**A corpus-based analysis of adjective amplification among native speakers and learners of English**

**Abstract**

This paper analyses adjective amplification by native speakers and learners of English with diverse language-backgrounds based on the *International Corpus of Learner English* and the *Louvain Corpus of Native English Essays*. Configuration frequency analysis (CFA) to find collocational differences between learners and native speakers and Multifactorial Prediction and Deviation Analysis Using Regression/Random Forests (MuPDARF) to evaluate which factors contribute to learners taking non-nativelike decisions with respect to the use of *very*. The CFA shows that the all-purpose amplifier *very* is rather inconspicuous while specific combinations (*extremely difficult* and *completely different*) are systematically overused. The MuPDARF provides a detailed picture of divergences between native speakers and learners and reports that learners are significantly more likely to use *very* in a non-nativelike manner in not primed contexts and when amplifying non-emotional adjectives. In addition, learners use *very* non-nativelike with low frequency adjectives in predicative contexts.

1. **Introduction**

The rise of learner corpus research in the past three decades or so has advanced reliable empirical research of language background specific difficulties that learners from diverse language backgrounds face. In addition, ever more sophisticated statistical modeling allows us to analyze systematic effects of the language background on the acquisition process. Arguably, statistical methods that appropriately model the cognitive mechanisms that underly the acquisition process, such as *Multifactorial Prediction and Deviation Analysis Using Regression/Random Forests* (MuPDARF; Gries and Deshors 2014; Heller, Bernaisch and Gries 2017) have only relatively recently been introduced in learner corpus research (see Gries 2018 for an elaboration of this argument). The present paper combines learner corpus research with recently introduced multivariate statistical methods to unearth systematic divergencies between native speakers and learners of English with diverse language backgrounds with respect to adjective amplification. As such, the present study serves as a case study that exemplifies not only how modern statistical methods can be used to detect and evaluate language background specific difficulties but also which language-internal and – external factors contribute to non-nativelike performance.

Adjective amplification (see (1)) as a linguistic phenomenon is particularly relevant from the perspective of a learner of English because amplification plays a crucial role in the social and emotional expression of speakers (Ito and Tagliamonte 2003: 258) and because nativelike use is difficult to acquire as amplifier systems are particularly prone to change (Quirk et al. 1985:590).

(1) a. […] because they were *very poor*. (ICLE-BG-SUN-0003.1)[[1]](#endnote-1)

b. At the same time they made drugs *really popular*. (ICLE-PO-POZ-0046.3)

c. […] it is *absolutely irresponsible* to fight against another denomination […]. (ICLE-GE-SAL-0009.3)

d. […] they will *completely disappear* in fifty years at latest. (ICLE-RU-MOS-0020.6)

e. The countries are *so different* (culture, language, tradition) […]. (ICLE-FR-UCL-0067.1)

f. In others words, Europe has to become a *very strong* power. (ICLE-FR-UCL-0066.1)

g. Also if they are lucky and she gets a *really suitable* husband […]. (ICLE-SW-LND-0023.8)

h. […] you will have an *absolutely excellent* view on the match. (ICLE-DN-NIJ-0007.2)

i. These two *completely antagonistic* characters have nothing to envy one another. (ICLE-CZ-PRAG-0006.1)

The paper takes a comparative approach to analyzing differences in amplifier use between native speakers and leaners of English from diverse language backgrounds and adds to existing research by using a substantially sized data base and advanced statistical modeling.

The following section surveys previous research on adjective amplification with a focus on adjective amplification among learners of English. The third section provides an overview of the data bases, the data processing steps, and the statistical methods. The fourth section presents the results of the analysis while the fifth section discusses the results in light of previous research and potential shortcomings of the present analysis. The sixth section offers an outlook for potential directions of future research.

1. **Previous Research**

Adjective amplification has received a substantial amount of attention, particularly from historical and functional perspectives (e.g., Bolinger 1972; Breban and Davidse 2016; Lorenz 2002; Méndez-Naya 2003, 2008; Méndez-Naya and Pahta 2010; Nevalainen 2008; Nevalainen and Rissanen 2002; Paradis 2008; Partington 1993; Peters 1992; Rissanen 2008). Furthermore, over the past two decades, sociolinguistic or variationist studies have provided fine-grained analyses of amplifier use across social strata and within various varieties of English (e.g., Bauer and Bauer 2002; D’Arcy 2015; Fuchs 2017; Fuchs and Gut 2016; Ito and Tagliamonte 2003; Macaulay 2002; Pertejo and Martínez 2014; Martínez and Pertejo 2012; Stenström 1999; Tagliamonte 2008; Tagliamonte and Denis 2014; Tagliamonte and Roberts 2005). However, despite having been extensively studied, relatively little attention has been placed on the role that the adjective in amplified adjective—notable exceptions are Wagner (2017) and Hendrikx, Van Goethem, and Wulff (2019)—or on how native speakers and learners of English differ with respect to their use of amplifiers.

The most extensive comparative study that has focused on adjective intensification among native speakers and learners of English is Lorenz (1999) and other studies by Lorenz which complement his monograph-length study (e.g., Lorenz 1998, 2002). Lorenz (1999: 198) interprets the ICLE data to show that German learners of English overuse the all-purpose intensifiers *very* and *really* which they use as in speech rather than academic written discourse (1999: 215). In a similar vein, Hinkel (2003) found that the patterning of adverbs in data coming from learners of English with Chinese, Japanese, Korean, and Indonesian language backgrounds mirrored the frequencies in conversational style, thus indicating that learners used colloquial style rather than a more appropriate formal style due to a limited lexical repertoire (2003: 1058).

Lorenz (1998) also found that learners exhibit a lower degree of cohesion in their use of amplifier-adjective collocations and amplify significantly more frequently than native speakers both relatively and in terms of absolute frequencies (Lorenz 1998: 64). While this finding contarsts with studies by Granger (1998) and Hendrikx, Van Goethem, and Wulff (2019) who found that native speakers amplify significantly more than learners, Lorenz (1998: 62-63) argues that the tendency among German learners to amplify more frequently is a simple misapplication and can be attributed to overall writing strategies of advanced learners. According to Lorenz (1998), this misapplication is caused by the learner’s desire for expressiveness. The points made by Lorenz (1998) are supported by a detailed analysis of a multitude of examples while his statistical analysis is less transparent because focuses on differences between semantic classes of intensifiers rather than individual amplifier variants.

Similar to Lorenz (1998, 1999, 2002), Granger (1998) has analysed amplifier-adjective combinations in written data from native English speakers and French learners of English. To complement corpus-based analysis, Granger (1998) uses additional word combination tasks to tap into intuitional differences between native speakers and French learners of English. Contrasting with Lorenz (1999), Granger (1998) found that native-speakers amplified substantially more than learners and the results of the word combination task indicated that learners showed significantly more variation in what they deemed acceptable pairs compared with native speakers. Granger (1998) argues that these results suggest that the sense of collocational salience was missing or misguided among (French) learners of English (1998: 152).

An attempt to partially replicate Granger’s (1998) study has been undertaken by Edmonds and Gudmestad (2014) who analyse intensifying adverbials in academic essays. Edmonds and Gudmestad (2014) study indicated that advanced learners of English preformed like native speakers of English while less advanced learners differed substantially from both advanced learners and native speakers. This study thus substantiates findings provided by Forsberg (2010) who focused on highly advanced learners of English and also found that differences between learners and native speakers of English wane off as learners become more proficient. Similarly, Hendrikx, Van Goethem, and Wulff (2019: 80-81) found that while learners amplified significantly less than native speakers, more advanced leaners approximate the rate of amplification exhibited by native speakers.

A study by Hendrikx, Van Goethem, and Wulff (2019) investigated the use of intensifying constructions among native speakers and learners of French, Dutch, and English. The aim of this study was to evaluate cross-linguistic influence and the impact of Content and Language Integrated Learning (CLIL) on the acquisition of intensifying constructions, including amplifier-adjective bigrams. While this study used (fictional) e-mails to a friend on the topics of a party or holidays as a corpus and thus investigated rather informal texts, it showed that students in CLIL produced more target-like intensification compared with students in non-CLIL. Target likeness was shown, e.g., by less overuse of all-round intensifiers and collostructions containing advanced vocabulary as well as informal intensifying language (Hendrikx, Van Goethem, and Wulff, 2019: 97). Despite being methodologically very well-designed, Hendrikx, Van Goethem, and Wulff (2019) is less relevant in the present context as the data is significantly less formal and because the outlook of that study differs substantially from the current analysis.

In their research on the expression of stance in academic writing by learners of English, Fallas Escobar and Chaves Fernández (2017) found that use of boosters (amplifiers), together with stance-taking strategies, enabled students to develop a stronger voice (Fallas Escobar and Chaves Fernández 2017: 118). Finally, Friginal et al. (2017), who analyzed learner and teacher talk, found that learners of English used significantly fewer booster types compared to their teachers while students and teachers did not differ substantially with respect to token frequencies (Friginal et al. 2017: 89). After reviewing the findings provided by previous research, we will now focus on the data sources and the methods applied in the present study.

1. **Data and Methodology**

The following subsection presents the data used in the present study, describes the data processing, and surveys the processed data while the next subsection describes the statistical methods that are used.

3.1 Data sources and processing

This study is based on two data sources. The first source of data is a subsection of the *International Corpus of Learner English* (ICLE, Granger, 2002; see Granger, 1993). The version of the ICLE used in the present analysis was published in 2002 and consists of argumentative writing by intermediate to advanced EFL learners with Bulgarian, Czech, Dutch, Finnish, French, German, Italian, Norwegian, Polish, Russian, Spanish, and Swedish language backgrounds, while data from speakers with Chinese, Japanese, Tswana, and Turkish were excluded because their data was not available for the present study.

The second data source is the *Louvain Corpus of Native English Essays* (LOCNESS) which has been specifically compiled to provide data comparable to the ICLE. The LOCNESS represents native English essays and consists of British pupils’ A level essays (60,209 words), British university students’ essays (95,695 words), and American university students’ essays (168,400 words).

The data processing of both corpora was done in the programming environment R (R Core Team 2008) and was identical for both corpora to avoid incomparability issues. In a first step, the data were cleaned and harmonized by removing, for instance, meta-data such as file identifiers. In a next step, the lexical richness of each essay in the data was calculated as in Tweedie and Baayen (1998) using the quanteda package (Benoit et al., 2018) which provides ten different measures of lexical richness (The ordinary Type-Token Ratio (TTR), Herdan’s C, Guiraud’s Root TTR, Carroll’s Corrected TTR, Dugast’s Uber Index, Summer’s index, Yule’s K, Simpson’s D, Herdan’s V\_m, and the Maas Index). Since the lexical richness measures were highly correlated (a principal component analysis showed that the first component explained 95.8 percent of the variance for the native speaker data and 98.4 percent of the non-native speaker data), the present study chose Yule’s K (Yule 1944) as a measure of lexical richness which serves as a proxy for linguistic proficiency in the present study. Then, the token frequency of adjective types was calculated separately for each language to obtain a frequency measure which can be used to control for frequency effects. The cleaned data were then part-of-speech tagged by implementing a maximum entropy tagger provided in the openNLP package (Hornik 2016). After part-of-speech-tagging, all adjectives (tag JJ) were extracted and it was determined for each adjective whether it was amplified and which lexical form served as an amplifier. In a next step, the syntactic function of the adjective (attributive or predicative) was determined and annotation was added that showed whether the same amplifier type had occurred within a span of up to three previous pre-adjectival slots to test potential persistence or priming effects (cf. Tulving and Schacter 1990: 301; Szmrecsanyi 2005: 113; also Szmrecsanyi 2006). After coding for priming, negated adjectives, misclassified items, as well as comparative and superlative forms were removed from the analysis. In addition, adjectives that were never amplified, or which were not amplified by at least two different amplifier types, were removed from the analysis. Then, a sentiment analysis was applied to the adjectives in the data using the syuzhet package in R (Jockers 2017). The sentiment analysis was implemented because the amplifier choice has been shown to be affected by the emotionality of the amplified adjective. The results of the sentiment analysis were used to determine whether a given adjective was positive (e.g. *happy*, *nice*), negative (e.g. *sad*, *angry*) or non-emotional (e.g. *rusty* or *flat*). Next, all remaining adjectives were classified semantically based on a simplified version of the classification provided by Dixon (1977, 2004; cf. also D’Arcy 2015; Tagliamonte 2008; Tagliamonte and Roberts 2005). In a next step, adjectives were coded for gradability. This means that each adjective was annotated as being gradable (in case the adjective denoted a scale as with the adjective *cool*), non-gradable (if the adjective denoted a state as with the adjective pregnant), or whether the adjective was neither clearly gradable nor non-gradable (as with the adjectives *subjective* or *disputable*). The coding of the gradability of adjectives were based on both data-driven criteria and human expert annotation. The previously described data processing produced the data set that is summarized in Table 1.

Table 1: Overview of adjective tokens, amplified adjective tokens, and percentage of amplification across languages.

|  |  |  |  |
| --- | --- | --- | --- |
| **Language** | **Adjective tokens** | **Amplified adj. tokens** | **Percent amplified adj. tokens** |
| Bulgarian | 8,594 | 455 | 5,3 |
| Czech | 8,345 | 648 | 7,8 |
| Dutch | 6,126 | 395 | 6,4 |
| English | 3,201 | 303 | 9,5 |
| Finnish | 7,114 | 453 | 6,4 |
| Flemish | 3,616 | 277 | 7,7 |
| French | 11,055 | 540 | 4,9 |
| German | 8,457 | 575 | 6,8 |
| Italian | 9,611 | 514 | 5,3 |
| Polish | 9,919 | 655 | 6,6 |
| Russian | 9,295 | 674 | 7,3 |
| Spanish | 7,169 | 529 | 7,4 |
| Swedish | 12,020 | 731 | 6,1 |
| **Total** | 104,522 | 6,749 | 6.46 |

Table 1 shows that the vast majority of adjectives are not amplified with an average of only 6.43 adjectives being amplified. In addition, the rate of amplification is relatively homogeneous with a range of 4.9 percent (French) to 9.5 percent (English). Thus, native English speakers amplify more compared to the English learners. Figure 1 shows the rates of amplification across languages in descending order from left to right.

|  |
| --- |
| Figure 1: Percentages of amplified adjectives in English by L1. |

Figure 1 shows that native English speakers have higher rates of amplification compared to the learners of English. In fact, the rate of amplification by native speakers is markedly higher than the rates of amplification by the learners – a trend which holds across syntactic contexts.

Before continuing with the final data processing steps, we will briefly check the proficiency scores across languages. Figure 2 shows that native speakers have the highest proficiency score. The score for native English speakers is, indeed, markedly higher compared to the scores of the learners. This is comforting as it suggests that the proficiency does indeed capture proficiency.

|  |
| --- |
| Figure 2: Proficiency scores derived from the lexical richness measure Yule’s K across language backgrounds. |

In a final step of data processing, all adjective tokens that were not amplified were removed from the analysis. The reasoning behind the decision to remove non-amplified tokens was that the linguistic variable can be defined as a situation in which “the speaker reaches a decision-point” (Wallenberg 2013; cited in Maddeaux and Dinkin 2017). The variable in this context is the decision whether or not to amplify an adjective given that the speaker has already decided to modify this adjective somehow. The variable context thus encompasses only amplified adjectives while leaving out zero context, i.e. contexts where the speaker could have amplified an adjective but did not. This means that not amplified adjectives had to be removed as to only focus on contexts that allow for amplification and thus represent a variable context. This final data set is summarized in Table 3.

Table 2: Overview of the absolute token frequencies and percentages of amplifier types across languages with infrequent amplifier types being collapsed into the category “other”.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Language** |  | **completely** | **extremely** | **other** | **really** | **so** | **very** | **Total** |
| Bulgarian | N | 16 | 26 | 114 | 28 | 59 | 212 | 455 |
| % | 3.5 | 5.7 | 25.1 | 6.2 | 13.0 | 46.6 |  |
| Czech | N | 31 | 14 | 98 | 50 | 82 | 373 | 648 |
| % | 4.8 | 2.2 | 15.1 | 7.7 | 12.7 | 57.6 |  |
| Dutch | N | 15 | 13 | 77 | 14 | 39 | 237 | 395 |
| % | 3.8 | 3.3 | 19.5 | 3.5 | 9.9 | 60.0 |  |
| English | N | 4 | 14 | 66 | 5 | 40 | 174 | 303 |
| % | 1.3 | 4.6 | 21.8 | 1.7 | 13.2 | 57.4 |  |
| Finnish | N | 11 | 26 | 102 | 13 | 63 | 238 | 453 |
| % | 2.4 | 5.7 | 22.5 | 2.9 | 13.9 | 52.5 |  |
| Flemish | N | 9 | 8 | 33 | 16 | 29 | 182 | 277 |
| % | 3.2 | 2.9 | 11.9 | 5.8 | 10.5 | 65.7 |  |
| French | N | 14 | 17 | 90 | 21 | 72 | 326 | 540 |
| % | 2.6 | 3.1 | 16.7 | 3.9 | 13.3 | 60.4 |  |
| German | N | 23 | 25 | 121 | 52 | 75 | 279 | 575 |
| % | 4 | 4.3 | 21 | 9 | 13 | 48.5 |  |
| Italian | N | 29 | 8 | 74 | 44 | 62 | 297 | 514 |
| % | 5.6 | 1.6 | 14.4 | 8.6 | 12.1 | 57.8 |  |
| Polish | N | 19 | 51 | 124 | 39 | 86 | 336 | 655 |
| % | 2.9 | 7.8 | 18.9 | 6 | 13.1 | 51.3 |  |
| Russian | N | 16 | 27 | 111 | 54 | 93 | 373 | 674 |
| % | 2.4 | 4 | 16.5 | 8 | 13.8 | 55.3 |  |
| Spanish | N | 23 | 19 | 52 | 27 | 57 | 351 | 529 |
| % | 4.3 | 3.6 | 9.8 | 5.1 | 10.8 | 66.4 |  |
| Swedish | N | 18 | 45 | 122 | 30 | 113 | 403 | 731 |
| % | 2.5 | 6.2 | 16.7 | 4.1 | 15.5 | 55.1 |  |
| Total | N | 228 | 293 | 1,184 | 393 | 870 | 3,781 | 6,749 |
| % | 3.4 | 4.3 | 17.5 | 5.8 | 13.0 | 56.0 |  |

The percentages presented in table 2 are visualized in figure 3.

|  |
| --- |
| Figure 3: Percentages of amplifier types of all amplifiers by language background and syntactic function. |

Figure 3 shows that the use of amplifiers across language backgrounds is relatively consistent with some divergence with respect to the percentages of *very*. After providing an overview of the data used in the present study, the following section provides information about the operationalization of variables.

3.2 Classification and coding of variables

The following section describes the variables included in the statistical analysis.

*very*: the dependent variable of the random forest models is the occurrence of amplifying, pre-adjectival *very*. Each adjective in the data was coded as 1 if it was amplified by very or as 0 if it was intensified by another amplifier.

*NonNativeLike*

NonNativeLike is the dependent variable of the mixed-effects binomial logistic regression and it reflects whether the observed behaviour of a learner differs from the prediction of what a native speaker would have done in the respective context based on the random forest model. If the observed choice of the learner agreed with the prediction, the instance is coded as 0 while it was coded as 1 if the observed choice and the prediction differ.

*Language*

Language refers to the language background of the speakers in the data (Bulgarian, Czech, Dutch, Finnish, French, German, Italian, Norwegian, Polish, Russian, Spanish, and Swedish).

*Priming*

Priming in the present study refers to instances of production priming (cf. Szmrecsanyi 2005: 113) and can therefore be defined as re-use of linguistic material that was used in the preceding discourse (cf. Tulving and Schacter 1990: 301). While there exists a substantial amount of research on priming both in psycholinguistics and corpus linguistics, various issues remain unclear. One such issue relates to the duration of priming effects as the decay time varies from milliseconds, in the case of syntactic, form, and production priming, to months or even years in cases of semantic or conceptual priming (Althaus and Kim 2006: 962). The present study uses a scope of three adjectival slots as a window in which priming may occur due to the fact that form priming is short-lived and disappears soon after exposure to the stimulus prime (Althaus and Kim 2006: 962). The current study thus assumes that priming is present if the same amplifier is reused in at least one out of the subsequent three pre-adjectival slots.

*Emotionality*

The emotionality of adjectives was coded by implementing a Sentiment Analysis using the syuzhet package in R (Jockers 2017). The sentiment analysis used in the current study relies on the Word-Emotion Association Lexicon (Mohammad and Turney 2013; cf. http://www.purl.org/net/NRCemotionlexicon), which comprises 10,170 terms based on 38,726 ratings from 2,216 raters. The emotion coding of this lexicon is based on ratings gathered through the crowed-sourced Amazon Mechanical Turk service. In this Turk survey, raters were asked a sequence of questions relating to whether a given word was associated with one of eight emotions (ANGER, ANTICIPATION, DISGUST, FEAR, JOY, SADNESS, SURPRISE, TRUST). Each term was rated at least five times and for 85 percent of words, four or more raters provided identical ratings. According to the ratings, words like *dark* or *tragic* are more readily associated with SADNESS and words such as *happy* or *beautiful* are associated with JOY while words like *cruel* or *outraged* are associated with ANGER. In the present study, adjective associated with ANGER, DISGUST, FEAR, or SADNESS are coded as NegativeEmotional while adjectives associated with ANTICIPATION, JOY, SURPRISE, or TRUST are coded as PositiveEmotional. Adjectives that are not associated with any emotional state are coded as NonEmotional.

*SemanticCategory*

The semantic classification of adjectives uses a simplified version of the semantic categories proposed by Dixon (1977, 2004). The reason for simplifying Dixon’s (1977, 2004) categorization relates to the uneven distribution of category members. Because the data contained hardly any age and colour terms, such adjective types as well as adjectives that could not be categorized (e.g. *familiar*, *genuine*, or *inadequate*) were assigned the label NoSemType. The resulting factor consisted of five levels (*Difficulty, Dimension*, *HumanPropensity*, *NoSemType*, *PhysicalProperty*, and *Value*).

*Gradability*

Gradability refers to the semantic property of adjectives to represent either points on a scale (cf. Quirk et al. 1985), in which case an adjective was coded as gradable (for instance *hot*, *low*, *big*, *narrow*, or *intelligent*), or limits of a scale, in which case the adjective was coded as being non-gradable (for instance *wooden* or *married*). Although amplifiers are commonly considered to co-occur only with gradable adjectives, they can also co-occur with non-gradable adjectives for pragmatic purposes, e.g. for emphasis. Furthermore, during its grammaticalization, *very* has spread from non-gradable to gradable adjectival contexts (Adamson and González-Díaz 2004 cited in Tagliamonte 2008) indicating that co-occurrence with non-gradable adjectives may be indicative of innovative amplifier variants. All adjectives that could not be classified as being gradable or non-gradable were coded as ‘GradabilityUndetermined’.

*Adjective*

Adjective refers to the adjective type that is amplified. Given the relatively moderate size of the data, adjectives that were not used at least three times in any of the language background specific subsections were collapsed. This resulted in all adjectives except for *different*, *good*, *hard*, *difficult*, and *important* being collapsed into a single category (*other*).

*Frequency*

Frequency refers to the frequency of the adjective types by language background (as stated above, adjectives except for *different*, *good*, *hard*, *difficult*, and *important* were collapsed into the category *other*). In addition, a first MuPDARF showed that the effect of frequency exhibited substantial curvature in its correlation with the use of *very* in the random forest models and with nativelike choices in the regression model. Therefore, frequency was transformed into a polynomial to the second degree to avoid curvature in effect of frequency (see Wulff and Gries to appear).

After reviewing the operationalization of the variable included in the statistical analysis, we now turn to an overview of the variables and their levels with respect to the frequency of *very*.

Table 3: Variables and variable levels with frequency and percentage of very included in the random forest analysis.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable** | **Scale** | **Level** | **very (N)** | **Adjectives (N)** | **very (%)** |
| **Dependent variable** | | | | | |
| very | nominal |  | 3,781 | 6,749 | 56.0 |
| **Independent variables** | | | | | |
| Language | categorical | English (reference var.) | 174 | 303 | 57.4 |
| Bulgarian | 212 | 455 | 46.6 |
| Czech | 373 | 648 | 57.6 |
| Dutch | 237 | 395 | 60.0 |
| Finnish | 238 | 453 | 52.5 |
| Flemish | 182 | 277 | 65.7 |
| French | 326 | 540 | 60.4 |
| German | 279 | 575 | 48.5 |
| Italian | 297 | 514 | 57.8 |
| Polish | 336 | 655 | 51.3 |
| Russian | 373 | 674 | 55.3 |
| Spanish | 351 | 529 | 66.4 |
| Swedish | 403 | 731 | 55.1 |
| Priming | nominal | NoPrime (ref. var.) | 3,379 | 6,242 | 54.1 |
| Prime | 402 | 507 | 79.3 |
| Emotionality | categorical | Negative (ref. var.) | 451 | 1025 | 44.0 |
| NonEmotional | 2260 | 3914 | 57.7 |
| Positive | 1070 | 1810 | 59.1 |
| Function | nominal | Attributive (ref. var.) | 1735 | 2659 | 65.3 |
| Predicative | 2046 | 4090 | 50.0 |
| SemanticCategory | categorical | Difficulty | 344 | 505 | 40.5 |
| Dimension | 570 | 767 | 42.6 |
| HumanPropensity | 467 | 876 | 34.8 |
| NoSemType (ref. var.) | 932 | 2086 | 30.9 |
| PhysicalProperty | 359 | 708 | 33.6 |
| Value | 1109 | 1807 | 38.0 |
| Gradability | categorical | GradabilityUndetermined (ref. var.) | 77 | 213 | 26.6 |
| Gradable | 1813 | 2988 | 37.8 |
| NotGradable | 1891 | 3548 | 34.8 |
| Adjective | categorical | different | 115 | 325 | 26.1 |
| difficult | 221 | 312 | 41.5 |
| good | 140 | 207 | 40.3 |
| hard | 86 | 133 | 39.3 |
| important | 441 | 567 | 43.8 |
| other (ref. var.) | 2778 | 5205 | 34.8 |
| Frequency | numeric |  |  |  |  |

After surveying the data sources and the variables included in the MuPDARF analysis, we will now turn to the statistical procedures used in the present analysis.

3.3 Statistical methods

The present study makes use of configural frequency analysis (CFA; Krauth and Lienert 1973), to determine which amplifier-adjective bigrams differ significantly between native speakers and learners of English and Multifactorial Prediction and Deviation Analysis Using Regression/Random Forests (MuPDARF; Gries and Deshors, 2014; Heller, Bernaisch, and Gries 2017), to provide a very fine-grained understanding of how native speakers, and learners of English differ. In the present case, we use the MuPDARF approach to find out about which factors contribute to non-native like choices of learners with respect to the use of *very*. In the following, these statistical methods are explained in order to facilitate the rationale behind these methods.

*Configuration Frequency Analysis*

Configuration frequency analysis or configural frequency analysis (Krauth and Lienert 1973; Bortz, Lienert and Boehnke 2008) allows us to detect combinations of amplifiers and adjectives (configurations) that occur significantly more or less frequently in learner data than would be expected given native speaker use. In contrast to the χ2-family of tests that only provide information about whether any of the configurations differ significantly from what would be expected by chance, CFAs enable us to determine exactly which configurations differ significantly between learners and native speakers. In other words, CFAs allow us to test whether an individual configuration occurs significantly more often (Type) or less often (Antitype) in the learner data than would be expected given the frequency of the configuration in the native speaker data. To control for α-error rate inflation, the present study has made use of Bonferroni corrections (see Gries 2018 for an argument why corrections are necessary).

*Multifactorial Prediction and Deviation Analysis Using Regression/Random Forests*

A Multifactorial Prediction and Deviation Analysis Using Regression/Random Forests (MuPDARF), for short, was only recently introduced to SLA research (see Gries and Adelman 2014; Gries and Deshors 2014; Heller, Bernaisch, and Gries 2017). MuPDARF enable us to investigate the difference between the choices of learners and native speakers in a much more detailed and fine-grained manner than CFAs. In the present case, the MuPDARF allows us to investigate exactly which factors that correlate with learners making non-nativelike choices and how strongly these factors affect the degree of non-nativeness. The MuPDARF analysis is fit to the data with the occurrence of *very* versus the occurrence of other amplifiers (*other*) being the dependent variable. This is particularly interesting in the because (i) *very* is the most frequently used amplifier and (ii) because it has the potential to show how much more detailed our understanding of what affects learners’ choices can become if we make use of advanced multivariate methods. A MuPDARF analysis consists of the following steps:

1. Fit a random forest to the native-speaker data to determine which factors contributed to what extend to the decision of a native speaker.
2. Apply the random forest obtained from the native speaker data to the learner data to predict which choice a native speaker would have made given the same conditions.
3. Save the predictions of the predicted choices in a vector and compare the choices actually made by the learners to the predicted choices.
4. Then, a second vector is created in which instances where the observed and the predicted choices are identical are coded as 0 and as 1 if the predicted and the observed choices differ.
5. This second vector serves as the dependent variable of a mixed-effects binomial regression model. The results of that mixed-effects regression model then inform which factors contribute to learners making nativelike choices.

The regression modelling, that was implemented during the last step of MuPARF, used both adjective type and language background as random effects. Other random effect structures were evaluated but resulted in models with a substantially higher Akaike Information Criterion (AIC) value, indicating a suboptimal model fit. Model fitting was done by following a step-wise step up procedure during which predictors (variables and interactions between variables) are consecutively added to a model. Predictors were retained if they significantly improved model fit and if their inclusion did not result in unacceptable multicollinearity. Multicollinearity was measured via variance inflation factors (VIFs). The cut-off point for VIFs was a maximum value of 3 (see Zuur et al. 2010). In cases where the VIFs were unacceptable, model fitting continued without that predictor. The model fitting aims to find the final minimal regression model which explains a maximum of variance with a minimum number of predictors. Once a final minimal model is found, this model provides information about which factors contribute to what extent to learners making non-nativelike choices.

1. **Results**

This section presents the results of the statistical analyses. We begin with the results of the CFA. Table 4 presents significant types and antitypes of amplifier-adjective bigrams. Types are amplifier-adjective combinations which are used significantly more frequently compared with their frequency in the native speaker data while antitypes indicate that an amplifier-adjective combination occurs significantly less often than in the native speaker data.

Table 4: Results of the covarying collexeme analysis.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Language** | **Amplifier** | **Adjective** | **Obs. N** | **Exp. N** | **z-value** | **Pr(>|z|)** | **Type** |
| Bulgarian | completely | different | 6 | 0.43 | 8.52 | 0.00000\*\*\* | Type |
| very | little | 28 | 11.92 | 4.66 | 0.00000\*\*\* | Type |
| very | other | 110 | 157.12 | 3.76 | 0.00009\*\*\* | Antitype |
| Czech | completely | different | 14 | 0.9 | 13.78 | 0.00000\*\*\* | Type |
| Dutch | completely | different | 8 | 0.59 | 9.69 | 0.00000\*\*\* | Type |
| extremely | difficult | 4 | 0.81 | 3.54 | 0.0002\*\*\* | Type |
| other | different | 11 | 4.41 | 3.14 | 0.00084\*\*\* | Type |
| Finnish | completely | different | 5 | 0.42 | 7.11 | 0.00000\*\*\* | Type |
| really | hard | 2 | 0.27 | 3.32 | 0.00045\*\*\* | Type |
| French | completely | different | 3 | 0.55 | 3.32 | 0.00046\*\*\* | Type |
| German | completely | different | 6 | 0.44 | 8.35 | 0.00000\*\*\* | Type |
| really | hard | 5 | 1.19 | 3.49 | 0.00024\*\*\* | Type |
| Italian | completely | different | 14 | 1.14 | 12.02 | 0.00000\*\*\* | Type |
| Polish | completely | different | 5 | 0.43 | 7 | 0.00000\*\*\* | Type |
| really | good | 6 | 0.94 | 5.21 | 0.00000\*\*\* | Type |
| extremely | difficult | 10 | 2.51 | 4.74 | 0.00000\*\*\* | Type |
| other | different | 12 | 3.53 | 4.51 | 0.00000\*\*\* | Type |
| Russian | completely | different | 3 | 0.38 | 4.24 | 0.00001\*\*\* | Type |
| other | different | 11 | 3.37 | 4.15 | 0.00002\*\*\* | Type |
| Spanish | completely | different | 10 | 0.76 | 10.57 | 0.00000\*\*\* | Type |
| very | important | 54 | 35.71 | 3.06 | 0.0011\*\*\* | Type |
| really | important | 5 | 2.18 | 1.91 | 0.02782\*\*\* | Type |
| Swedish | completely | different | 5 | 0.57 | 5.86 | 0.00000\*\*\* | Type |
| extremely | difficult | 6 | 1.73 | 3.24 | 0.0006\*\*\* | Type |

Table 5 shows that learners tend to significantly overuse collocations while they only rarely underuse collocations – the only exception being *very* collocating with low frequency adjectives in Bulgarian. The most striking observation that become apparent in table 5 is that non-native speakers appear to diverge from native speakers only with respect to a limited set of adjectives, in particular *different* and *difficult*. The results also confirm that across all languages, learners of English use the combination *completely* plus *different* significantly more compared with native speakers of English. Another collocation that is overused by non-native speakers having a Dutch, Polish, and Swedish background is *extremely difficult*. Non-native speakers also overuse *really* but the collocation differs across language backgrounds. While *really* co-occurs significantly more frequently than expected, given the native speaker data, with *hard* among speakers with Finnish and German language backgrounds, it is significantly overused with *good* by speakers with Polish and with *important* by speakers with Spanish language backgrounds.

The CFA has allowed us to find significant differences in the use of amplifier-adjective bigrams between native speakers and learners of English. To get very fine-grained understanding of differences between native speakers and learners of English, we now turn to the MuPDARF analysis. The initial random forest analysis used the native speaker data with *very* as the dependent variable and was fit with 3000 trees while considering two randomly selected variables at each split. The out-of-bag error rate was relatively high with 34.98 percent but still achieved an accuracy of 77.56 percent for the training set (70 percent of the data) and an accuracy of 76.53 for a test set (30 percent of the data). Also, initial random forest model performed significantly better than a base-line model which only predicted 57.42 percent of the data correctly and improved prediction accuracy by 34.5 percent. An overview of the variable importance of the random forest model fit to the native speaker data is provided in figure 4.

|  |
| --- |
| Figure 4: Variable importance plot of the random forest fit to the native speaker data with use of very as dependent variable. |

The prediction of random forest for learner data with *very* being the dependent variable only achieved an accuracy of 62.05 percent accuracy. However, this still represents a significant improvement compared to a base-line accuracy of 55.96 percent by a factor of 10.89 percent.

The model fitting process which included the additional variables *Language* and *NonNativeLike* (the dependent variable for the mixed-effects modelling) arrived at a final minimal model that performed significantly better than a base-line model (L.R. χ2: 230.95, DF 8, p < .001\*\*\*). As significant predictors, the final minimal model contains the syntactic context, priming, adjective type frequency, and emotionality as main effects and an interaction between syntactic function and frequency (see table 5). The model statistics, in particular the very low pseudo-R2, C, and Somers’ Dxy values, give reasons for concern and indicate a very poor, yet significant fit. The moderate model fit requires additional attention and is addressed in the discussion section.

Table 5: Results of the final minimal mixed-effects binomial logistic regression model of the MuPDARF procedure.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Random effects** | **Group(s)** | **Variance** | **Std. Dev.** | **L.R. χ2** | **DF** | **Pr(>|z|)** |
| Random Effect 1 | Language | 0.12 | 0.14 | 128.67 | 2 | p < .001\*\*\* |
| Random Effect 2 | Adjective | 0.22 | 0.61 | p < .001\*\*\* |
| **Fixed effects** | **Estimate** | **VIF** | **OddsRatio** | **Std. Error** | **z value** | **Pr(>|z|)** |
| (Intercept) | -0.84 |  | 0.43 | 0.22 | -3.76 | p < .001\*\*\* |
| Function:Predicative | 0.28 | 1.01 | 1.32 | 0.06 | 5.09 | p < .001\*\*\* |
| Priming:Prime | -0.65 | 1.00 | 0.52 | 0.11 | -5.90 | p < .001\*\*\* |
| Frequency | 25.5 | 1.48 | >10 | 4.14 | 6.17 | p < .001\*\*\* |
| Emotionality:NonEmotional | 0.17 | 1.68 | 1.19 | 0.08 | 2.25 | p = .0244\* |
| Emotionality:PositiveEmotional | 0.22 | 1.65 | 1.25 | 0.09 | 2.37 | p = .0179\* |
| Function:Predicative::Frequency | -15.4 | 1.47 | 0.00 | 5.10 | -3.01 | p = .0026\*\* |
| **Model statistics** |  |  |  |  |  | **Value** |
| Number of Groups |  |  |  |  |  | 12 |
| Number of cases in model |  |  |  |  |  | 6,446 |
| Observed successes |  |  |  |  |  | 3,997 |
| Residual deviance |  |  |  |  |  | 8,329.7 |
| R2 (Nagelkerke) |  |  |  |  |  | 0.056 |
| R2 (Hosmer and Lemeshow) |  |  |  |  |  | 0.032 |
| R2 (Cox and Snell) |  |  |  |  |  | 0.041 |
| C |  |  |  |  |  | 0.620 |
| Somers’ Dxy |  |  |  |  |  | 0.234 |
| AIC |  |  |  |  |  | 8,347.7 |
| BIC |  |  |  |  |  | 8,408.6 |
| Prediction accuracy |  |  |  |  |  | 63.22% |
| **Model Likelihood Ratio Test** |  | **L.R. χ2: 230.95** | | | **DF: 8** | **p < .001\*\*\*** |

In the following, the individual factors reported as significant by the regression model are visualized and briefly summarized. We begin with the intercept adjustments (random effects) of the regression model. Figure 5 shows that the intercept adjustments differ substantially across adjective types. This indicates that the proportion of nativelike choices by learners varies across adjective types with the most severe problems being associated with the adjective *different*. The most nativelike choices are made with respect to *important* and *difficult* while infrequent adjective types (*other*), *hard*, and *good* require relatively little adjustment.

|  |
| --- |
| Figure 5: Adjustment by adjective type. |

Figure 6 shows that the adjustments for language are remarkably homogenous with only minor adjustments being necessary. This indicates that the language background of speakers is substantially less important than the adjective that is amplified. With respect to the minor adjustment that are necessary, speakers with Czech, Polish, and Swedish language backgrounds differ most from native speakers while speakers with Spanish, Italian, and French language backgrounds make the most nativelike choices.

|  |
| --- |
| Figure 6: Adjustment by language background. |

With respect to fixed-effects, the model reports priming, the emotionality and frequency of the adjective, the syntactic context and an interaction between syntactic context and adjective frequency as significant predictors. Neither the semantic class of the adjective nor the proficiency of the speaker or any other interaction between predictors correlated significantly with the probability to make non-nativelike choices. Since adjective frequency and syntactic context are part of a significant interaction, their main effects will not be interpreted.

According to the regression results, learners make significantly more non-native like choices in non-primed contexts (see figure 7) as the predicted probabilities of making a non-nativelike choice is significantly lower in primed contexts. Also, learners are more likely to make non-nativelike choices if the adjective that is intensified is not associated with a core emotion compared with adjectives that are either associated with positive or negative core emotions (see figure 8).

|  |  |
| --- | --- |
| Figure 7: Predicted probability of non-nativelike choice by priming. | Figure 8: Predicted probability of non-nativelike choice against emotionality of adjective. |

According to figure 9, learners are only more likely to make non-nativelike choices in predicative contexts when the intensified adjective is relatively infrequent while the syntactic context does not appear to matter for frequent adjectives. Among frequent adjectives, the rate of non-nativelike amplification is very high regardless of the syntactic context. It is also important to note that the increase in non-nativelike choices levels off among high-frequency adjectives (frequency values above 0.4).

|  |
| --- |
| Figure 9: Predicted probability of non-nativelike choice against the frequency of amplified adjective by syntactic function. |

After reviewing the results of the statistical analysis, the following section discusses the findings in light of the relevant literature and problematizes potential issues of the statistical procedures.

1. **Discussion and Outlook**

The analysis presented here offer surprising and interesting insights. For instance, the results of the configural frequency analysis show that, when it comes to academic writing, learners have problems with a rather specific set of adjectives, in particular *different* and *difficult*, rather than across the board (see table 4). The same holds true for amplifiers: while learners overuse *completely*, *extremely*, and *really*, they rarely overuse the general, all-purpose amplifier *very* – notable exceptions being speakers with Bulgarian and Spanish language backgrounds who overuse *very* with *little* and *important* respectively. This finding is unexpected because it contrasts with previous research that suggested that learners significantly overuse *very* (Lorenz 1999: 198). While the findings presented here cannot substantiate overuse of *very*, all significant divergences that are detected by the configural frequency analysis represent instances of overuse, i.e. cases in which the learners used combinations of amplifiers and adjectives significantly more frequently than would be expected given the frequencies of these combinations among native speakers (see table 4). This means that learners have a significant tendency to overuse combinations compared to native speakers. This trend holds for all amplifier-adjective bigrams with the only exception being *very* cooccurring with infrequent adjectives among speakers with a Bulgarian language background.

What is striking about the significant divergences that the CFAs have detected is that they suggest that learners do not simply misapply amplifiers but that learners appear to either transfer usage patterns acquired in spoken conversation to academic discourse – as in the case of *really* – or overuse rather specific combinations (*extremely difficult*, *completely different*). Indeed, *really*, in particular, has been shown to occur predominantly in spoken rather than with written discourse (Biber et al. 2007: 565) and has replaced *very* as the dominant amplifier in informal spoken conversation across various geographically distinct varieties of English (cf. D’Arcy 2015 for NZE; Ito and Tagliamonte 2003 as well as Barnfield and Buchstaller 2010 for North East British English; Tagliamonte 2008 as well as Tagliamonte and Denis 2014 for Toronto English; and Tagliamonte and Denis 2014 for South Eastern Ontario English).

The MuPDARF results provide very detailed insights into which factors correlate with non-nativelike use of *very* among learners of English from diverse language backgrounds. Indeed, these results highlight what can be gained from using advanced multivariate statistical methods in learner corpus research (see also Gries 2018). The results of the MuPDARF show that the language background is substantially less important compared with the effect of adjective type. This finding supports recent research on adjective amplification which has shifted the focus from the amplifiers themselves to an approach which more thoroughly considers the interdependency of amplifiers and adjectives (see, e.g., Wagner 2017 or Hendrikx, Van Goethem, and Wulff 2019). In addition, the results of the MuPDARF procedure exemplify how learner corpus research may profit from including cognitive factors such as priming. While being unsurprising in itself, arguably because both learners and native speakers are similarly affected by priming, the finding that learners are significantly more likely to behave like native speakers in primed contexts confirms the need for cognitive factors to be integrated into language learning and language acquisition research. To elaborate, if priming had not been part of the statistical modelling, then this would have led to an overestimation of non-cognitive factors. This is so because the amount of variance that is explained by priming would could been attributed to confounding factors that are part of the statistical model.

Concerning the association of adjectives with core emotional states, the statistical analysis indicates that learners appear to have more severe issues when dealing with non-emotional adjectives compared with emotive adjectives. This finding extends previous research that shows that emotive adjectives are intensified more frequently than non-emotive adjectives (Boucher & Charles, 1969; Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). The present study indicates that emotive adjectives are not only intensified more frequently than non-emotive ones but that learners acquire amplifying strategies more readily for emotive adjectives. It seems plausible that learners are better at targeting nativelike intensification strategies when dealing with emotional adjectives because these are likely to be more cognitively salient than non-emotional adjectives are. This is relevant from a theoretical perspective because it expands previous research which showed that physical form, learner attention, and instructional focus, rather than cognitive-emotional salience, positively affected L2-acquisiton (see Cintrón-Valentín and Ellis 2016).

What is interesting about the interaction between adjective frequency and syntactic context is that the effect of syntactic context only applies to low frequency adjectives. As adjective frequency increases, the rate of non-nativelike use of *very* increases regardless of syntactic context before the effect of frequency wanes off among high-frequency adjectives. The positive correlation between non-nativelike language use and adjective frequency appears puzzling because learners should be better at nativelike production of elements that they encounter frequently. However, the opposite appears to hold true. While the view that frequency information is crucial in both first and second language acquisition has gained notable traction over the past two decades or so (see, for instance, Gass and Mackey 2002; Tomasello 2009; or Bybee 2003), it is important to understand what is meant by “frequency” in this context. Frequency should not be understood as raw frequency but rather as conditional probability, i.e., the occurrence of something given something else. This means that mere frequency of occurrence is insufficient because learners acquire linguistic structures by exploiting statistical cues in the input (see Ellis 2002). Thus, learners require additional cues for statistical learning to take effect (see Saffran et al. 1997; Saffran 2003). In the present case, the cue is clearly the adjective that is modified. With this in mind, what appears to be happening is that that the frequency effect is piggybacking on the effect of adjective type. The learners struggle with adjectives that neither infrequent nor highly frequent because in this mid-frequency-range, the learners overgeneralize to avoid acquiring errors. The strategy causes issues only among mid-frequency-range combinations because learners did not have sufficient input to fully acquire the systematicity underlying these combinations. Once the learner has had more exposure to input – as with high frequency combinations- then the error rate stagnates and eventually wanes off because the learner encountered a sufficiently large number of cues to fully acquire the systematicity of the variation.

With respect to methodology, the present study draws attention to the advantages of making use of advanced multifactorial statistics. While previous research on amplification in learner data offers very thorough and detailed qualitative insights, multivariate statistics allow us to unearth quantitative differences between learners and native speakers that have so far gone unnoticed. In elaboration of Gries (2018), the present paper exemplifies that learner corpus research can profit from carefully defining the control against which learner output is evaluated. The present study serves as a case study on why the concepts of over- and underuse can be misleading if not used in the sense of significant deviations from expected frequencies given the native speaker output (see Gries 2018).

Despite their advantages, certain issues of the statistical analysis require additional attention. The mixed-effects model has only little explanatory power and model fit criteria are below common cut-off values. Somers’Dxy, for example, ranges from -1 to 1 where values around 0 indicate pure chance. A value of merely 0.234 thus shows that the model presented here performs rather poorly. This is substantiated by a Harroll’s C value of 0.62. Harroll’s C rangers from 0 to 1 and it has been suggested that models have real predictive capacity with values above .8 (Baayen 2008: 204). However, while model fit is clearly suboptimal, the model still performs significantly better than chance (L.R. χ2: 230.95, DF 8, p < .001\*\*\*) and increase the prediction accuracy by a factor of 10.89 percent. An additional issue that requires discussion is the operationalization of frequency in the mixed-effects modelling. Frequency was operationalized as the polynomial to the second degree of the relative frequency. This was necessary because the effect of frequency was non-linear and exhibited curvature (see Wulff and Gries to appear). If frequency had not been transformed, then the mixed-effects model would have ignored the curvature in the effect of frequency. While the transformation was justified given the data at hand, the interpretation of the effect is more difficult and does not allow simple inferences of the form “if the frequency of an adjective increases x this leads to a decrease in the error rate of the learner by y”. Thus, the interpretation of the interaction between adjective frequency and syntactic function remains relatively coarse-grained – the improved fit to the data hence comes at the cost of the interpretability of the model. Also, it should be kept in mind that the corpus data used in the present study were collected about 20 years ago and future research would profit from using more recent data as well as data representing learners of non-European languages. A potential remedy for these- issues could be to use the more recent, second version of the ICLE (ICLEv2).

**References**

Adamson, S., & González-Díaz, V. 2004. Back to the very beginning: The development of intensifiers in Early Modern English. Paper presented at the *Thirteenth International Conference on English Historical Linguistics*, University of Vienna, 23-28 August 2004.

Althaus, S. L., & Kim, Y. M. 2006. “Priming Effects in Complex Information Environments: Reassessing the Impact of News Discourse on Presidential Approval”, *Journal of Politics* 68(4), 960-976.

Baayen, R. H. 2008. *Analyzing Linguistic Data. A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press.

Barnfield, K., & Buchstaller, I. 2010. “Intensifiers on Tyneside: Longitudinal developments and new trends”, *English World-Wide* 31(3), 252-287.

Bauer, L., & Bauer, W. 2002. “Adjective boosters in the English of Young New Zealanders”, *Journal of English Linguistics* 30(3), 244–257.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C. & Vohs, K. D. 2001. „Bad Is Stronger than Good”, *Review of General Psychology* 5(4), 323–370.

Biber, D., Johansson, S., Leech, G. Conrad, S., & Finegan, E. 2007. *Longman grammar of spoken and written English*. London: Longman.

Bolinger, D. 1972. *Degree words*. The Hague: Mouton.

Bortz, J., Lienert, G. A., & Boehnke, K. 2008. *Verteilungsfreie Methoden in der Biostatistik*. Heidelberg: Springer.

Boucher, J. & Charles, E. O. 1969. “The Pollyana Hypothesis”, *Journal of verbal learning and verbal behavior* 8, 1–8.

Breban, T. & Davidse, K. 2016. “The history of very: the directionality of functional shift and (inter)subjectification”, *English Language and Linguistics* 20(2), 221–249.

Bybee, J. 2003. *Phonology and Language Use*. Cambridge: Cambridge University Press.

Cintrón-Valentín, M. C. & Ellis, N. C. 2016. “Salience in Second Language Acquisition: Physical Form, Learner Attention, and Instructional Focus”, *Frontiers in Psychology* 7, 1284.

D’Arcy, A. F. 2015. “Stability, stasis and change – the longue durée of intensification”, *Diachronica* 32(4), 449–493.

Dixon, R. M. W. 1977. “Where have all the adjectives gone?”, *Studies in Language* 1, 19-80.

Dixon, R. M. W. 2004. “Adjective classes in typological perspective”. In R. M. W. Dixon & A. Y. Aikhenvald (Eds.), *Adjective classes. A cross-linguistic typology*. Oxford: Oxford University Press, 1–49.

Edmonds, A., & Gudmestad, A. 2014. “Your Participation Is greatly/highly Appreciated: Amplifier Collocations in L2 English”, *The Canadian Modern Language Review*/*La Revue canadienne des* *langues vivantes* 70(1), 76–102.

Ellis, N. C. 2002. “Frequency Effects in Language Processing”, *Studies in Second Language* *Acquisition* 24(2): 143–188.

Fallas Escobar, C., & Chaves Fernández, L. 2017. “EFL Learners’ Development of Voice in Academic Writing: Lexical Bundles, Boosters/Hedges and Stance-Taking Strategies”, *Education and Learning Research Journal* 15, 96-124.

Forsberg, F. 2010. “Using conventional sequences in L2 French”, *International Review of Applied* *Linguistics in Language Teaching* 48(1), 25–51.

Friginal, E., Lee, J. J., Polat, B., & Roberson, A. 2017. *Exploring Spoken English Learner Language* *Using Corpora. Learner talk*. Cham: Palgrave Macmillan.

Fuchs, R. & Gut, U. 2016. “Register variation in intensifier usage across Asian Englishes”. In H. Pichler (Ed.), *Discourse-Pragmatic Variation and Change: Insights from English*. Cambridge: Cambridge University Press, 185–213.

Fuchs, R. 2017. “Do women (still) use more intensifiers than men? Recent change in the sociolinguistics of intensifiers in British English”, *International Journal of Corpus Linguistics* 22(3), 345–374.

Gass, S. M. & Mackey, A. 2002. “Frequency Effects and Second Language Acquisition”, *Studies in* *Second Language Acquisition* 24(2), 249–260.

Granger, S. 1993. “The International Corpus of Learner English”. In J. Aarts, P. de Haan, & N. Oostdijk (Eds.), *English Language Corpora: Design, Analysis and Exploitation*. Amsterdam: Rodopi, 57-69.

Granger, S. 1998. “Prefabricated patterns in advanced EFL writing: Collocations and formulae”. In A. P. Cowie (Ed.), *Phraseology: Theory, analysis, and applications*. Oxford, UK: Clarendon, 145–160.

Gries, S. T. 2018. “On over- and underuse in learner corpus research and multifactoriality in corpus linguistics more generally”, *Journal of Second Language Studies* 1(2), 277–309.

Gries, S. T. & Adelman, A. S. 2014. “Subject realization in Japanese conversation by native and non-native speakers: exemplifying a new paradigm for learner corpus research”. In J. Romero-Trillo (Ed.), *Yearbook of Corpus Linguistics and Pragmatics 2014: New empirical and theoretical paradigms*. Cham: Springer, 35-54.

Gries, S. T. & Deshors, S. C. 2014. “Using regressions to explore deviations between corpus data and a standard/target: two suggestions”. *Corpora* 9(1), 109-136.

Heller, B., Bernaisch, T., & Gries, S. T. 2017. “Empirical perspectives on two potential epicenters: The genitive alternation in Asian Englishes”. *ICAME Journal* 41, 111-144.

Hendrikx, I., Van Goethem, K., & Wulff, S. 2019. “Intensifying constructions in French speaking L2 learners of English and Dutch. Cross-linguistic influence and exposure effects”. *International Journal of Learner Corpus Research* 5(1), 63–103.

Hinkel, E. 2003. “Adverbial markers and tone in L1 and L2 students’ writing”, *Journal of Pragmatics* 35(7), 1049-1068.

Ito, R. & Tagliamonte, S. 2003. “Well weird, right dodgy, very strange, really cool: Layering and recycling in English intensifiers”, *Language in Society* 32, 257–279.

Krauth, J., & Lienert, G. A. 1973. *Die Konfigurationsfrequenzanalyse und ihre Anwendung in* *Psychologie und Medizin*. Freiburg: Alber.

Lorenz, G. R. 1998. “Overstatement in advanced learners’ writing: stylistic aspects of adjective intensification”. In S. Granger (Ed.), *Learner English on computer*. London: Longman, 53-66.

Lorenz, G. R. 1999. *Adjective Intensification – Learners Versus Native Speakers: A Corpus Study of Argumentative Writing.* Amsterdam: Rudopi.

Lorenz, G. R. 2002. “Really worthwhile or not really significant: A Corpus-based Approach to the Delexicalisation and Grammaticalisation of Adverbial Intensifiers in Modem English”. In I. Wischer & G. Diewald (Eds.), *New Reflections on Grammaticalization*. Amsterdam: John Benjamins, 143-161.

Maddeaux, R., & Dinkin, A. 2017. “Is like like like?: Evaluating the same variant across multiple variables”, *Linguistics Vanguard*, 3(1). Available at: https://www.degruyter.com/view/j/lingvan.2017.3.issue-1/lingvan-2015-0032/lingvan-2015-0032.xml?format=INT (accessed January 2019).

Macaulay, R. 2002. “Extremely interesting, very interesting, or only quite interesting: Adverbs and social class”, *Journal of Sociolinguistics* 6(3), 398–417.

Méndez-Naya, B. 2003. “On intensifiers and grammaticalization: The case of SWIÞE”, *English Studies* 84(4), 372–391.

Méndez-Naya, B. 2008. “On the history of downright”, *English Language and Linguistics* 12(2), 267-287.

Méndez-Naya, B. & Pahta, P. 2010. “Intensifiers in competition: The picture from early English medical writing”. In I. Taavitsainen & P. Pahta (Eds.), *Early Modern English medical texts: Corpus* *description and studies*. Amsterdam: John Benjamins, 191–214.

Mohammad, S. M., & Turney, P. D. 2013. “Crowd sourcing a word-emotion association lexicon”, *Computational Intelligence* 29(3), 436-465.

Nevalainen, T. 2008. “Social variation in intensifier use: Constraint on -ly adverbialization in the past?”, *English Language and Linguistics* 12(2), 289–315.

Nevalainen, T. & Rissanen, M. 2002. “Fairly pretty or pretty fair? On the development and grammaticalization of English downtoners”, *Language Sciences* 24, 359–380.

Pertejo, P. N. & Martínez, I. M. P. 2014. “That’s absolutely crap, totally rubbish. The use of intensifiers absolutely and totally in the spoken language of British adults and teenagers”, *Functions* *of Language* 21(2), 210-237.

Martínez, I. M. P., & Pertejo, P. N. 2012. „He’s absolutely massive. It’s a super day. Madonna, she is a wicked singer. Youth language and intensification: A corpus-based study”, *Text and Talk* 32(6). 773–796.

Paradis, C. 2008. “Configurations, construals and change: Expressions of DEGREE”, *English Language and Linguistics* 12(2), 317-343.

Partington, A. 1993. “Corpus evidence of language change: The case of intensifiers”. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In honour of John Sinclair*. Amsterdam: John Benjamins, 177-192.

Peters, H. 1992. “English boosters: Some synchronic and diachronic aspects”. In G. Kellermann & M. D. Morrissey (Eds.), *Diachrony within synchrony: Language history and cognition*. Frankfurt: Peter Lang, 529–545.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. 1985. *A Comprehensive Grammar of the English Language*. London and New York: Longman.

Rissanen, M. 2008. “From ‘quickly’ to ‘fairly’: On the history of *rather*”, *English Language and Linguistics* 12(2), 345–359.

Saffran, J. R. 2003. “Statistical language learning: mechanisms and constraints”, *Current Directions in Psychological Science* 12(4), 110–114.

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. 1997. “Incidental Language Learning: Listening (and Learning) Out of the Corner of Your Ear”, *Psychological Science* 8(2), 101–105.

Stenström, A.-B. 1999. “He was really gormless – she’s bloody crap: Girls, boys and intensifiers”. In H. Hasselgård & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson*. Amsterdam: Rodopi, 69-78.

Szmrecsanyi, B. 2005. “Language Users as Creatures of Habit: A corpus-based Analysis of Persistence in Spoken English”, *Corpus Linguistics and Linguistic Theory* 1(1), 113-150.

Szmrecsanyi, B. 2006. *Morphosyntactic Persistence in Spoken English: A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis*. Berlin & New York: Walter de Gruyter.

Tagliamonte, S. 2008. “So different and pretty cool! Recycling intensifiers in Toronto, Canada”, *English Language and Linguistics* 12(2), 361-394.

Tagliamonte, S. & Denis, D. 2014. “Expanding the transmission/diffusion dichotomy: Evidence from Canada”, *Language* 90(1), 90-136.

Tagliamonte, S. & Roberts, C. 2005. “So weird; so cool; so innovative: The use of intensifiers in the television series Friends”, *American Speech* 80(3), 280-300.

Tomasello, M. 2009. “The Usage-based Theory of Language Acquisition”. In E. L. Bavin (Ed.), *The Cambridge Handbook of Child Language*. Cambridge: Cambridge University Press, 69–87.

Tulving, E., & Schacter, D. L. 1990. “Priming and Human Memory Systems”, *Science* 247(4940), 301-306.

Tweedie, F. J. & Baayen, H. R. 1998. „How Variable May a Constant Be? Measures of Lexical Richness in Perspective”, *Computers and the Humanities* 32(5), 323–352.

Wagner, S. 2017. “*Totally new* and *pretty awesome*: Amplifier-adjective bigrams in GloWbE”, *Lingua* 200, 63-83.

Wallenberg, J. C. 2013. A unified theory of stable variation, syntactic optionality, and syntactic change. Paper presented at *15th Diachronic Generative Syntax Conference*, University of Ottawa, 1-3 August 2013.

Wulff, S., & Gries, S. T. to appear. “Exploring individual variation in learner corpus research: some methodological suggestions”. In B. S. W. LeBruyn & M. Paquot (Eds.). *Learner corpus research and second language acquisition*. Cambridge: Cambridge University Press.

Yule, G. U. 1944. *The Statistical Study of Literary Vocabulary*. Cambridge: Cambridge University Press.

Zuur, A. F., Ieno, E. N., & Elphick, C. S. 2010. “A protocol for data exploration to avoid common statistical problems2, *Methods in Ecology and Evolution* 1(1), 3-14.

**Software**

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. 2018. “quanteda: An R package for the quantitative analysis of textual data”, *Journal of Open Source Software* 3(30), 774.

Hornik, K. 2016. “*openNLP: Apache OpenNLP Tools Interface. Version 0.2-6”*. Available at: https://cran.r-project.org/web/packages/openNLP/openNLP.pdf (accessed January 2019).

Jockers, Matthew. 2017. “*syuzhet: Extracts Sentiment and Sentiment-Derived Plot Arcs from Text*. Version 1.0.1”. Available at: https://github.com/mjockers/syuzhet (accessed January 2019).

R Core Team. 2008. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

**Corpora**

Granger, S. 2002. *The International Corpus of Learner English (ICLE)*. Centre for English Corpus Linguistics, Université Catholique de Louvain.

The *Louvain Corpus of Native English Essays (LOCNESS)*. Centre for English Corpus Linguistics, Université Catholique de Louvain. Available at: https://uclouvain.be/en/research-institutes/ilc/cecl/locness.html (accessed January 2019).

1. In examples, the code ICLE stands for the corpus (ICLE = International Corpus of Learner English). The next two capital letters represent a country or language code (BG = Bulgaria, CZ = Czech Republic, DB = Dutch (Belgium), DN = Dutch (Netherlands), FI = Finland, FR = France, GE = Germany, IT = Italia, PO = Poland, RU = Russia, SP = Spain, SW = Swedish). The next three capital letters refer to the city in which the data was collected, for example LND for Lund, NIJ for Nijmegen, POZ for Poznan, etc. The numbers at the end of the code identify the essay and the author of the essay. [↑](#endnote-ref-1)