

Idiomatic or literal? Production of collocations in German read speech

Frank Zimmerer, Erika Brandt, Bistra Andreeva, Bernd Möbius

Saarland University, Saarbrücken, Germany

zimmerer@coli.uni-saarland.de

Abstract

Collocations have been identified as an interesting field to study the effects of frequency of occurrence in language and speech. We report results of a production experiment including a duration analysis based on the production of German collocations. The collocations occurred in a condition where the phrase was produced with a literal meaning and in another condition where it was idiomatic. A durational difference was found for the collocations, which were reduced in the idiomatic condition. This difference was also observed for the function word *und* ('and') in collocations like *Mord und Totschlag* ('murder and manslaughter'). However, an analysis of the vowel /ʊ/ of the function word did not show a durational difference. Some explanations as to why speakers showed different patterns of reduction (not all collocations were produced with a shorter duration in the idiomatic condition by all speakers) and why not all speakers use the durational cue (one out of eight speakers produced the conditions identically) are proposed.

Index Terms: collocations, idiomatic expressions, speech production.

1. Introduction

In recent years, an increasing amount of evidence has been presented suggesting that usage frequency has an important impact on several aspects of language and speech. These aspects include sound change, pronunciation variation, phonetic reduction, or the perception of these patterns (e.g., [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18]).

Syntactic and morphological co-occurrence have been identified as pivotal factors in predicting reaction times when modeling human performance computationally [2]. A study of three reduction phenomena investigating various factors, including different frequency measures, found that both prior probability (frequency of occurrence) and conditional probabilities (contextual predictability) affected temporal reduction: high-frequency and highly predictable items tend to be shortened [9].

Similarly, a stronger degree of palatalization has been reported for /d/ across word boundaries in cases like *did you* compared to items like *good you*, where the first /d/ in *did* has a higher conditional probability than the /d/ in *good you* [4]. There is additional evidence of usage frequency, indicating that contractions in English are frequency-dependent [12].

Collocations and idiomatic expressions can be considered as a special case of words co-occurring with a higher frequency than expected by chance. They have received considerable interest by various researchers (e.g., [3, 19, 20, 6, 21, 22, 23, 9, 24, 16, 25, 26]). Collocations and idiomatic expressions exhibit characteristics that make them a prime field of study for the effects of frequency of usage on speech production and perception.

Collocations are difficult to model but crucial for successful natural language processing systems [26]. They show patterns that make them stand out compared to the standard usage of

words. This special status affects their usage in written text and speech.

For example, Bybee and Scheibman [6] investigated different realizations of *don't*. Their results suggest that speakers reduce *don't* more if it occurs in frequent phrases (e.g., *I don't know*, or *I don't think*) compared to *don't* if it occurs in less frequent phrases (e.g., *don't assume I'm guilty*). These results have been interpreted as evidence for assuming a lexical storage of phrases [10].

Lieberman [16], using fast read speech material, investigated the production and perception of what he called *redundant* and *non-redundant* words in phrases, i.e., whether a word can be predicted from its context or not. The results suggest that, on the one hand, speakers produced redundant words with a different amplitude pattern and, on the other hand, intelligibility was inversely related to predictability.

One of the main characteristics that sets idiomatic expressions apart from the literal use of words with high co-occurrence probability, is the degree to which they can be interpreted literally. Idioms can be defined as fixed expressions that do not have a literal meaning, or whose literal meaning is not necessarily compositional. Consider an idiom like *bread and butter* in a sentence like *Acoustic analyses are the bread and butter of phoneticians*: the sentence does not refer to actual bread or butter. This factor has also been found to have an impact on pronunciation.

For instance, Hay and Bresnan [10], investigating the phrase *giving a hand*, showed that the /æ/ of *hand* was affected by sound change to a smaller extent when the phrase was used more idiomatically, whereas the vowel /ɪ/ in *give* was more advanced with respect to sound change when the phrase was used to refer to abstract themes compared to its use expressing a transfer of possession. The authors argue that this can also be seen as an argument for the storage of units larger than the word, as did Schweitzer and colleagues [27]—but see [9] for results that suggest a more complex picture for collocation storage, where semantic relation needs to be incorporated too.

Van Lancker and colleagues [13] investigated ditropically ambiguous sentences in perception and production, where sentences like *She was to keep a stiff upper lip* could have an idiomatic or a literal interpretation. If speakers produced the sentences under normal conditions, listeners were not able to distinguish the two meanings above chance level. However, when the speakers were asked to make the difference clearly, listeners were able well above chance level to perceive the intended meaning, i.e. whether the collocation was intended to have a literal or an idiomatic interpretation. When analyzing the production of these ditropically ambiguous sentences, Van Lancker and colleagues [14] found various prosodic cues, such as sentence duration, pausing, or word duration, as well as some segmental differences, that differentiated literal from idiomatic meanings of sentences. Their data suggest that the literal meaning can be expected to be produced more clearly and with more emphasis on the respective constituents. The authors also found

that different strategies were used by different speakers.

The use of segmental and prosodic cues has been shown to affect foreign (or second) language learners, too [15]. Native speakers of American English were able to perceive the differences between the idiomatic and the literal meaning of English utterances best, followed by native speakers of other varieties of English, who in turn were better than fluent non-native speakers of English. Learners of English as a second language were not able to distinguish the differences above chance level. These results are in line with an earlier study by Cronk and Schweigert, who reported evidence that familiarity with collocations is crucial for their processing [21].

This overview of the literature suggests that speakers can be expected to produce differences between collocations and literal meanings, at least in some conditions. Individual speakers may show different patterns in producing these differences. However, the cues differentiating between the two meanings are not very strong, as indicated by some difficulties for non-native speakers to actually learn these patterns.

We aimed at replicating and extending some of these findings with respect to the production of collocations and idiomatic expressions. To this end, we selected collocations which show different degrees of idiomatic use, to investigate the following research questions: Do speakers produce differences between co-occurring words with literal or idiomatic meaning, even if they are not explicitly instructed to make the difference?

2. Method and Data

2.1. Materials

For the production experiment, 20 collocations were selected, which could be used in both an idiomatic expression and in a literal reading. Although all the expressions could be used literally and idiomatically, there was a considerable difference with respect to their status as idiomatic expression, that is, some of the combinations can be considered to be used almost exclusively in an idiomatic expression (e.g., *Kraut und Rüben* – idiomatic: ‘pell-mell’ – literally: ‘cabbage and beets’), whereas other collocations are overwhelmingly used literally (e.g., *Stadt und Land* – ‘city and country’).

All the 20 collocations had the structure ‘X and Y’ (e.g., *Mord und Totschlag* – idiomatic: ‘blood and thunder’ – literally: ‘murder and manslaughter’) – where X and Y were either nouns (15 pairs) or adjectives (5 pairs).

For each of these 20 collocations, a preceding and a following sentence was created to ensure a plausible context which made either the literal or idiomatic reading the only possibility. The target collocations always occurred in the second sentence. We ensured that the sentence in which the collocations occurred had the same number of syllables in both conditions (mean number of syllables in the second sentence: 14.65 syllables, range 10–27 syllables, SD 4.37).

Furthermore, all collocations occurred in final position of the sentences in both conditions. Additionally, 20 three-sentence texts were constructed that were used as fillers. The fillers were similar in their structure to the target items, but they did not include any high frequency collocations. Of these 60 three-sentence texts (20 idiomatic, 20 literal, 20 fillers) two lists of 30 three-sentence texts were created. In each list, every sentence with a target collocation occurred once, all conditions occurred 10 times per list.

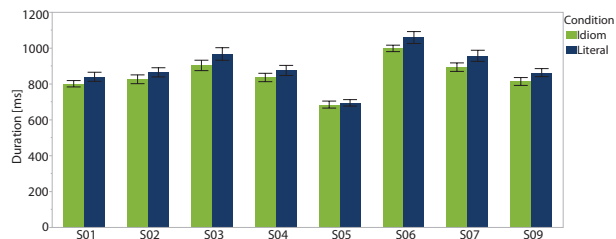


Figure 1: Duration (in ms) of the collocations by condition for the eight speakers.

2.2. Recordings

Recordings were made in a sound proof room. Participants stood in front of a computer screen, from which they read the sentences into a microphone. The experimental lists were presented with a PowerPoint presentation, and after reading a three-sentence text, participants clicked to see the next text. There was no time pressure. If participants made an error, they could repeat the sentence as often as they liked. The experimenter monitored the recordings in an adjacent room to adjust the recording level if necessary. Recordings were made on a desktop computer with the sampling rate set to 48 kHz.

For the duration analysis, the medial sentences where the collocations occurred were excised with PRAAT [28]. Subsequently, in a first step, automatic segmentation and labelling of the sentences was performed with WebMAUS [29]. In a second step, manual correction of the words of the collocation was performed by a trained phonetician. The statistical analyses were calculated with JMP [30].

2.3. Participants

In total, 8 female students between 19 and 30 years of age (mean 24.75; all native speakers of German) from Saarland University in Saarbrücken participated in this experiment. Only female speakers were chosen to avoid a confounding influence of the factor gender. They received monetary compensation for their participation and gave written consent to participate in this study. They were naïve with respect to the goal of the experiment.

3. Results

3.1. Duration analysis

We analyzed the duration of the collocations to investigate whether the speakers produced differences in duration depending on whether the collocation was used in the literal, or in the idiomatic meaning. Duration cues were previously found to differentiate between these two meanings (e.g., [14]), at least when speakers were explicitly asked to make this difference. We were interested in the extent to which our speakers use durational cues even in the absence of explicit instructions.

Of the 320 items (8 speakers * 20 collocations * 2 conditions), 4 items had to be excluded due to reading errors (for 3 items, the number of syllables in the two conditions differed) and to a technical problem with the recording (1 item). All counterparts of the same collocation pair by that speaker were also excluded to ensure a balanced comparison. All other items (312) were subsequently analyzed.

Overall, the idiomatic expressions had a mean duration of 845 ms (SD 130), whereas the collocations in the literal condi-

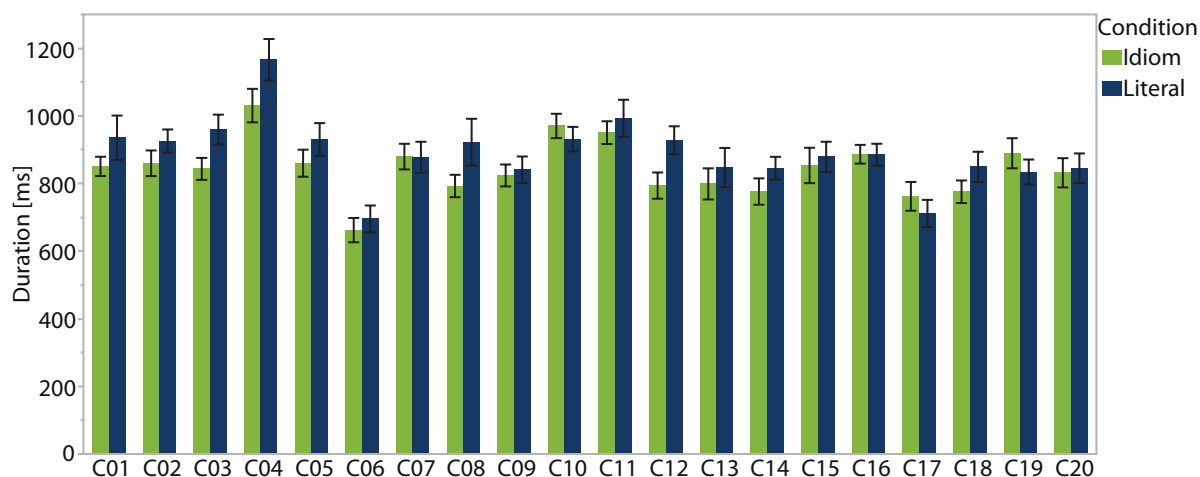


Figure 2: Duration of the different collocation pairs, idiomatic meaning in green, literal meaning in blue.

tion had a mean duration of 890 ms (SD 157). Figure 1 illustrates this pattern, viz. longer duration for collocations in the literal meaning, which can be observed for all speakers. However, the extent to which speakers differed with respect to the two meanings is rather different (e.g., Speaker 05 (S05) vs. S06).

With respect to the different collocations, out of the 20 collocation pairs, 15 were produced with a shorter duration for the idiomatic meaning, in 5 cases the collocation with the literal meaning showed a shorter duration. This pattern is depicted in Figure 2.

Subsequently, we calculated a linear mixed model with SPEAKER as random factor, COLLOCATION DURATION as dependent factor, as well as COLLOCATION and CONDITION (*literal, idiomatic*) and their interaction as independent factors. The results indicated that COLLOCATION ($F(19,265)=24.4$, $p<0.001$) and CONDITION ($F(1,265)=44$, $p<0.001$) were significant factors, as was their interaction ($F(19,265)=2.89$, $p<0.001$). This showed that, overall, the speakers produced the collocations significantly shorter, when they occurred in the idiomatic condition. However, because not all of the collocations were produced with this pattern, we observed the interaction.

A second analysis was performed on the duration of *und* 'and' of all collocations. The results indicate that COLLOCATION ($F(19,265)=3.38$, $p<0.001$) and CONDITION ($F(1,265)=6.18$, $p<0.01$) are significant factors, but the interaction was not. Function words are most likely to be reduced, and anecdotally, collocations like *bread and butter* or *rock'n'roll* show this pattern arguably even in a lexicalized way when produced idiomatically. In the idiomatic condition, *und* had a mean duration of 140 ms (SD 34), whereas in the collocations produced literally *und* had a mean duration of 149 ms (SD 37). As can be seen in Figure 3, apart from Speaker 05 (S05) who produced *und* in both conditions with the same mean duration, all speakers produced *und* with a shorter mean duration in the idiomatic condition.

After examining the complete word, the vowel /ʊ/ in the function word was analyzed separately. The mean duration of the vowel in the idiomatic use was 49.2 ms, whereas the vowel in the literal condition had a mean duration of 48.9 ms. If we split up the data by speaker, we see in Figure 4 that 4 out of 8 speakers produced the vowel with a longer mean duration in the idiomatic compared to the literal condition, three speakers

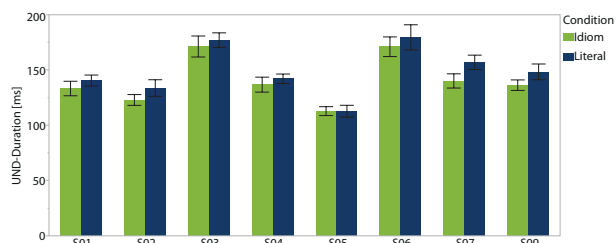


Figure 3: Duration of the function word 'und', the second word of the collocations for the eight speakers, by condition in ms.

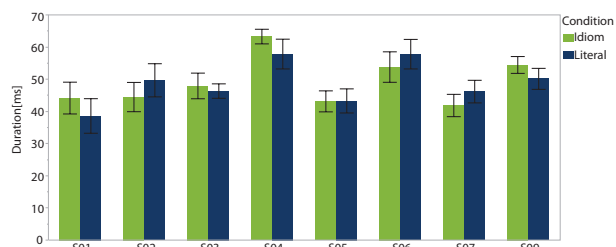


Figure 4: Duration of the vowel in the function word 'und' for the eight speakers, by condition in ms.

showed the opposite tendency, and one speaker (S05) produced the vowels with the same mean duration in both conditions.

A final analysis was carried out to investigate the difference of the duration of the vowel /ʊ/ in the function word. The linear mixed model included DURATION of vowel /ʊ/ as dependent variable, SPEAKER as random factor as well as COLLOCATION and CONDITION (*idiom, literal*) and their interaction as fixed factors. The results indicate that there was a significant difference between the COLLOCATIONS ($F(19,263)=2.18$, $p<0.01$), but neither CONDITION nor the interaction were significant.

4. Discussion and Conclusions

The results of the study presented here indicate that, overall, speakers seem to produce collocations with their literal meaning or as idiomatic expressions with different durations. Duration

has been identified as one of the possible cues used by speakers to differentiate collocations in their literal and idiomatic use (e.g., [14]). We conclude that speakers tend to reduce words in collocations if they occur in idiomatic expressions.

The instructions to the participants of the experiment made no mention of this difference, indeed, no mention of ‘collocation’ was made at all, to keep the speakers uninformed about the subject of the productions, in contrast to [13, 14, 15]. In an informal interview after the experiment, none of the participants were aware of the purpose of the sentence triplets.

This point is crucial, since one of our research questions aimed to find out whether speakers produce differences if they are not explicitly instructed to do so. As can be seen in the results section, this difference was not huge, and not every speaker produced the difference. However, apart from one speaker (S05), all participants produced the durational difference. Overall, the significant tendency of reduction for idioms, which was also found in [14], was replicated, even with this rather uncontrolled set of collocations.

The same pattern can also be seen in the duration of the function word *und*. Overall, speakers reduced the duration of *und* in the idiomatic condition. Despite the small numeric difference in the conditions, this difference proved to be robust enough to reach significance.

Whereas the overall mean duration for the idiomatic condition was smaller, there was one speaker (again, S05) who did not produce this difference.

The duration analysis of the vowel /*ʊ*/ in the function word *und*, however, does not show this pattern. This finding can be seen as somewhat surprising, because the duration of vowels is comparatively long and would be a good candidate for reduction. After all, grammaticalized reductions occur in idiomatic expressions like *bread and butter*. Reduction, however, could also be produced with respect to the formant targets of vowels. Thus, in subsequent studies it will be important to also analyze formant structures of vowels in idiomatic and literal collocations.

On a general level, there are further possible explanations as to why speakers were producing only small differences. In our experiment, speakers read the items at their ‘normal’, or comfortable rate (for instance in contrast to [16]). There are two possible factors that may have diminished stable cues for differentiating between idiomatic and literal meanings. First, read speech has been shown in many cases to result in less reduction than conversational speech [31, 32] and, second, fast speech may have also led to more reduction than the comfortable speed in this experiment. Furthermore, the context of the three sentence items, which was constructed to differentiate between the meanings in the first place, could have made speakers reluctant to additionally emphasize the meaning in which the collocation was produced.

The results reported here also hint at the necessity to investigate other cues, such as F0 contours, or the use of creaky voice [14]. Because speakers differ with respect to their strategies, it is also important to investigate the extent to which durational cues were used by individual speakers in combination with other cues and to analyze possible trade-off relations.

Speaker S05 produced the collocations with the highest speech rate, and in her productions, no difference was found. One possible explanation for her behavior is the speech rate itself. If the speech rate is (too) high, a further increase in speech rate could lead to reduction effects which are not characteristic for read speech. Therefore, the lack of durational differences

could also be, at least in part, explained by the task of the experiment.

Nonetheless, the results reported here are promising. Even with the relatively small sample size of eight female speakers, some trends in duration differences could be observed (for 7 out of 8 speakers). Future studies might investigate the effect of the task, that is, whether different speaking styles, or different contexts have an effect on the patterns [32]. Furthermore, experiments in which the collocation are more tightly controlled [13, 14, 15] may be important. Finally, more speakers are needed to investigate the individual strategies that are used by various speakers, especially if several cues are investigated and compared at the same time.

To conclude, the extent to which listeners are able to use the cues speakers produce to differentiate the two meanings is crucial. Because the differences are rather small, it is not obvious whether listeners are able to exploit these differences to discriminate collocation readings perceptually.

5. Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG) as part of SFB 1102 ‘Information Density and Linguistic Encoding’ at Saarland University.

6. References

- [1] M. Aylett and A. E. Turk, “The Smooth Signal Redundancy Hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech,” *Language and Speech*, vol. 47, no. 1, pp. 31–56, 2004.
- [2] R. H. Baayen, “Demythologizing the word frequency effect: A discriminative learning perspective,” *The Mental Lexicon*, vol. 5, pp. 436–446, 2010.
- [3] A. Bell, D. Jurafsky, E. Fosler-Lussier, C. Girand, M. Gregory, and D. Gildea, “Effects of disfluencies, predictability, and utterance position on word form variation in English conversation,” *Journal of the Acoustical Society of America*, vol. 113, no. 2, pp. 1001–1024, 2003.
- [4] N. Bush, *Frequency effects and word-boundary palatization in English*. Amsterdam: John Benjamins, 2001, pp. 255–280.
- [5] J. Bybee, “Phonological evidence for exemplar storage of multiword sequences,” *Studies in Second Language Acquisition*, vol. 24, no. 2, pp. 215–221, 2002.
- [6] J. Bybee and J. Scheibmann, “The effect of usage on degrees of constituency: the reduction of don’t in English,” *Linguistics*, vol. 37, no. 4, pp. 575–596, 1999.
- [7] U. Cohen Priva, “Informativity affects consonant duration and deletion rates,” *Laboratory Phonology*, vol. 6, no. 2, pp. 243–278, 2015.
- [8] S. Gahl, “‘Thyme’ and ‘time’ are not homophones. The effect of lemma frequency on word durations in spontaneous speech,” *Language*, vol. 84, no. 3, pp. 474–496, 2008.
- [9] M. L. Gregory, W. D. Raymond, A. Bell, E. Fosler-Lussier, and D. Jurafsky, “The effects of collocational strength and contextual predictability in lexical production,” in *Proceedings of the 35th meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 1999, Book Section, pp. 151–166.
- [10] J. Hay and J. Bresnan, “Spoken syntax: The phonetics of giving a hand in New Zealand English,” *The Linguistic Review*, vol. 23, pp. 321–349, 2006.
- [11] F. T. Jaeger and E. Buz, “Signal reduction and linguistic encoding,” in *Handbook of Psycholinguistics*, E. M. Fernández and H. S. Cairns, Eds. Oxford, United Kingdom: Wiley-Blackwell, accepted.

- [12] M. Krug, "String frequency: A cognitive motivating factor in coalescence, language processing, and linguistic change," *Journal of English Linguistics*, vol. 26, pp. 286–320, 1998.
- [13] D. Van Lancker and G. J. Canter, "Idiomatic versus literal interpretations of ditropically ambiguous sentences," *Journal of Speech and Hearing Research*, vol. 24, pp. 64–69, 1981.
- [14] D. Van Lancker, G. J. Canter, and D. Terbeek, "Disambiguation of ditropic sentences: Acoustic and phonetic cues," *Journal of Speech and Hearing Research*, vol. 24, pp. 330–335, 1981.
- [15] D. Van Lancker-Sidtis, "Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one," *Applied Psycholinguistics*, vol. 24, pp. 45–57, 2003.
- [16] P. Lieberman, "Some effects of semantic and grammatical context on the production and perception of speech," *Language and Speech*, vol. 6, no. 3, pp. 172–187, 1963.
- [17] E. Schulz, Y. M. Oh, Z. Malisz, B. Andreeva, and B. Möbius, "Impact of prosodic structure and information density on vowel space size," in *Proceedings of Speech Prosody 2016*, Boston (MA), 2016, Conference Proceedings, pp. 350–354.
- [18] Y. Xu and D. Reitter, "Information density converges in dialogue: Towards an information-theoretic model," *Cognition*, vol. 170, pp. 147–163, 2018. [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed/29024915>
- [19] S. Benor and R. Levy, "The chicken or the egg? A probabilistic analysis of English binomials," *Language*, vol. 82, no. 2, pp. 233–278, 2006.
- [20] V. Brezina, T. McEnery, and S. Wattam, "Collocations in context: A new perspective on collocation networks," *International Journal of Corpus Linguistics*, vol. 20, no. 2, pp. 139–173, 2015.
- [21] B. C. Cronk and W. A. Schweigert, "The comprehension of idioms: The effects of familiarity, literalness, and usage," *Applied Psycholinguistics*, vol. 13, no. 2, pp. 131–146, 1992.
- [22] P. Durrant and A. Doherty, "Are high-frequency collocations psychologically real? investigating the thesis of collocational priming," *Corpus Linguistics and Linguistic Theory*, vol. 6, no. 2, pp. 125–155, 2010.
- [23] S. Evert, "The statistics of word cooccurrences word pairs and collocations," PhD Thesis, Universität Stuttgart, 2005.
- [24] S. T. Gries, "50-something years of work on collocations: What is or should be next ..." *International Journal of Corpus Linguistics*, vol. 18, no. 1, pp. 137–166, 2013.
- [25] L. Michelbacher, S. Evert, and H. Schütze, "Asymmetry in corpus-derived and human word associations," *Corpus Linguistics and Linguistic Theory*, vol. 7, no. 2, 2011.
- [26] V. Seretan, "On collocations and their interaction with parsing and translation," *Informatics*, vol. 1, no. 1, pp. 11–31, 2013.
- [27] K. Schweitzer, M. Walsh, S. Calhoun, H. Schütze, B. Möbius, A. Schweitzer, and G. Dogil, "Exploring the relationship between intonation and the lexicon: Evidence for lexicalised storage of intonation," *Speech Communication*, vol. 66, pp. 65–81, 2015.
- [28] P. Boersma and D. Weenink, *Praat: doing phonetics by computer. Version 6.0.25*, 2017. [Online]. Available: <http://www.praat.org/>
- [29] T. Kisler, U. Reichel, and F. Schiel, "Multilingual processing of speech via web services," *Computer Speech & Language*, vol. 45, pp. 326 – 347, 2017.
- [30] SAS Institute Inc, *JMP, Version13*, Cary (NC), 2017.
- [31] M. Ernestus, I. Hanique, and E. Verboom, "The effect of speech situation on the occurrence of reduced word pronunciation variants," *Journal of Phonetics*, vol. 48, pp. 60–75, 2015.
- [32] P. Wagner, J. Trouvain, and F. Zimmerer, "In defense of stylistic diversity in speech research," *Journal of Phonetics*, vol. 48, pp. 1–12, 2015.