese		**−0.07** *0.30*	**0.57** *0.33*	**−0.21** *0.30*	**1.06** *0.34*	**−0.91** *0.29*	**0.61** *0.31*	**0.97** *0.40*	**0.40** *0.28*	**0.65** *0.31*	**0.50** *0.31*	**1.13** *0.35*	**0.73** *0.34*	**0.58** *0.30*	**−0.35** *0.28*	**0.30** *0.30*
Chns−Cntns			**0.64** *0.37*	**−0.14** *0.32*	**1.13** *0.32*	**−0.84** *0.34*	**0.68** *0.30*	**1.04** *0.41*	**0.47** *0.24*	**0.72** *0.27*	**0.57** *0.34*	**1.20** *0.37*	**0.80** *0.35*	**0.65** *0.25*	**−0.27** *0.22*	**0.37** *0.33*
Czech				**−0.78** *0.26*	**0.49** *0.33*	**−1.48** *0.24*	**0.04** *0.31*	**0.40** *0.39*	**−0.17** *0.29*	**0.08** *0.32*	**−0.07** *0.28*	**0.56** *0.33*	**0.16** *0.33*	**0.01** *0.31*	**−0.92** *0.32*	**−0.27** *0.27*
Dutch					**1.27** *0.29*	**−0.70** *0.21*	**0.82** *0.27*	**1.18** *0.36*	**0.61** *0.27*	**0.86** *0.27*	**0.71** *0.24*	**1.34** *0.30*	**0.94** *0.29*	**0.79** *0.27*	**−0.13** *0.28*	**0.51** *0.25*
Finnish						**−1.97** *0.30*	**−0.45** *0.32*	**−0.09** *0.40*	**−0.66** *0.31*	**−0.41** *0.30*	**−0.56** *0.31*	**0.07** *0.36*	**−0.33** *0.35*	**−0.48** *0.30*	**−1.41** *0.29*	**−0.76** *0.30*
French							**1.52** *0.26*	**1.88** *0.34*	**1.31** *0.26*	**1.56** *0.29*	**1.41** *0.21*	**2.04** *0.29*	**1.64** *0.28*	**1.49** *0.27*	**0.56** *0.29*	**1.21** *0.23*
German								**0.36** *0.38*	**−0.20** *0.27*	**0.05** *0.30*	**−0.11** *0.28*	**0.52** *0.34*	**0.12** *0.33*	**−0.02** *0.27*	**−0.95** *0.28*	**−0.30** *0.28*
Italian									**−0.57** *0.38*	**−0.32** *0.39*	**−0.47** *0.36*	**0.16** *0.41*	**−0.24** *0.40*	**−0.39** *0.38*	**−1.31** *0.39*	**−0.67** *0.37*
Japanese										**0.25** *0.28*	**0.10** *0.28*	**0.73** *0.32*	**0.33** *0.31*	**0.18** *0.26*	**−0.75** *0.24*	**−0.10** *0.26*
Norwegian											**−0.15** *0.30*	**0.48** *0.34*	**0.08** *0.33*	**−0.07** *0.27*	**−1.00** *0.24*	**−0.35** *0.30*
Polish												**0.63** *0.32*	**0.23** *0.31*	**0.08** *0.28*	**−0.85** *0.31*	**−0.20** *0.26*
Russian													**−0.40** *0.35*	**−0.55** *0.34*	**−1.47** *0.34*	**−0.83** *0.31*
Spanish														**−0.15** *0.32*	**−1.07** *0.32*	**−0.43** *0.31*
Swedish															**−0.93** *0.25*	**−0.28** *0.27*
Tswana																**0.65** *0.26*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0
bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 45: Pairwise Differences in L1 Parameter Values for Will Future Simple Passive Neg. Binomial Type I Regression.

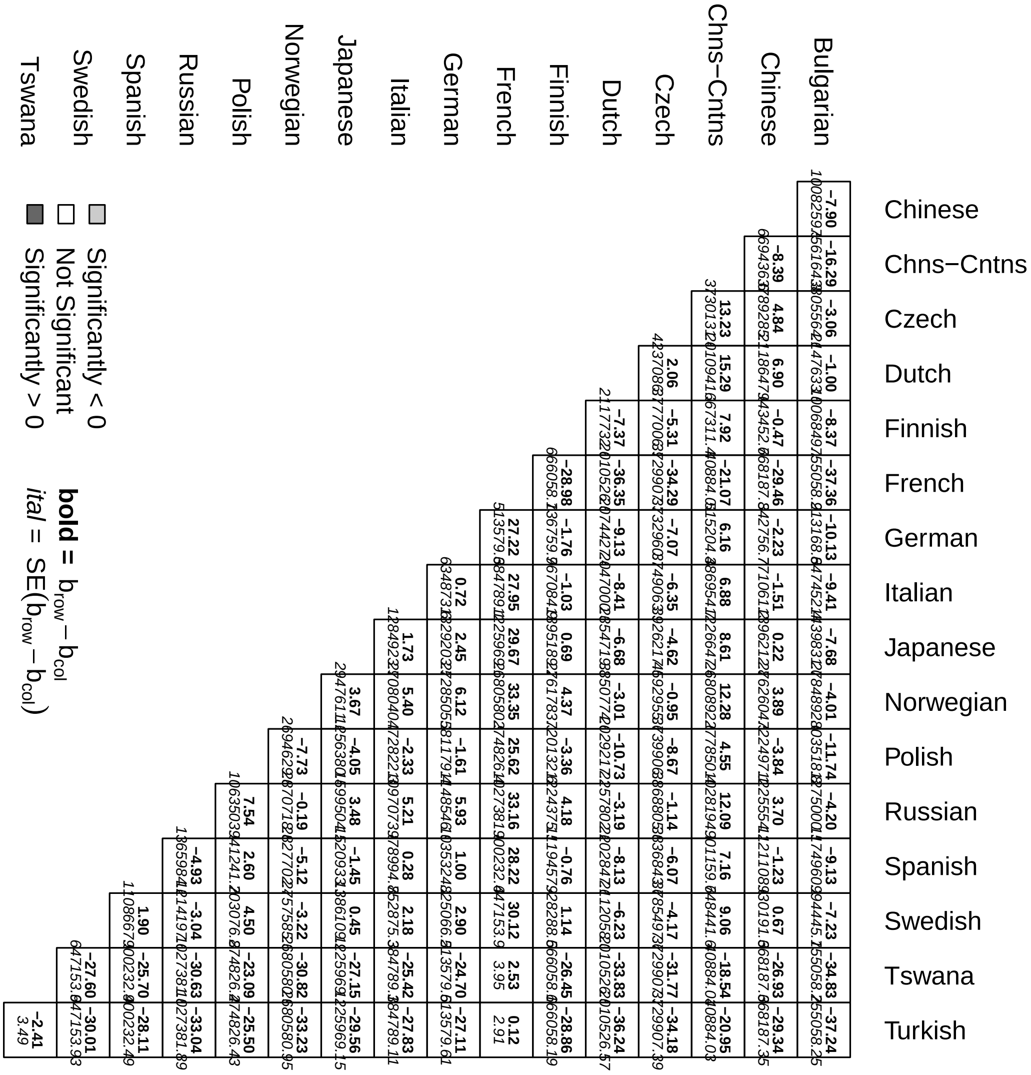
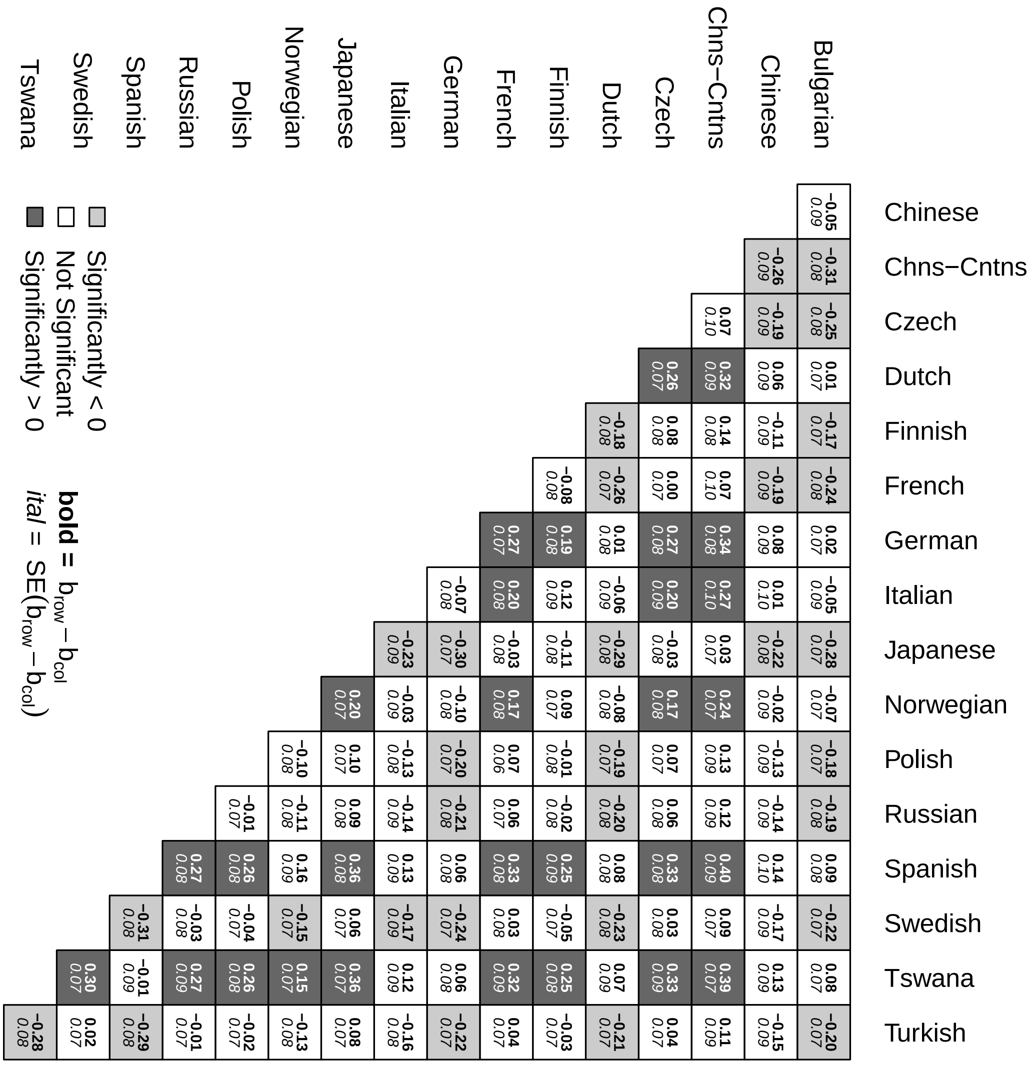


Figure 46: Pairwise Differences in L1 Parameter Values for Will Future Perfect Passive Neg. Binomial Type I Regression.

7.5.7 The Conditional Active

Out of 5 possible Tense and Aspect combinations the pairwise differences in L1 parameter values for 5 combinations could have been plotted. To assure readability the plots of the combined constructions have been omitted.

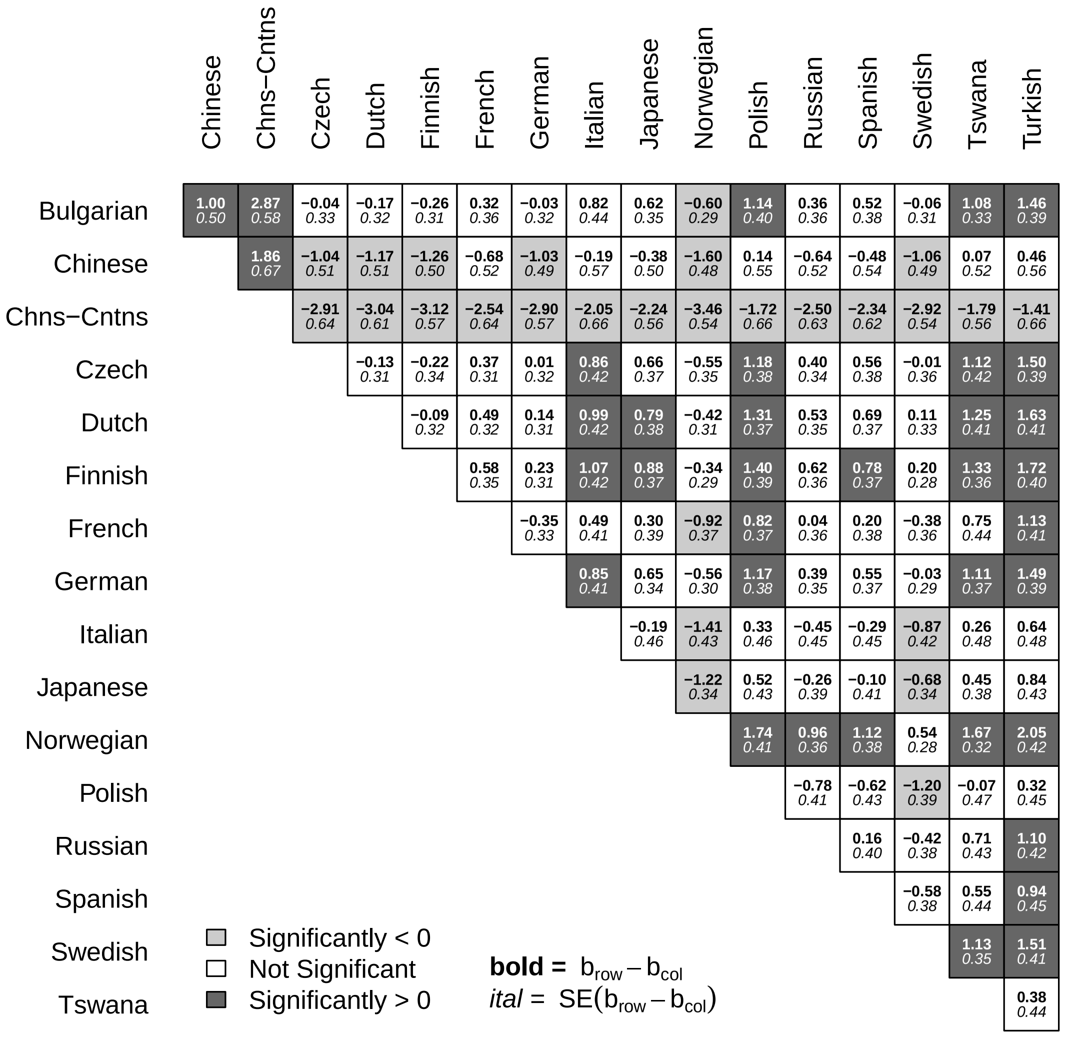


	Bulgarian	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana
Chinese	**−0.05** *0.09*															
Chns−Cntns	**−0.31** *0.08*	**−0.26** *0.09*														
Czech	**−0.25** *0.08*	**−0.19** *0.09*	**0.07** *0.10*													
Dutch	**0.01** *0.07*	**0.06** *0.09*	**0.32** *0.09*	**0.26** *0.07*												
Finnish	**−0.17** *0.07*	**−0.11** *0.09*	**0.14** *0.08*	**0.08** *0.08*	**−0.18** *0.08*											
French	**−0.24** *0.08*	**−0.19** *0.09*	**0.07** *0.10*	**0.00** *0.07*	**−0.26** *0.07*	**−0.08** *0.08*										
German	**0.02** *0.07*	**0.08** *0.09*	**0.34** *0.08*	**0.27** *0.08*	**0.01** *0.08*	**0.19** *0.08*	**0.27** *0.07*									
Italian	**−0.05** *0.09*	**0.01** *0.10*	**0.27** *0.10*	**0.20** *0.09*	**−0.06** *0.09*	**0.12** *0.09*	**0.20** *0.08*	**−0.07** *0.08*								
Japanese	**−0.28** *0.07*	**−0.22** *0.08*	**0.03** *0.07*	**−0.03** *0.08*	**−0.29** *0.08*	**−0.11** *0.08*	**−0.03** *0.08*	**−0.30** *0.07*	**−0.23** *0.09*							
Norwegian	**−0.07** *0.07*	**−0.02** *0.09*	**0.24** *0.07*	**0.17** *0.08*	**−0.08** *0.08*	**0.09** *0.07*	**0.17** *0.08*	**−0.10** *0.08*	**−0.03** *0.09*	**0.20** *0.07*						
Polish	**−0.18** *0.07*	**−0.13** *0.09*	**0.13** *0.09*	**0.07** *0.07*	**−0.19** *0.07*	**−0.01** *0.08*	**0.07** *0.06*	**−0.20** *0.07*	**−0.13** *0.08*	**0.10** *0.07*	**−0.10** *0.08*					
Russian	**−0.19** *0.08*	**−0.14** *0.09*	**0.12** *0.09*	**0.06** *0.08*	**−0.20** *0.08*	**−0.02** *0.08*	**0.06** *0.07*	**−0.21** *0.08*	**−0.14** *0.09*	**0.09** *0.08*	**−0.11** *0.08*	**−0.01** *0.07*				
Spanish	**0.09** *0.08*	**0.14** *0.10*	**0.40** *0.09*	**0.33** *0.08*	**0.08** *0.08*	**0.25** *0.09*	**0.33** *0.08*	**0.06** *0.08*	**0.13** *0.09*	**0.36** *0.08*	**0.16** *0.09*	**0.26** *0.08*	**0.27** *0.08*			
Swedish	**−0.22** *0.07*	**−0.17** *0.09*	**0.09** *0.07*	**0.03** *0.08*	**−0.23** *0.08*	**−0.05** *0.07*	**0.03** *0.08*	**−0.24** *0.07*	**−0.17** *0.09*	**0.06** *0.07*	**−0.15** *0.07*	**−0.04** *0.07*	**−0.03** *0.08*	**−0.31** *0.08*		
Tswana	**0.08** *0.07*	**0.13** *0.09*	**0.39** *0.07*	**0.33** *0.09*	**0.07** *0.09*	**0.25** *0.08*	**0.32** *0.09*	**0.06** *0.08*	**0.12** *0.09*	**0.36** *0.07*	**0.15** *0.07*	**0.26** *0.08*	**0.27** *0.09*	**−0.01** *0.09*	**0.30** *0.07*	
Turkish	**−0.20** *0.07*	**−0.15** *0.09*	**0.11** *0.09*	**0.04** *0.07*	**−0.21** *0.07*	**−0.03** *0.07*	**0.04** *0.07*	**−0.22** *0.07*	**−0.16** *0.08*	**0.08** *0.07*	**−0.13** *0.08*	**−0.02** *0.07*	**−0.01** *0.07*	**−0.29** *0.08*	**0.02** *0.07*	**−0.28** *0.08*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 47: Pairwise Differences in L1 Parameter Values for Conditional Simple Neg. Binomial Type I Regression.



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**1.00** *0.50*	**2.87** *0.58*	**−0.04** *0.33*	**−0.17** *0.32*	**−0.26** *0.31*	**0.32** *0.36*	**−0.03** *0.32*	**0.82** *0.44*	**0.62** *0.35*	**−0.60** *0.29*	**1.14** *0.40*	**0.36** *0.36*	**0.52** *0.38*	**−0.06** *0.31*	**1.08** *0.33*	**1.46** *0.39*
Chinese		**1.86** *0.67*	**−1.04** *0.51*	**−1.17** *0.51*	**−1.26** *0.50*	**−0.68** *0.52*	**−1.03** *0.49*	**−0.19** *0.57*	**−0.38** *0.50*	**−1.60** *0.48*	**0.14** *0.55*	**−0.64** *0.52*	**−0.48** *0.54*	**−1.06** *0.49*	**0.07** *0.52*	**0.46** *0.56*
Chns−Cntns			**−2.91** *0.64*	**−3.04** *0.61*	**−3.12** *0.57*	**−2.54** *0.64*	**−2.90** *0.57*	**−2.05** *0.66*	**−2.24** *0.56*	**−3.46** *0.54*	**−1.72** *0.66*	**−2.50** *0.63*	**−2.34** *0.62*	**−2.92** *0.54*	**−1.79** *0.56*	**−1.41** *0.66*
Czech				**−0.13** *0.31*	**−0.22** *0.34*	**0.37** *0.31*	**0.01** *0.32*	**0.86** *0.42*	**0.66** *0.37*	**−0.55** *0.35*	**1.18** *0.38*	**0.40** *0.34*	**0.56** *0.38*	**−0.01** *0.36*	**1.12** *0.42*	**1.50** *0.39*
Dutch					**−0.09** *0.32*	**0.49** *0.32*	**0.14** *0.31*	**0.99** *0.42*	**0.79** *0.38*	**−0.42** *0.31*	**1.31** *0.37*	**0.53** *0.35*	**0.69** *0.37*	**0.11** *0.33*	**1.25** *0.41*	**1.63** *0.41*
Finnish						**0.58** *0.35*	**0.23** *0.31*	**1.07** *0.42*	**0.88** *0.37*	**−0.34** *0.29*	**1.40** *0.39*	**0.62** *0.36*	**0.78** *0.37*	**0.20** *0.28*	**1.33** *0.36*	**1.72** *0.40*
French							**−0.35** *0.33*	**0.49** *0.41*	**0.30** *0.39*	**−0.92** *0.37*	**0.82** *0.37*	**0.04** *0.36*	**0.20** *0.38*	**−0.38** *0.36*	**0.75** *0.44*	**1.13** *0.41*
German								**0.85** *0.41*	**0.65** *0.34*	**−0.56** *0.30*	**1.17** *0.38*	**0.39** *0.35*	**0.55** *0.37*	**−0.03** *0.29*	**1.11** *0.37*	**1.49** *0.39*
Italian									**−0.19** *0.46*	**−1.41** *0.43*	**0.33** *0.46*	**−0.45** *0.45*	**−0.29** *0.45*	**−0.87** *0.42*	**0.26** *0.48*	**0.64** *0.48*
Japanese										**−1.22** *0.34*	**0.52** *0.43*	**−0.26** *0.39*	**−0.10** *0.41*	**−0.68** *0.34*	**0.45** *0.38*	**0.84** *0.43*
Norwegian											**1.74** *0.41*	**0.96** *0.36*	**1.12** *0.38*	**0.54** *0.28*	**1.67** *0.32*	**2.05** *0.42*
Polish												**−0.78** *0.41*	**−0.62** *0.43*	**−1.20** *0.39*	**−0.07** *0.47*	**0.32** *0.45*
Russian													**0.16** *0.40*	**−0.42** *0.38*	**0.71** *0.43*	**1.10** *0.42*
Spanish														**−0.58** *0.38*	**0.55** *0.44*	**0.94** *0.45*
Swedish															**1.13** *0.35*	**1.51** *0.41*
Tswana																**0.38** *0.44*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 48: Pairwise Differences in L1 Parameter Values for Conditional Perfect Neg. Binomial Type I Regression.

	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**30.51** *93515.8*	**−0.28** *0.54*	**0.33** *0.51*	**0.41** *0.46*	**−0.55** *0.41*	**1.53** *0.71*	**0.48** *0.49*	**1.99** *1.09*	**0.15** *0.46*	**0.56** *0.45*	**0.15** *0.48*	**0.43** *0.53*	**0.09** *0.52*	**0.16** *0.45*	**0.29** *0.41*	**0.41** *0.47*
Chinese		**−30.79** *93515.8*	**−30.18** *93515.8*	**−30.10** *93515.8*	**−31.05** *93515.8*	**−28.98** *93515.8*	**−30.03** *93515.8*	**−28.51** *93515.8*	**−30.36** *93515.8*	**−29.95** *93515.8*	**−30.35** *93515.8*	**−30.08** *93515.8*	**−30.42** *93515.8*	**−30.35** *93515.8*	**−30.22** *93515.8*	**−30.09** *93515.82*
Chns−Cntns			**0.61** *0.67*	**0.69** *0.61*	**−0.26** *0.49*	**1.81** *0.83*	**0.76** *0.56*	**2.28** *1.13*	**0.43** *0.46*	**0.84** *0.48*	**0.44** *0.62*	**0.71** *0.65*	**0.37** *0.60*	**0.44** *0.45*	**0.57** *0.44*	**0.69** *0.63*
Czech				**0.08** *0.49*	**−0.88** *0.50*	**1.20** *0.69*	**0.15** *0.53*	**1.66** *1.09*	**−0.18** *0.53*	**0.23** *0.56*	**−0.18** *0.48*	**0.10** *0.54*	**−0.24** *0.54*	**−0.17** *0.54*	**−0.04** *0.59*	**0.08** *0.50*
Dutch					**−0.96** *0.44*	**1.12** *0.69*	**0.07** *0.50*	**1.58** *1.08*	**−0.26** *0.52*	**0.15** *0.49*	**−0.26** *0.46*	**0.02** *0.52*	**−0.32** *0.51*	**−0.25** *0.48*	**−0.12** *0.55*	**0.00** *0.51*
Finnish						**2.07** *0.69*	**1.03** *0.46*	**2.54** *1.06*	**0.70** *0.46*	**1.11** *0.41*	**0.70** *0.45*	**0.98** *0.51*	**0.64** *0.49*	**0.71** *0.38*	**0.84** *0.44*	**0.96** *0.46*
French							**−1.05** *0.71*	**0.46** *1.18*	**−1.38** *0.72*	**−0.97** *0.74*	**−1.38** *0.66*	**−1.10** *0.72*	**−1.44** *0.72*	**−1.37** *0.72*	**−1.24** *0.78*	**−1.12** *0.70*
German								**1.51** *1.08*	**−0.33** *0.49*	**0.08** *0.50*	**−0.33** *0.48*	**−0.05** *0.55*	**−0.39** *0.53*	**−0.32** *0.47*	**−0.19** *0.52*	**−0.07** *0.51*
Italian									**−1.84** *1.08*	**−1.43** *1.09*	**−1.84** *1.06*	**−1.56** *1.10*	**−1.90** *1.08*	**−1.83** *1.07*	**−1.70** *1.11*	**−1.58** *1.08*
Japanese										**0.41** *0.48*	**0.00** *0.50*	**0.28** *0.54*	**−0.06** *0.53*	**0.01** *0.46*	**0.14** *0.46*	**0.26** *0.50*
Norwegian											**−0.41** *0.53*	**−0.13** *0.56*	**−0.47** *0.54*	**−0.40** *0.44*	**−0.27** *0.43*	**−0.15** *0.55*
Polish												**0.28** *0.51*	**−0.06** *0.50*	**0.01** *0.48*	**0.14** *0.57*	**0.26** *0.48*
Russian													**−0.34** *0.56*	**−0.27** *0.55*	**−0.14** *0.59*	**−0.02** *0.53*
Spanish														**0.07** *0.52*	**0.20** *0.57*	**0.32** *0.54*
Swedish															**0.13** *0.45*	**0.25** *0.50*
Tswana																**0.12** *0.52*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 49: Pairwise Differences in L1 Parameter Values for Conditional Progressive Neg. Binomial Type I Regression.

	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**2.14** *18003.4*	**−19.05** *9673.50*	**−15.93** *9673.50*	**−17.41** *9673.50*	**1.55** *15719.7*	**4.63** *13851.6*	**1.94** *14368.8*	**2.54** *17467.1*	**2.29** *13559.7*	**−18.58** *9673.50*	**−16.97** *9673.50*	**3.33** *16852.0*	**3.78** *15617.6*	**−19.18** *9673.50*	**−19.55** *9673.50*	**2.82** *14203.18*
Chinese		**−21.19** *15183.8*	**−18.07** *15183.8*	**−19.55** *15183.8*	**−0.59** *19598.0*	**2.49** *18133.9*	**−0.20** *18532.0*	**0.40** *21025.5*	**0.15** *17911.9*	**−20.72** *15183.8*	**−19.11** *15183.8*	**1.19** *20517.4*	**1.64** *19516.2*	**−21.32** *15183.8*	**−21.69** *15183.8*	**0.68** *18403.87*
Chns−Cntns			**3.12** *3.03*	**1.64** *2.66*	**20.60** *12390.9*	**23.68** *9914.26*	**20.99** *10624.8*	**21.59** *14543.8*	**21.34** *9502.06*	**0.46** *2.13*	**2.08** *2.77*	**22.38** *13799.1*	**22.83** *12261.1*	**−0.13** *2.04*	**−0.50** *2.18*	**21.86** *10399.69*
Czech				**−1.48** *1.67*	**17.48** *12390.9*	**20.56** *9914.26*	**17.87** *10624.8*	**18.47** *14543.8*	**18.23** *9502.06*	**−2.65** *2.00*	**−1.04** *1.63*	**19.26** *13799.1*	**19.71** *12261.1*	**−3.24** *1.92*	**−3.62** *2.72*	**18.75** *10399.69*
Dutch					**18.96** *12390.9*	**22.04** *9914.26*	**19.35** *10624.8*	**19.95** *14543.8*	**19.70** *9502.06*	**−1.17** *1.71*	**0.44** *1.63*	**20.74** *13799.1*	**21.19** *12261.1*	**−1.76** *1.66*	**−2.14** *2.47*	**20.23** *10399.69*
Finnish						**3.08** *15869.0*	**0.39** *16322.4*	**1.00** *19106.4*	**0.75** *15614.8*	**−20.13** *12390.9*	**−18.52** *12390.9*	**1.78** *18545.9*	**2.24** *17431.8*	**−20.72** *12390.9*	**−21.10** *12390.9*	**1.27** *16176.78*
French							**−2.69** *14532.0*	**−2.09** *17601.5*	**−2.34** *13732.5*	**−23.22** *9914.26*	**−21.60** *9914.26*	**−1.30** *16991.4*	**−0.85** *15767.9*	**−23.81** *9914.26*	**−24.18** *9914.26*	**−1.82** *14368.24*
German								**0.60** *18011.3*	**0.36** *14253.9*	**−20.52** *10624.8*	**−18.91** *10624.8*	**1.39** *17415.5*	**1.84** *16224.1*	**−21.11** *10624.8*	**−21.49** *10624.8*	**0.88** *14867.43*
Italian									**−0.25** *17372.7*	**−21.13** *14543.8*	**−19.51** *14543.8*	**0.79** *20048.4*	**1.24** *19022.5*	**−21.72** *14543.8*	**−22.10** *14543.8*	**0.27** *17879.51*
Japanese										**−20.88** *9502.06*	**−19.26** *9502.06*	**1.04** *16754.2*	**1.49** *15512.0*	**−21.47** *9502.06*	**−21.85** *9502.06*	**0.52** *14086.97*
Norwegian											**1.62** *2.00*	**21.91** *13799.1*	**22.37** *12261.1*	**−0.59** *1.29*	**−0.97** *1.96*	**21.40** *10399.69*
Polish												**20.30** *13799.1*	**20.75** *12261.1*	**−2.21** *1.75*	**−2.58** *2.68*	**19.78** *10399.69*
Russian													**0.45** *18459.4*	**−22.50** *13799.1*	**−22.88** *13799.1*	**−0.51** *17279.16*
Spanish														**−22.96** *12261.1*	**−23.34** *12261.1*	**−0.97** *16077.60*
Swedish															**−0.38** *1.96*	**21.99** *10399.69*
Tswana																**22.37** *10399.69*

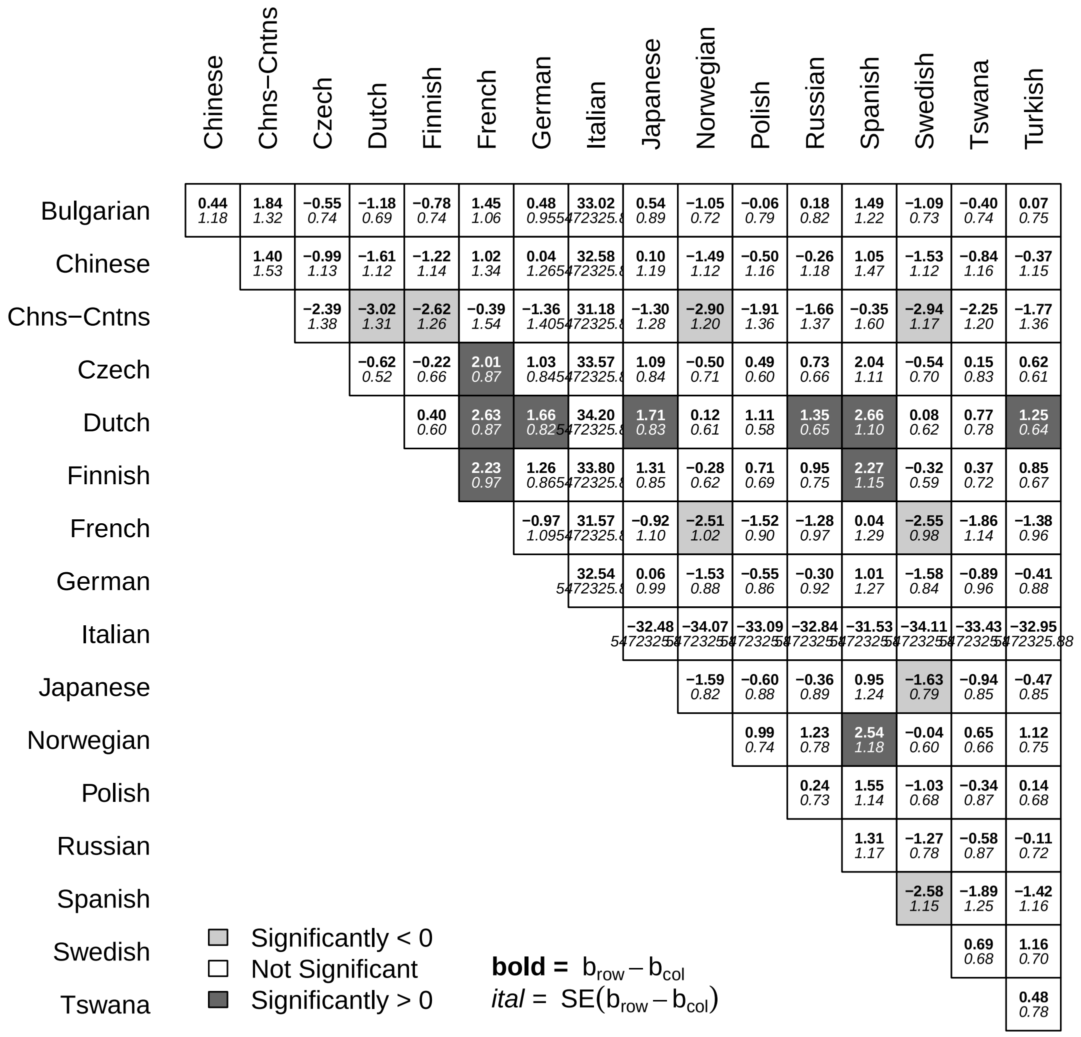
▨ Significantly < 0
□ Not Significant
■ Significantly > 0

bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 50: Pairwise Differences in L1 Parameter Values for Conditional Perfect Progressive Neg. Binomial Type I Regression.

7.5.8 The Conditional Passive

Out of 5 possible Tense and Aspect combinations the pairwise differences in L1 parameter values for 1 combinations could have been plotted. To assure readability the plots of the combined constructions have been omitted.



	Chinese	Chns−Cntns	Czech	Dutch	Finnish	French	German	Italian	Japanese	Norwegian	Polish	Russian	Spanish	Swedish	Tswana	Turkish
Bulgarian	**0.44** *1.18*	**1.84** *1.32*	**−0.55** *0.74*	**−1.18** *0.69*	**−0.78** *0.74*	**1.45** *1.06*	**0.48** *0.95*	**33.02** *5472325.8*	**0.54** *0.89*	**−1.05** *0.72*	**−0.06** *0.79*	**0.18** *0.82*	**1.49** *1.22*	**−1.09** *0.73*	**−0.40** *0.74*	**0.07** *0.75*
Chinese		**1.40** *1.53*	**−0.99** *1.13*	**−1.61** *1.12*	**−1.22** *1.14*	**1.02** *1.34*	**0.04** *1.26*	**32.58** *5472325.8*	**0.10** *1.19*	**−1.49** *1.12*	**−0.50** *1.16*	**−0.26** *1.18*	**1.05** *1.47*	**−1.53** *1.12*	**−0.84** *1.16*	**−0.37** *1.15*
Chns−Cntns			**−2.39** *1.38*	**−3.02** *1.31*	**−2.62** *1.26*	**−0.39** *1.54*	**−1.36** *1.40*	**31.18** *5472325.8*	**−1.30** *1.28*	**−2.90** *1.20*	**−1.91** *1.36*	**−1.66** *1.37*	**−0.35** *1.60*	**−2.94** *1.17*	**−2.25** *1.20*	**−1.77** *1.36*
Czech				**−0.62** *0.52*	**−0.22** *0.66*	**2.01** *0.87*	**1.03** *0.84*	**33.57** *5472325.8*	**1.09** *0.84*	**−0.50** *0.71*	**0.49** *0.60*	**0.73** *0.66*	**2.04** *1.11*	**−0.54** *0.70*	**0.15** *0.83*	**0.62** *0.61*
Dutch					**0.40** *0.60*	**2.63** *0.87*	**1.66** *0.82*	**34.20** *5472325.8*	**1.71** *0.83*	**0.12** *0.61*	**1.11** *0.58*	**1.35** *0.65*	**2.66** *1.10*	**0.08** *0.62*	**0.77** *0.78*	**1.25** *0.64*
Finnish						**2.23** *0.97*	**1.26** *0.86*	**33.80** *5472325.8*	**1.31** *0.85*	**−0.28** *0.62*	**0.71** *0.69*	**0.95** *0.75*	**2.27** *1.15*	**−0.32** *0.59*	**0.37** *0.72*	**0.85** *0.67*
French							**−0.97** *1.09*	**31.57** *5472325.8*	**−0.92** *1.10*	**−2.51** *1.02*	**−1.52** *0.90*	**−1.28** *0.97*	**0.04** *1.29*	**−2.55** *0.98*	**−1.86** *1.14*	**−1.38** *0.96*
German								**32.54** *5472325.8*	**0.06** *0.99*	**−1.53** *0.88*	**−0.55** *0.86*	**−0.30** *0.92*	**1.01** *1.27*	**−1.58** *0.84*	**−0.89** *0.96*	**−0.41** *0.88*
Italian									**−32.48** *5472325*	**−34.07** *5472325*	**−33.09** *5472325*	**−32.84** *5472325*	**−31.53** *5472325*	**−34.11** *5472325*	**−33.43** *5472325*	**−32.95** *5472325.88*
Japanese										**−1.59** *0.82*	**−0.60** *0.88*	**−0.36** *0.89*	**0.95** *1.24*	**−1.63** *0.79*	**−0.94** *0.85*	**−0.47** *0.85*
Norwegian											**0.99** *0.74*	**1.23** *0.78*	**2.54** *1.18*	**−0.04** *0.60*	**0.65** *0.66*	**1.12** *0.75*
Polish												**0.24** *0.73*	**1.55** *1.14*	**−1.03** *0.68*	**−0.34** *0.87*	**0.14** *0.68*
Russian													**1.31** *1.17*	**−1.27** *0.78*	**−0.58** *0.87*	**−0.11** *0.72*
Spanish														**−2.58** *1.15*	**−1.89** *1.25*	**−1.42** *1.16*
Swedish															**0.69** *0.68*	**1.16** *0.70*
Tswana																**0.48** *0.78*

▨ Significantly < 0
□ Not Significant
■ Significantly > 0
bold = $b_{row} - b_{col}$
ital = $SE(b_{row} - b_{col})$

Figure 51: Pairwise Differences in L1 Parameter Values for Conditional Perfect Passive Neg. Binomial Type I Regression.

7.6 Source Codes

7.6.1 Empirical Distributions

Source Code 48: Empirical Distributions and their respective plots

```
x=seq(-10,10,0.1)
xint=seq(-10,10,1)

norm=(1/sqrt(2*pi))*exp(-(((x-5)^2)/2))
log.norm=((((log(exp(1)))/(x*sqrt(2*pi)))*exp(((log(x-1)^2)/(4))
   )))
exponential=exp(-5*x)
cauchy=(1/pi)*(1/(((x-5)^2)+(1)^2))
poisson=exp(-5)*((5^as.integer(x))/factorial(as.integer(x)))
neg.bin=(choose(as.integer(x)+5-1,as.integer(x)))*0.5^5*(1-0.5)
   ^(as.integer(x))

plot(x,norm,type="l",ylab="p",xlim=c(0,10),ylim=c(0,0.4))
lines(x,cauchy,lty=2)
lines(x,log.norm,lty=4)
legend("topright",c("Normal","Log.␣Normal","Cauchy"),lty=c
   (1,4,2))

plot(x,exponential,type="l",lty=1,ylab="p",xlim=c(0,10),ylim=c
   (0,0.4))
lines(x,poisson,lty=2)
lines(x,neg.bin,lty=4)
legend("topright",c("Exponential","Poisson","Neg.␣Binomial"),lty
   =c(1,2,4))
```

7.6.2 Descriptive Statistics of Tense and Aspects

Source Code 49: Descriptive Statistics of Tense and Aspects

```
results.table=function(x){
  results=data.frame("Tense/Aspect"=rep(0,57),"sum"=rep(0,57),"min"=rep(0,5
  k=1:57
  results[,1]=names(x)
  for(i in k){
    results[i,2]=sum(x[,i])
  }
  for(i in k){
    results[i,3]=min(x[,i])
  }
  for(i in k){
    results[i,4]=max(x[,i])
  }
  for(i in k){
    results[i,5]=median(x[,i])
  }
  for(i in k){
    results[i,6]=mean(x[,i])
  }
  for(i in k){
    results[i,7]=var(x[,i])
  }
  for(i in k){
    results[i,8]=sd(x[,i])
  }

  for(i in k){
    results[i,9]=skewness(x[,i])
  }
  for(i in k){
    results[i,10]=kurtosis(x[,i])
  }
  return(results)
}
```

7.6.3 Tests for Empirical Distributions of Tense and Aspect

Source Code 50: Tests for Empirical Distributions of Tense and Aspect

```
results.distr=function(x){
  results=data.frame(”Tense/Aspect”=rep(0,57),”Normal”=rep(0,57),”Cauchy”=r
  k=1:57
  results[,1]=names(x)
  for(i in k){
    temp=ks.test(x[,i],”pnorm”)
    results[i,2]=temp$statistic
```

```
}
for(i in k){
  temp=ks.test(x[,i],”pcauchy”)
  results[i,3]=temp$statistic
}
for(i in k){
  temp=ks.test(x[,i],”pexp”)
  results[i,4]=temp$statistic
}
for(i in k){
  temp=ks.test(x[,i],”plnorm”)
  results[i,5]=temp$statistic
}
for(i in k){
  temp=fitdist(x[,i],”pois”,method=”mme”)
  temp1=ks.test(x[,i],”ppois”,temp$estimate)
  results[i,6]=temp1$statistic
}
for(i in k){
  #test1=fitdist(x[,i],”nbinom”,method=”mle”)
  #probp=x[,i]/x[,2]
  xtemp=data.frame(”TA”=rep(0,4424))
  xtemp[,1]=x[,i]
  y=subset(xtemp,TA>0,select=TA)
  if(sum(y[,1])>0){
  neg.binI=function(size,prob) {
    R=dnbinom(y[,1],size,prob)
    -sum(log(R))
  }
  try1=mle(neg.binI,start=list(prob=0.1,size=0.1))
  temp=ks.test(y[,1],”pnbinom”,prob=try1@coef[2],size=try1@coef[1])
  results[i,7]=temp$statistic}else{results[i,7]=0}
    }
for(i in k){
  results[i,8]=names(which.min(results[i,2:7]))
}
for(i in k){
  results[i,9]=if(max(x[,i])<=5){”checkmark”}else{”X”}
  }

  return(results)
}
```

7.6.4 Model Selection with Adjusted R^2 and AIC Values

Source Code 51: Model Selection with Adjusted R^2 and AIC Values

```
model.selection=function(input.table.red){
  library(gamlss)
  results=data.frame("Tense␣and␣Aspect"=rep(0,103),"Adj.␣R-Sq"=
      rep(0,103),"Linear␣Model"=rep(0,103),"Log.␣Normal"=rep
      (0,103),"Poisson"=rep(0,103),"Neg.␣Binomial␣I"=rep(0,103),"
      Neg.␣Binomial␣II"=rep(0,103),"Best␣Model"=rep(0,103))
  results[,1]=names(input.table.red)
  k=c(47:81,83:84,88:90,93:96,98,99,100,103)
  #linear model
  for (i in k){
    temp=lm(input.table.red[,i]~input.table.red[,3]+input.table.
        red[,6]+input.table.red[,7]+input.table.red[,8]+input.
        table.red[,9]+input.table.red[,10]+input.table.red[,11]+
        input.table.red[,14]+input.table.red[,15]+input.table.red
        [,16]+input.table.red[,17]+input.table.red[,18]+input.
        table.red[,19]+input.table.red[,20]+input.table.red[,21]+
        input.table.red[,22]+input.table.red[,23]+input.table.red
        [,24]+input.table.red[,25]+input.table.red[,26]+input.
        table.red[,27]+input.table.red[,28]+input.table.red[,29]+
        input.table.red[,30]+input.table.red[,34]+input.table.red
        [,35]+input.table.red[,36]+input.table.red[,37]+input.
        table.red[,41]+input.table.red[,42]+input.table.red[,43]+
        input.table.red[,45]+input.table.red[,46])
    temp1=AIC(temp)
    temp2=summary(temp)
    results[i,2]=temp2$adj.r.squared
    results[i,3]=temp1
  }
  #poisson
  for (i in k){
    temp=gamlss(input.table.red[,i]~input.table.red[,3]+input.
        table.red[,6]+input.table.red[,7]+input.table.red[,8]+
        input.table.red[,9]+input.table.red[,10]+input.table.red
        [,11]+input.table.red[,14]+input.table.red[,15]+input.
        table.red[,16]+input.table.red[,17]+input.table.red[,18]+
        input.table.red[,19]+input.table.red[,20]+input.table.red
        [,21]+input.table.red[,22]+input.table.red[,23]+input.
        table.red[,24]+input.table.red[,25]+input.table.red[,26]+
        input.table.red[,27]+input.table.red[,28]+input.table.red
        [,29]+input.table.red[,30]+input.table.red[,34]+input.
        table.red[,35]+input.table.red[,36]+input.table.red[,37]+
        input.table.red[,41]+input.table.red[,42]+input.table.red
        [,43]+input.table.red[,45]+input.table.red[,46],family=PO
        (mu.link = "log"))
    temp1=AIC(temp)
    results[i,4]=temp1
  }
  #neg.bin1
  for (i in k){
   temp=gamlss(input.table.red[,i]~input.table.red[,3]+input.
      table.red[,6]+input.table.red[,7]+input.table.red[,8]+
      input.table.red[,9]+input.table.red[,10]+input.table.red
      [,11]+input.table.red[,14]+input.table.red[,15]+input.
      table.red[,16]+input.table.red[,17]+input.table.red[,18]+
      input.table.red[,19]+input.table.red[,20]+input.table.red
      [,21]+input.table.red[,22]+input.table.red[,23]+input.
      table.red[,24]+input.table.red[,25]+input.table.red[,26]+
      input.table.red[,27]+input.table.red[,28]+input.table.red
      [,29]+input.table.red[,30]+input.table.red[,34]+input.
      table.red[,35]+input.table.red[,36]+input.table.red[,37]+
      input.table.red[,41]+input.table.red[,42]+input.table.red
      [,43]+input.table.red[,45]+input.table.red[,46],family=NBI
      (mu.link="log",sigma.link="log"))
   print(i)
   temp1=AIC(temp)
   results[i,5]=temp1
  }
  #neg.bin2
```

```
for (i in k){
   temp=gamlss(input.table.red[,i]~input.table.red[,3]+input.
      table.red[,6]+input.table.red[,7]+input.table.red[,8]+
      input.table.red[,9]+input.table.red[,10]+input.table.red
      [,11]+input.table.red[,14]+input.table.red[,15]+input.
      table.red[,16]+input.table.red[,17]+input.table.red[,18]+
      input.table.red[,19]+input.table.red[,20]+input.table.red
      [,21]+input.table.red[,22]+input.table.red[,23]+input.
      table.red[,24]+input.table.red[,25]+input.table.red[,26]+
      input.table.red[,27]+input.table.red[,28]+input.table.red
      [,29]+input.table.red[,30]+input.table.red[,34]+input.
      table.red[,35]+input.table.red[,36]+input.table.red[,37]+
      input.table.red[,41]+input.table.red[,42]+input.table.red
      [,43]+input.table.red[,45]+input.table.red[,46],family=
      NBII(mu.link=”log”,sigma.link=”log”))
   temp1=AIC(temp)
   print(i)
   results[i,6]=temp1
 }
 #best model
 for (i in k){
   if(min(results[i,3:6]>0)){
   results[i,7]=names(which.min(results[i,3:7]))}else{”/”}
 }
 return(results)
 }
```

Glossary

Aspect In English there are four aspects referring to the temporal manner of the action. The perfect aspect refers to an action which happened before another action. The progressive refers to an ongoing action. The perfect progressive is a combination of both, whereas the simple aspect rejects the perfect, progressive and perfect progressive aspect (Pullum & Huddleston 2002, Huddleston & Pullum 2007).

Contrastive Analysis The Contrastive Analysis compares two versions of the same text. The first version is kept in the native language of its writer while the second is translated into another language.

Contrastive Interlanguage Analysis The Contrastive Interlanguage Analysis contrasts the inter-language performance of non-native speakers with the language performance of native speakers. Additionally, it compares the inter-languages of various non-native speaker groups with each other (Granger 1996).

English as a Foreign Language English as a Foreign Language refers to the performance of English being learnt and not acquired.

Interlanguage According to Larry Selinker (1972) learners of a L2 or foreign language who have not yet achieved a full native-like language performance use a ”separate linguistic system” (214) that neither completely follows the rules of the L1 nor the target language.

Mode In English there are two modes: the active and the passive mode. Since, the mode is the central topic of this paper the reader is referred to the corresponding chapters 2 and 3.3.

Native Speaker The term native speaker refers to a person speaking a language which is her first language.

Non-Native Speaker The term non-native speaker refers to person speaking a language which is not her first language.

R R is a programming language like Java, C or Python as well as a programme and is distributed by the Comprehensive R Archive Network (CRAN). It is mostly used for statistical analyses.

Tenses In English there are three tenses referring to the present, the past and the future (Pullum & Huddleston 2002, Huddleston & Pullum 2007).

Voice In English there are two voice. The indicative refers to a true statement while the conditional refers to a hypothetical statement (Pullum & Huddleston 2002, Huddleston & Pullum 2007).

Plagiatserklärung

Hiermit versichere ich, Tobias Gärtner, dass ich die anliegende Arbeit selbst angefertigt und alle für die Arbeit verwendeten Quellen und Hilfsmittel vollstä angegeben habe.

Ich habe die Arbeit noch nicht zum Erwerb eines anderen Leistungsnachweises eingereicht.

Mit der Übermittlung meiner Arbeit auch an externe Dienste zur Plagiatsprüfung bin ich einverstanden.

Ort, Datum Unterschrift