

Schwa Optionality in Spontaneous Speech in German: A Meta-Study of Verbal Inflection in Three Corpora

Christine Mooshammer, Marie-Theres Weißgerber, Robert Lange

Institute for German Studies and Linguistics, Humboldt-Universität zu Berlin, Berlin, Germany

christine.mooshammer@hu-berlin.de

Abstract

Standard grammars of German prescribe verb-final schwa in the first person singular for many verbs, e.g. [ha:b@] for *habe* ‘have’. This variant, however, is mostly realised without the schwa in spontaneous speech, e.g. [ha:p]. The aim of this study is to investigate the conditions promoting the realisation of schwa in this position. We compare the effect of the segmental context, position within the phrase, speech rhythm, speech rate and situation in three corpora of spontaneous speech. Overall, final schwa was realised only in 21% of all potential cases, with more schwas in IP final position and formal situations.

Index Terms: spontaneous speech, phonetic reduction, German, situational variation

1. Introduction

As has been found in several corpus studies of English, between 25 and 32% of words in spontaneous speech are missing one or more phones compared to the canonical forms (see e.g. [1]). Phonetic reductions are even more common in function words and inflections, and are also found in read speech. As noted by Tucker and Mukai [2], spontaneous speech is a broad term that generally refers to unprepared speech and encompasses a wide range of different speaking styles that vary according to situation, task, addressee etc. The aim of this study is to investigate spontaneous speech phenomena in a wider range of speaking styles (or registers) (see e.g. [3]) and other factors that have been shown to contribute to phonetic variation. In the present study, we focus on the realisation of verb-final schwa inflection in three corpora of spontaneous speech in German. In written German, 1st person singular present tense weak verbs are written with a final <-e> that is pronounced as a schwa [ə], e.g. in *ich mache* ‘I make’ the canonical pronunciation is [ɪç maχə]. In spontaneous speech, the realisation of the verb-final schwa is optional, and for high frequent verbs, such as *hab-e*, *seh-e*, *glaub-e* ‘have, see, believe’, rather rare (see [4]). In the present tense, dropping the final schwa does not change the meaning (unlike the 3rd person singular in the preterite).

The distribution of this seemingly free variation between e.g. [ha:bə] and [ha:p] ‘(I) have’ is partly determined by several linguistic and situational factors. First, we will consider how suprasegmental aspects influence reduction phenomena in German. For example, [5] found in several experiments that the speech rhythm, with an alternation between strong and weak syllables, affects whether the final <-e> in adverbs with two alternative spellings, such as *gern*, *gerne* ‘gladly’, is realised. Speakers produce schwas more often before stressed syllables in order to avoid stress clashes (see e.g. [6]). Kentner [5], p. 117, identifies a link between the concept of “rhythmic alternation” and the impacts of “stress clash avoidance” and “stress

lapse avoidance”. A strengthening effect has been found in [7] for following prosodic boundaries with more frequent final schwa in <-en> endings in phrase-final position compared to phrase-medial position. Fast speech rates lead to more frequent reductions, as predicted by Lindbloms H&H Theory [8]. At the segmental level, the following context affects the likelihood of schwas, as found, by Kohler et al. [9] for example, with less frequent final schwas when the following word starts with a vowel. Apart from these phonetic and phonological aspects the semantic load of the inflection also plays a role. As Zimmerer et al. (2014) [10] found for German, reductions of <-t> are less likely when they distinguish between word forms. For verb-final schwa this is the case in the 3rd person singular, preterite, as in *hatte* ‘had’ which would be homophonic with *hat* ‘has’ 3rd person singular, present tense without the schwa.

Word frequency, discourse mention, neighbourhood density and predictability also affect how explicit words are pronounced. Clopper & Turnbull (2018) [11] attribute these effects to the smaller processing demands for listeners and speakers of frequent, given and predictable words from high density neighborhoods, which lead to more reduced variants. Based on these factors, they distinguish between ‘easy’ and ‘hard’ contexts leading to more or less phonetically reduced forms, respectively. Different speaking styles are also subsumed under this dichotomy by Clopper & Turnbull (2018) [11]. Compared to plain speech, clear speech is easier for listeners to process, but imposes a higher cognitive load on speakers. In the current study, we examine two speaking styles that elicit clear speech phenomena: foreigner-directed and formal speech. Foreigner-directed speech, also known as non-native addressee register (NNAR), is characterised by slower speech rates, hyperarticulation and greater vowel dispersion (see [12] for a review, and e.g. [13]). Definitions of formal speech frequently refer to Labov’s notion of “attention given to the speech process” [14]. It is distinctive from informal speech in that speakers often adhere closely to written language [3] and to the standard [15].

In this study, we compare the frequency of verb-final schwa realisations in three corpora of spontaneous German, varying the elicited registers and tasks, as described in Section 2.1. Since NNAR and formal speech are so-called hard registers in the terminology of Clopper and Turnbull (2018) [11], we assume that verb-final schwas are realised more frequently than in informal speech or when speaking to L1 addressees. We also compare free conversations with task-based dialogues (spot the difference task), where the predictions are more exploratory. In terms of processing load for the speaker, the task-based dialogues should elicit more explicit word forms. However, in a previous study [4], we found a slightly but significant increase of verb-final schwas in free conversations compared to task-based dialogues. Furthermore, we will address a subset of the

above mentioned factors: phrasal strengthening (schwa adjacent to a boundary), speech rhythm (following word stress), following phonetic context and speech rate. Predictability will be operationalised in a limited context: since word order is rather flexible in German, 1st person singular verb forms are often followed by *ich*, e.g. *habe ich* ‘have I’, making the co-occurrence of this verb form and the pronoun highly predictable. By comparing these factors in three corpora of spontaneous speech we aim to gain a better understanding of how pronunciation variants are affected by different speaking styles.

2. Method

2.1. Corpora

2.1.1. BeDiaCo: Berlin Dialogue Corpus

The Berlin Dialogue Corpus (BeDiaCo_{main}, [16]) contains spontaneous speech with two different tasks and word lists read aloud. It includes eight face-to-face dialogues by 16 speakers (10 male, 6 female) from Northern Germany between the ages of 18 and 31. Prior to the experiment the participants did not know each other. The two tasks consisted of a free conversation about a topic of their choice for about 15 minutes and to solve two spot-the-difference-tasks, referred to as Diapix tasks (about 8 minutes each). The Diapix task [17] was developed as a dialogue elicitation procedure in which the interlocutors collaborate to find differences between two highly similar pictures without seeing each other’s versions. For the recordings, the participants were seated in a sound-attenuated booth and equipped with two headsets from Beyerdynamics (Headset Opus 54). Both microphones were connected to a preamplifier, and one channel was assigned to each speaker.

2.1.2. CoNNAR: Corpus of Non-native Addressee Register

The CoNNAR_{videocall} corpus was recorded with the aim to elicit foreigner-directed speech with highly proficient learners of German [18]. This sub-corpus has a similar structure to BeDiaCo_{main}. The difference to BeDiaCo is that each participant communicates with an L1 speaker and an L2 speaker, the tasks are the same. 20 German L1 speakers (10 female; age 20–38 years, mean age = 26 years, sd = 4.5 years) were recorded in two sessions with instructed confederates: Once with another German L1 speaker (L1 confederate, n = 4) and once with an English L1 speaker (L2 confederate, n = 4, from the UK and the USA) with self-reported mid to high proficiency in German. Each confederate participated in five sessions with the experimental participants. The order of L1 and L2 confederate experiments was counterbalanced. As in BeDiaCo, the recordings of the 40 experimental sessions consisted of word lists, two Diapix tasks (8 minutes each) and an 8 minute free conversation. Due to the COVID-19 pandemic the interlocutors were placed in neighboring rooms and connected via Zoom. The participants were seated in a sound-attenuated booth, the confederates in an office. The data was recorded using directional microphones (Sennheiser) via Audacity [19] as stereo WAV files. Only the data of the experimental participants are considered here.

2.1.3. RUEG: Research Unit “Emerging Grammars in Language Contact Situations”

The RUEG subcorpus *RUEG-DE German* [20] contains spontaneous speech recordings of native speakers of German. Participants watched a short video of an accident and were asked to summarise the events. The experiment comprises two task

set-ups, with one formal and one informal setting, which are intended to yield two different speech registers. The experimental set-ups differed from one another in terms of both the design of the experiment room [20] and the task. In the formal condition, subjects were instructed to report the accident they had just observed to a police officer. In the informal condition, they provided a report of the events to a friend. In both tasks, participants were instructed to imagine the conversation partner and to deliver the report via a voice message. 94 recordings of 47 subjects (25 female, between 13 and 37 years) are analysed.

2.2. Annotation

Table 1: *Potential predictors of schwa realisation*

Predictor	Levels
Following context	<i>Obstruent/ Sonorant/ Vowel</i>
Stress of the follow. syllable	<i>Stressed/ Unstressed/ Pause</i>
Global articulation rate	<i>Numeric</i>
Following I	<i>Yes/ No</i>
Lemma frequency	<i>Numeric</i>

Annotations were carried out automatically and corrected manually in Praat [22] on different tiers (see Figure 1 for RUEG) in all three corpora. One tier contains text transcriptions of the recordings, tokenised on the word-level. An extra tier was inserted to specify the stimulus, which is defined as a first-person singular verb with an optional schwa-suffix. In RUEG data annotations, first-person singular irregular preterite verbs such as “wollte” (‘wanted’) are included. The third tier contains annotations of the realisation of schwas with binary labels for being present or not. The following context of the potential schwa location was annotated on a different tier: pauses, obstruents, sonorants and vowels. In the case of BeDiaCo and CoNNAR, these annotations were done automatically, and for RUEG manually. The last tier contains labels for the stress of the following syllable, with manually annotated values for unstressed and stressed syllables, based on the auditory impression. In RUEG, the occurrence of “ich” (‘I’) before or after the target word was annotated manually, and in BeDiaCo as well as CoNNAR automatically extracted by the query function of the emuR database system [23]. The articulation rate was calculated by enumerating the syllables of all instances of the pronunciation-based transliteration per speaker for each task using the R package *syly* 0.1-6 [24] with *syly.de* 0.1-2 [25] and dividing this value by total length of articulation time. Silent pauses, and extra- and paralinguistic events such as laughing, clicks, and background noises as well as pseudonymised tokens were excluded. Lemma frequency contains the absolute frequency of verbs, centred and logarithmised to the natural base.

2.3. Statistics

We used a binary logistic regression analysis (R packages *lme4* [26] and *lmerTest* [27]) to test which factors influence verb-final schwa realisation. Schwa realisation (reference level with schwa vs. without schwa) was included as the dependent variable, and speakers and lemmas as random intercepts to account for individual and word differences. Each corpus was tested separately. Two sets of models were computed because the predictors following segmental context and stress both include the factor level *Pause*. Therefore, the first set of models includes task and/or register, stress, centered articulation rate and

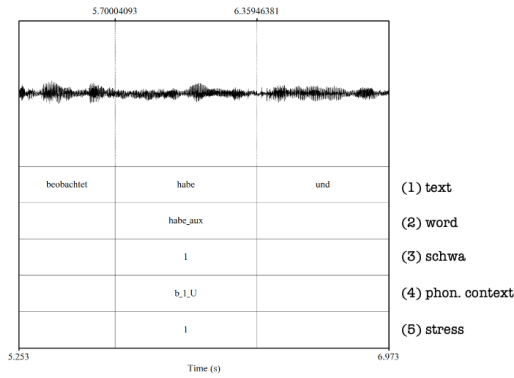


Figure 1: Annotation layers for the analysis of the RUEG data.

lemma frequency (see Table 1). For the second set of models, all items with a following pause were deleted and the predictors were following segmental context, following I, task and/or register, centered articulation rate and lemma frequency. By model comparisons it was tested whether inclusion of the interactions improved the models. However, none of the models improved. Significant results are presented in the text. The data and a script with the statistical models can be found at <https://osf.io/hgv9p/>.

3. Results

3.1. Overview

In all three corpora, there are 2624 finite verb forms with a potential final schwa. Only 21% of these are realised. Table 2 gives an overview of the corpora with the number of participants and the amount of available data. The number of data points for RUEG is lower (194) because the recordings per participant are much shorter (74.4 seconds) than for the other corpora (about 48 minutes for CoNNAR and 30 minutes for BeDiaCo). Regarding the percentage of realised schwas (last lines in Table 2), RUEG shows the highest percentage of realised schwas in verb-final position with 33.5%. The lowest percentage of realised schwas (18.3%) is found in the CoNNAR corpus, followed by BeDiaCo with 23.7%.

Table 2: Overview of corpora.

	BeDiaCo	CoNNAR	RUEG
Participants N	16	20	47
Duration (min)	180	410	101
Tokens	41036	85949	17538
Tasks	Diapix free conv.	Diapix free conv.	description –
Channel	face-to-face	video	”imagined”
Addressee	L1	L1	friend
	–	L2	police officer
Potential schwas N	855	1575	194
- with schwa %	23.7	18.3	33.5
- without schwa %	76.3	81.7	66.5

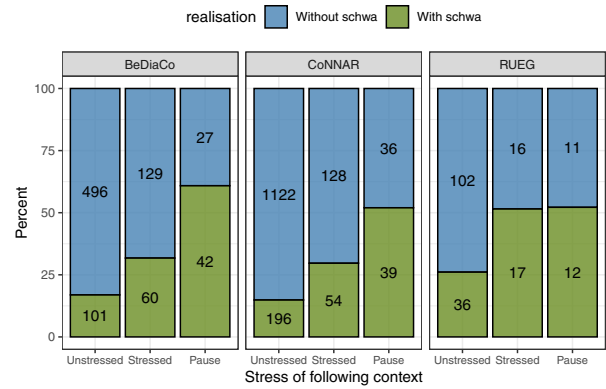


Figure 2: Effect of prosodic context, percentage and number of tokens.

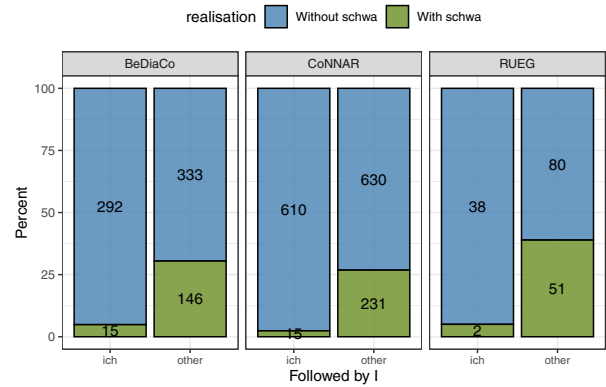


Figure 3: Followed by ich (excluding following pause).

3.2. Phrasing, stress and speech rate

Since not all corpora contain prosodic annotations, phrase boundaries are identified by a following pause as a first approximation. Figure 3.2 shows the effect of stress and phrasing. For all three corpora, significantly more schwa are realised before pauses with over 50%, confirming phrase-final strengthening (only compared to unstressed for RUEG). Schwas are more often realised when followed by a stressed syllable compared to following unstressed syllables (significant for BeDiaCo and CoNNAR), which may be a mechanism to avoid stress clashes. Articulation rate has no significant effect on the occurrence of final schwa in all three corpora.

3.3. Frequency and predictability

Lemma frequency is only significant for BeDiaCo with more final schwas in less frequent verbs, as expected. For CoNNAR and RUEG the statistical models do not converge when lemma frequency is included (for the data-set excluding 167 items followed by a pause). Figure 3.3 shows the number of realised schwas depending on whether the verb is followed by ‘I’ or not. Very few schwas (n=32 or 3.3% of the verbs) are realised when followed by ‘I’. This effect is significant for all three corpora.

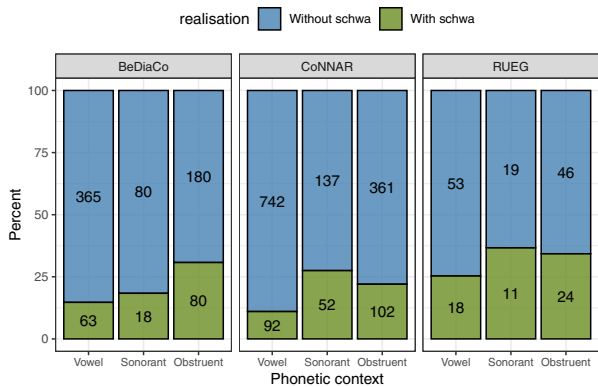


Figure 4: Effect of the following phonetic context (excluding following pause).

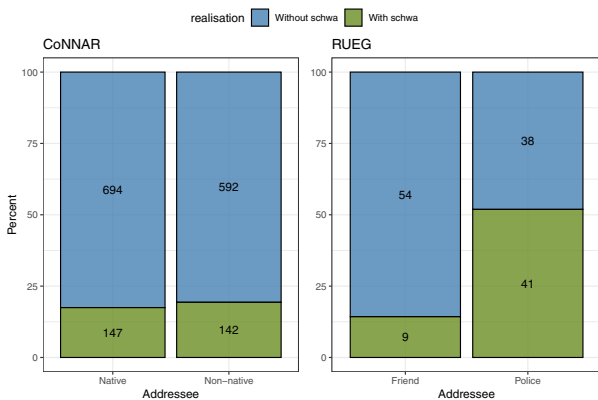


Figure 5: Addressee comparison in CoNNAR and RUEG.

3.4. Phonetic context

Significantly fewer schwas are realised when the following word starts with a vowel than with a sonorant for CoNNAR or with an obstruent for BeDiaCo. For RUEG the phonetic context does not show a significant difference, but a similar trend can be observed in Figure 3.4, on the right side.

3.5. Situational variation

Figure 3.5 compares the effect of the addressee on the realisation of schwa for native vs. non-native addressees in CoNNAR (left) and for an imagined friend vs. an imagined police officer in RUEG (right). In CoNNAR the native language of the addressee does not significantly affect the number of schwa realisations whereas in RUEG the more formal situation significantly increases the number of schwa realisations.

4. Discussion

By comparing three corpora, we have identified several consistent factors that affect the realisation of schwa in 1st person singular verb forms in German spontaneous speech. The number of schwa realisations is influenced by phrase boundaries, alternating rhythm, following ‘I’ and the phonetic context, with

more schwa realisations before pauses, stressed syllables, sonorants and obstruents. Another fairly consistent effect is caused by adjacent ‘I’ with virtually no schwa when followed by ‘I’. *Ich* in German is unstressed by default (only 22 *ich* tokens are stressed in our dataset) and starts with a vowel. Excluding all tokens followed by ‘I’ (37% of all cases) changes the results drastically: for verb forms followed by words other than ‘I’, the percentage of schwa realisations is not affected by the following stress patterns in all three corpora and even slightly increases for words starting with vowels compared to sonorants and obstruents. Therefore, the stress and the phonetic context effects are mainly driven by the very frequent sequence of verbs in the 1st person singular and ‘I’. As the pronoun *ich* is the second most frequent word in spontaneous German [28], this co-occurrence is highly predictable and therefore prone to phonetic reduction. As has been suggested in [29] sequences such as *habe ich* ‘have I’ are contractions or clitics that are resyllabified as in [ha.biç]. Note that for verbs with stem-final voiced obstruents, the deletion of schwa leads to word-final devoicing in German, e.g. [hap rç]. In most cases, however, the resyllabified variant is found in [29], a process similar to ‘enchaînement consonantique’ in French. We plan to investigate the phonetic details of these contractions further.

Another factor influencing phonetic reduction phenomena is speaking style or register. The corpora examined in this study vary in communicative tasks and addressees. As mentioned in the introduction, in [4] we found slightly but significantly more frequent schwa realisations in free conversations than in Diapix tasks. This could not be replicated for the CoNNAR corpus, which generally showed the largest reduction rate (see Table 2). Due to the restrictions during the pandemic, this corpus was not recorded face-to-face in one room, but via Zoom in two adjacent rooms. As Belz et al. 2023 [30] found for read speech, speakers reduce their vowel space in video-conferences compared to co-present situations. They argue that this may be due to the reduced involvement of the speakers in video situations. Contrary to previous findings on NNAR, our participants did not speak more clearly with more frequent final schwa realisations when speaking to non-native speakers. Several reasons could explain this null finding. First, as mentioned above, the speakers were in a video-conference environment. This may have reduced their ability to adapt to their interlocutors. Secondly, the non-native confederates are medium to high-level learners of German with an audible English accent, whereas previous studies have investigated NNAR towards non-native speakers with lower levels of proficiency (see e.g., [12]). The clearest differences between registers are found for formality in the RUEG corpus. Imagining talking to a police officer significantly increased the number of schwa realisations compared to talking to a friend. This striking effect also explains why overall RUEG has the highest percentage of realised schwa. Since these results are based on a subset of the data available in RUEG, we are currently annotating more speakers.

5. Conclusions

Nearly 80% of the possible verb-final schwas are not realised in spontaneous speech in German. This makes variants such as *ich find, ich glaub, ich hab* ‘I find, I believe, I have’ the standard that should be taught to learners of German. Whether more explicit pronunciations contribute to a foreign accent in German will be investigated in the near future.

6. Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1412, 416591334. We thank Torben Schilling, Miriam Müller, Megumi Terada and Bianca Sell for their annotation work.

7. References

- [1] Johnson, K., “Massive reduction in conversational American English”. In *Spontaneous speech: data and analysis. Proceedings of the 1st session of the 10th international symposium*, 29–54, 2004.
- [2] Tucker, B. V. and Mukai, Y., “Spontaneous speech”, Cambridge University Press, 2023.
- [3] Lüdeling, A., Alexiadou, A. and Adli, A. et al., “Register: Language users’ knowledge of situational-functional variation: Frame text of the first phase proposal for the CRC 1412.”, in *Register Aspects of Language Situation (REALIS)*, (1), 1-58, 2022.
- [4] Lange, R., Sell, B., Terada, M., Belz, M., Mooshammer, C. and Lüdeling, A., “Schwa realisation in verbal inflection in two dialogue registers of German spontaneous speech”, in *Zeitschrift für Sprachwissenschaft*, 1-30, 2024.
- [5] Kentner, G., “Schwa optionality and the prosodic shape of words and phrases”, in *Ulbrich, C., Werth, A. and Wiese, R., Empirical Approaches to the Phonological*, 121-151, 2018.
- [6] Selkirk, E., “Phonology and Syntax: The Relation between Sound and Structure”, Cambridge, MA: MIT Press, 1984.
- [7] Niebuhr, O., Görs, K. and Graupe, E., “Speech reduction, intensity, and F0 shape are cues to turn-taking”, in *Proceedings of the SIGDIAL Conference*, 261–269, 2013.
- [8] Lindblom, B., “Explaining phonetic variation: A sketch of the H&H theory”, in *W. J. Hardcastle and A. Marchal [Ed], Dordrecht: Springer*, 403-439, 1990.
- [9] Kohler, K. J. and Rodgers, J., “Schwa deletion in German read and spontaneous speech”, in *Spontaneous German speech: Symbolic structures and gestural dynamics*, 35: 97-123, 2001.
- [10] Zimmerer, F., Scharinger, M. and Reetz, H., “Phonological and morphological constraints on German/t/-deletions”, *Journal of Phonetics* 45, 64–75, 2014.
- [11] Clopper, C. G. and Turnbull, R., “Exploring variation in phonetic reduction: Linguistic, social, and cognitive factors.” in *Cangemi, F., Clayards, M., Niebuhr, O., Schuppler, B. and Zellers, M., Rethinking Reduction: Interdisciplinary Perspectives on Conditions, Mechanisms, and Domains for Phonetic Variation*, 25-72, 2018.
- [12] Piazza, G., Martin, c. and Kalashnikova, M., “The Acoustic Features and Didactic Function of Foreigner-Directed Speech: A Scoping Review” in *Journal of Speech, Language, and Hearing Research* 65.8, 2896–2918, 2022.
- [13] Uther, M., Knoll, M. A., and Burnham, D., “Do you speak E-NG-LI-SH? A comparison of foreigner-and infant-directed speech”, in *Speech communication*, 49(1), 2-7, 2007.
- [14] Labov, W., “The linguistic variable as a structural unit”, New York: ERIC, 1966.
- [15] Heylighen, F. and Dewaele, J.-M., “Formality of language: definition, measurement and behavioral determinants”, in *Interne Bericht, Center “Leo Apostel”, Vrije Universiteit Brussel*, 4(1), 1999.
- [16] Belz, M. and Mooshammer, C., “Berlin Dialogue Corpus (BeDiaCo) Version 1”, *Medien-Repositorium Humboldt-Universität zu Berlin*, URL: <https://rs.cms.hu-berlin.de/phon>, accessed on 25 May 2020.
- [17] Baker, R., and Hazan, V., “DiapixUK: task materials for the elicitation of multiple spontaneous speech dialogs.” *Behavior research methods*, 43, 761-770, 2011.
- [18] Lüdeling, A., Mooshammer, Ch., Lange, R., Sell, B. M., & Terada, M., “Corpus of Non-Native Addressee Register (CoNNAR)” Version 1, 2023, URL: <https://rs.cms.hu-berlin.de/phon/pages/home.php>.
- [19] Audacity team, 2020. Audacity (version 2.4.2). <https://www.audacityteam.org/>
- [20] Lüdeling, A. et al. “RUEG Corpus”, Version 1.0, <https://zenodo.org/records/11234583>, 2024.
- [21] Wiese, H., “Language Situations: A method for capturing variation within speakers’ repertoires”, in *Methods in dialectology XVI*, 45: 105-117, Frankfurt Main: Peter Lang, 2020.
- [22] Boersma, P. and Weenink, D., “Praat: doing phonetics by computer”, Version 6.0.43, <http://www.praat.org/>, 2023.
- [23] Winkelmann, R., Jaensch, K., Cassidy, S. and Harrington, J., “emuR: Main package of the EMU speech database management system”, R package version 2.3.0, 2021.
- [24] Michalke, M., “syllly.de: Language support for sylly package: German (Version 0.1-2)”, 2017, URL: <https://github.com/unDocUMeantIt/sylly>, accessed on 23 August 2022.
- [25] Michalke, M., “syllly: Hyphenation and syllable counting for text analysis (Version 0.1-6)”, 2020, URL: <https://github.com/unDocUMeantIt/sylly>, accessed on 23 August 2022.
- [26] Bates, D., Mächler, M., Bolker, B. and Walker S., “Fitting linear mixed-effects models using lme4”, in *Journal of Statistical Software* 67(1), 1–48, 2015.
- [27] Kuznetsova, A., Brockhoff, P. and Christensen, R., “lmerTest package: Tests in linear mixed effects models”, *Journal of Statistical Software* 82(13), 1–26, 2017.
- [28] Brackhane, F., “Beobachtungen zu Frequenz und Funktionen von ja in deutscher Spontansprache.” in *Deutsche Sprache* 4/2022, 335-363, 2022.
- [29] Trouvain, J., Mooshammer, C., Belz, M., Lange, R., “Merging verb forms with “ich” to enchaînement consonantique in German”, *Book of Abstracts ISSP 2024 - 13th International Seminar on Speech Production*, 2024.
- [30] Belz, M., Ebert, M., Müller, M., Sun, J., Terada, M., and Xia, Q., “Reduced vowel space in video conferences via Zoom: Evidence from read speech.”, *JASA Express Letters*, 3(10), 2023.