Segmental effects on prosody: Modelling German argument structure

Petra Augurzky[a]*, Arndt Riester[b], Fabian Tomaschek[a]

*Corresponding author: petra.augurzky@uni-tuebingen.de, Tel: +49 7071 29-73961
[a]Seminar für Sprachwissenschaft, Universität Tübingen, Wilhelmstr. 19, 72074 Tübingen, Germany
[b]Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, Pfaffenwaldring. 5b, 70569 Stuttgart, Germany

Abstract
Research on prosody-segment interactions usually focuses on the effects of prosody on segment realisation. In our study, we follow a different path by examining whether segmental cues affect the placement of prosodic boundaries. Though recent studies have shown that multiple factors contribute to the prosodic outcome of an utterance (e.g. Breen et al., 2011), the role of segmental information on boundary realization has not yet been addressed. The present corpus study thus examines whether segment clashes at word boundaries increase the probability of prosodic breaks and whether this proposed process contributes to a cognitive processing model predicting argument structure. In our first analysis, we examine whether consonant clashes affect speech *production* by analysing the segmental environment of potential boundary positions in a radio news corpus. In our second analysis, we use these production results as a background for predicting the *comprehension* of argument structures.
Our results show that segment clashes affect both of these modalities: The presence of consonant clashes significantly increases the number of prosodic boundaries in speech production. In comprehension modelling, segment clashes reduce the perceptual uncertainty when classifying two-argument structures, but increase it for one-argument structures.

Keywords
Segment effects, prosody, argument structure, modelling, corpus analyses, naïve discrimination learning

Segmental effects on prosody: Modelling German argument structure

1.0 Introduction

Recent studies have shown that multiple factors contribute to the prosodic phrasing of an utterance, amongst them syntactic structure, sentence meaning, constituent length, and prosodic balance (e.g. Breen et al., 2011). Thus, prosodic phrase boundaries are not exclusively informative for signalling the intended meaning of an utterance. For instance, suprasegmental information like eurhythmic constraints and balance (e.g. Gee and Grosjean, 1983; Selkirk, 2000) affect the prosodic phrasing of an utterance, and thus indirectly contribute to the interpretation of ambiguous structures (e.g. Fodor, 2001). In a combined production-perception experiment in German, prosodic balance was shown to affect both the production and the comprehension of ambiguous argument structures (Augurzky, 2008).

Whereas suprasegmental cues like balance and constituent length have been investigated in various languages, the role of segmental information on boundary realization has not been addressed yet. Indeed, research on prosody-segment interactions usually examines the effects of prosody on segment realization and not vice versa. For instance, these studies address the question how phenomena like boundary realization, speech rhythm, tempo and speech style affect segment production (e.g. Byrd & Choi, 2006; Cho & McQueen, 2005; Keating, 2006).

In our study, we follow a different path by examining whether and how subtle segmental cues themselves affect prosodic phrasing, and whether such processes occur independently of structural disambiguation and sentence meaning, comparably to prosodic balance. The rationale behind this idea was the following: we assume that if speakers envisage articulation difficulties, the production system will have to slow down, thus leading to segment lengthening and/or the production of silent phases in the signal. Both of these parameters are also involved in boundary realisation. As a result, purely motor-driven phenomena might trigger comprehension difficulties by being misconstrued as signalling a (potentially informative) prosodic boundary.

In the present study, we first investigated whether segment clashes at word boundaries increase the probability of prosodic breaks in speech *production*. Based on the results from this production analysis, we then tested whether the inclusion of clash information improves the performance of a cognitive processing model predicting argument structure *comprehension*. To this end, we examined the prosodic realization of German argument/modifier phrases in a German radio news corpus (Eckart, Riester & Schweitzer, 2012):

(1) ... [die Gewerkschaf**t**]$_{\mathrm{NOM}}$ [**d**en Druck$_{\mathrm{ARG}}$]$_{\mathrm{AKK}}$...
... the trade union the pressure ...

(2) ... [der Geschicht**e**$_{\mathrm{GEN}}$ [**d**er Menschheit$_{\mathrm{ADJ}}$]$_{\mathrm{GEN}}$]...
... the history of-the mankind ...

In the examples in (1) and (2), two subsequent noun phrases differ with respect to their argument structure properties. Whereas in (1), both noun phrases are arguments of the verb (*two-argument structure*), the second noun phrase in (2) is a possessive modifier of the first noun phrase, and thus only one verbal argument is realized (*one-argument structure*). In reading studies, two-argument structures comparable to those in (1) have been shown to be separated by a prosodic boundary, whereas in one-argument structures as in (2), both noun phrases are usually included within a single prosodic phrase (Augurzky, 2008; Gollrad, 2008).

The main aim of the present study was to replicate previous findings from experimental settings by considering spoken corpus data. Additionally, we determined whether prosodic phrasing is mediated by segment information. To this end, we automatically extracted constructions as in (1) and (2) from the corpus and classified word boundaries at the edges between the two noun phrases into *potential boundary positions* (*PBP*s). Segments around *PBPs* were then classified according to whether or not their succession led to a segment clash.

The hypotheses for the production analysis can be summarized as follows: Based on the literature, we expected an increased number of boundaries between the two noun phrases for two-argument structures as opposed to one-argument structures. If the presence of a consonant clash has an effect on phrasing, we additionally expected that more boundaries were realized in the presence of a clash compared to the non-clash cases. Although such a specific effect has not yet been shown in the literature, similar phonological effects have: Jaeger (2012) showed that speakers avoided overlapping word onsets between two consecutive words on speech production. Since Munson (2011) showed that cluster probability correlates negatively with cluster durations, we hypothesized that more frequent segment combinations within words and across words reduce the number of boundary realizations. This is simply because their pronunciation is better trained.

Studies with infants have shown that we strongly rely on phonotactics when segmenting words (Mattys, 1999). This suggests that such information, which is represented by consonant clashes in our study, is also used during perception. In our comprehension modelling, we tested how clash information affects the performance of a learning algorithm trained to classify argument structure. As consonant clashes are expected to result in an increased amount of boundary productions, clashes co-occurring with one-argument structures should result in misleading information and thus to a decrease in model performance. If that were the case, this would be a starting point for future behavioural experiments, testing whether clash information induces a bias of choosing a particular syntactic template.

2.0 The Corpus: DIRNDL

The corpus used in the experiments presented below is the DIRNDL database (*Discourse Information Radio News Database for Linguistic Analysis*, see Eckart et al., 2012). It consists of three days of hourly radio news broadcasts from a German radio station *(Deutschlandfunk)* and comprises approximately five hours of read speech uttered by nine speakers (five male, four female). The corpus was manually labelled with respect to prosody according to GToBI(S) (Mayer, 1995). The speech signal was automatically aligned with word, syllable and phoneme boundaries using forced alignment (Rapp, 1995). The speech and label files have been processed with the Edinburgh Speech Tools Library (Taylor, Caley, Black, & King, 1999) to calculate information about e.g. position in the prosodic phrase.

The transcript was processed with various language processing tools, one of them the XLE parser (Crouch et al., 1993-2011) using a German LFG grammar (Rohrer & Forst, 2006). POS tags, morphology, and dependency trees are also available. A particular advantage of the DIRNDL database is its alignment of prosodically annotated spoken and syntactically analysed written data, which are accessible via queries to an SQL database management system.

3.0 Clash Classification

As a first step, we extracted all structures in the corpus that were either syntactically labelled as "sequential" (corresponding to a succession of two verbal arguments as in (1)) or "embedded" (corresponding to a modifier structure as in (2)). After manual verification and a re-classification of a few wrong parses, we obtained 1643 embedded structures, and 1231 sequential structures. We then determined consonant sequences around potential boundaries (i.e. the final segment of the first phrase and the first segment of the second phrase) in two separate data sets. The first data set was based on the orthographic transcriptions yielded by the automatic extraction process (*orthographic data*), and the second set was based on the manual annotation of the actually realized segments, including information about phonological processes such as t-deletion, glottalisation or r-vocalisation (*phonological data*).

In the orthographic data set, certain rule-based phonological processes were considered that occur reliably in German. For instance, the final segment in "Prag" was labelled as [k] due to final obstruent devoicing. Vowels were excluded from this analysis, since the annotation of the glottal stop was inconsistent. After the exclusion of vowels and missing values, 1452 PBPs were available for analysis: 947 (65.2%) one-argument, 505 (34.7%) two-argument structures. For the phonological data set, the actually produced segments around the PBPs served as a basis for clash classification. Vowels were included into this analysis, resulting in a total of 2874 structures. For both data sets, sequences were classified as clashes if their syllable-internal co-occurrence was prohibited by German phonotactic rules (e.g. [*tp*]), unless their combination did not result in a succession of two voiced segments (e.g. [*ld*]). Both legal clusters and voiced combinations were labelled as No-Clash. Based on the literature on speech production (see

Cholin, Schiller & Levelt, 2004, and references therein), we hypothesize that sequences that may co-occur syllable-internally are more likely to be planned as part of a production unit than illegitimate sequences and thus indirectly reflect articulatory ease in speech production. The introduced classification criteria allow us to capture a large part of the data according to our intuition about whether or not a segment combination can be produced smoothly. Here are some examples from our corpus:

(3) *Clash*

a.	... die Gewerkschaft den Druck ... *... the trade-union the pressure ...*	→ Gewekschaf[**t d**]en → [**t d**]
b.	... in Prag beide Tore ... *... in Prague both goals ...*	→ Pra[**k b**]eide → [**k b**]

(4) *No Clash*

a.	... dem Deckmantel der zivilen Nutzung ... *... the cloak of-the civil use ...*	→ Deckmante[**l d**]er → [**l d**]
b.	... als Glücksfall für Europa und die Welt ... *... as stroke of luck for Europe and the world ...*	→ Glücksfal[**l f**]ür → [**r n**]

Finally, PBPs were checked with respect to whether or not a prosodic phrase boundary was actually realized. This information can be extracted from the corpus by considering the prosodic labelling. We included both intermediate phrase boundaries and intonational phrase boundaries into our analysis. If one of those two categories was realized, the PBP was labelled as a boundary, otherwise it was labelled as a non-boundary.

4.0 Analyses and results

4.1 Production analysis

In our first analysis, we investigated whether the presence of consonant clashes generally affects the prosodic phrasing of argument structure. All analyses were carried out using R, Version 3.0 (R_Development_Core_Team, 2010). The production analysis was based on the orthographic data set.[1] For statistical evaluation, generalized linear models were fitted to predict the probability of boundary production for one-argument structures and two-argument structures. The feature encoding segment clashes was taken from the classifying procedure described above. The model controls for several other features that can be accessed in the corpus. An overview upon all features used as predictors in the model is shown in the following:

Clash:	encodes the presence or absence of a consonant clash at the PBP.
Status:	encodes the argument status of the construction (one argument vs. two arguments).
NwordsLeft:	the number of words within the syntactic phrase left of the potential boundary. Ranges from 0 to 7. Due to sparse data, counts above 7 are collapsed to 7.
NwordsRight:	the number of words within the syntactic phrase right to the potential boundary. Ranges from 0 to 12. Due to sparse data, counts above 7 are collapsed to 7.
AccentsLeft:	pitch accent realization on the word preceding the potential boundary, classified into *Rise*, *Fall* or *None* according to Schweitzer, Riester, Walsh & Dogil (2009)
AccentsRight:	pitch accent realization on the word following the potential boundary, classified into *Rise*, *Fall* or *None*

[1]Note that we decided not to use the phonological data set as a basis for the production analysis because segments tend to be strengthened at the edges of prosodic boundaries in spoken language (e.g. Fougeron & Keating, 1997). Thus, an increase in the number of boundaries in clash environments might have been *caused* by a segment clash, or, alternatively, the boundary itself might be the reason for the production of non-reduced segments. For our first analysis, we thus focused on the idealized orthographic transcriptions, which we assume to be free of such confounds and to reflect planning processes of the underlying form more directly.

interFreq:	the frequency of segment combinations across word boundaries calculated in the SDEWAC corpus (Shaoul &Tomaschek, 2013)
intraFreq:	the frequency of segment combination with a word calculated in the SDEWAC corpus

We included accent patterns into the analysis since we found in our exploratory analysis that they affected significantly the boundary production. Figure 1 illustrates the predictions by the optimal model found by stepwise inclusion and exclusion of predictors and interactions and compared by Chisquare-tests. Y-values represent the log(P/P-1) where 0 equals 50%. Table 1 presents the results of a Chisquare-test applied to the generalized linear model.

Table 1

Chisquare-results for the generalized linear model

	Predictor	**DF**	**Deviance**	**P**
a	*Status*	1	240.7	< 0.001
b	*Clash*	1	13.4	< 0.001
c	*InterFreq*	1	10.9	< 0.001
d	*IntraFreq*	1	10.0	0.002
e	*AccentsLeft*	1	122.3	< 0.001
f	*Clash*InterFreq*	1	10.7	0.001
g	*NwordsRight*AccentsLeft*	2	27.2	<0.001

Status yielded significance: We replicated previous findings that a prosodic boundary cued argument phrases in ~80%, and adjunct phrases in ~20% of the cases. The occurrence of a prosodic break was significantly affected by the predictor *Clash* (Figure 1a, Table 1a&b). *NwordsLeft* missed significance and therefore was exluded from the model. *NwordsRight* did not show any main effect, but interacted significantly with *AccentsLeft:* after a Rise, the increase of the probability was steeper than after a Fall (Figure 1b, Table 1g). Rise patterns resulted higher probability of phrase boundaries than Fall patterns. Furthermore, *AccentsLeft* produced a significant main effect on its own (Figure 1c, Table 1e). However, the levels Fall and None did not differ significantly and could therefore be collapsed, leading to an improvement of the fit (i.e. the model with the higher number of AccentsLeft levels was not significantly better). Finally, sequence frequencies affected significantly the production of boundaries (Figure 1d). *InterFreq* showed a negative slope, indicating that the higher the across words sequence frequency, the lower the probability of a boundary (Table 1c). It furthermore interacted significantly with *Clash* (Table 1f). Under no clash, the slope of *InterFreq* was steeper than under clash, producing less boundaries for lower frequencies. Interestingly, *IntraFreq* resulted in an opposite pattern (Table 1d): The higher the sequence frequency within the word, the higher the probability of a boundary.

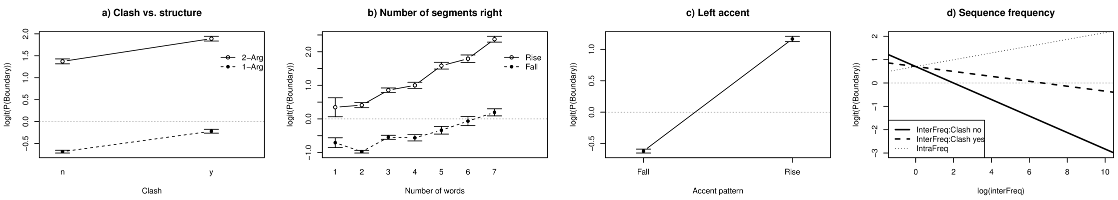

Figure 1

Probabilities of prosodic boundaries depending on: a) one/two-argument structures and clash; b) Number of words on the right side of the boundary in interaction with the left accent pattern; c) accent pattern on th left side; d) on sequence frequencies. Y-scales represent log(P/1-P) where 0 equals 50% chance.

4.2 Modelling Comprehension via NDL

In our second analysis, we investigated how these cues contribute to the parsing of the speech signal as an adjunct or argument phrase prior to a semantic disambiguation. For this analysis, we analyzed both the orthographic and the phonological data set. Whereas the former gives us results that are more directly comparable to our production results, the latter is probably more informative with respect to the cues that are actually available in the signal for the comprehender. Parsing was modelled with the *naive discriminative learner (NDL)* classifier (Baayen, 2011a). The idea behind NDL is that different outcomes (such syntactic structures) are cued by certain cue combinations (such as phonemes, or pauses, accents etc.). Given various environments such as reduction, one outcome can be cued by different cue combinations. E.g the word *singen* 'to sing' can be signalled by *[zɪŋən]*, *[zɪŋə]* or *[zɪŋ]*. NDL simulates the cognitive process of learning by calculating how strong different cue combinations activate an outcome. Activation strength depends on how often cue and outcome co-occur. NDL's mathematics, which is based on Rescorla-Wagner equations, can be found in Baayen (2011). NDL has repeatedly been shown to be a valid tool to correctly predict human linguistic behaviour (e.g. Baayen, 2011). This is why, here, the model results themselves are presented without any additional experimental data. We were interested in how well non-semantic cues prior to the phrase boundary indicate what syntactic structure should follow. This is why, we used the following predictors as cues: *NwordsLeft, AccentsLeft, Boundary, and Clash.*

Since we were interested in segmental effects, we focused our analysis on the *Clash* cue. For this, we trained NDL 100 times with every possible cue combination using randomly selected 75% of the corpus data for classifying the presented cues into one/two-argument structures. In sum, 31 cue combinations were used, concatenating one to five cues. Cue level combinations amounted to 1117, summed up over all cue combinations. The learned model was tested with 25% of the data. In the second step, we calculated the percentage of correct one/two-argument structure classifications. Figure 2a illustrates the difference between cue combinations containing and those not containing Clash. Results from the phonological data set are not illustrated, as model performance was comparable for both data sets. Descriptively, Clash (adjusted R^2: 0.9) resulted in a slightly better model fit (Phonoclash adjusted R^2: 0.88).

Table 2

Linear model results for classification of the syntactic structure by NDL.

		Estimate	Std. Error	T-value	P
a	Intercept: Structure: one-argument	23.8 %	0.3	80.2	<0.001
b	Structure: two-argument	69.7 %	0.4	166.3	<0.001
c	Clash: No clash	-5.4 %	0.4	-12.6	<0.001
d	Two-arguments*noClash	4.8 %	0.6	7.9	<0.001

Clash significantly increases the reliability with which the presented cues can be classified as a given structure (Diff = 3.3%, t=13.3, p<0.001). We furthermore investigated how the correct classification was affected by syntactic structure. We thus modelled the percentage of correct classification with a linear model using syntactic structure and clash as independent factors (R-squared=0.8, F(3,6196)=19120, p<0.001). The results are presented in Table 2 and in Figure 2b. Two-argument structures were significantly more often correctly classified than one-argument structures (Table 2 a.+b). One-argument structures were less often correctly classified when no clash occurred (Table 2 a+c). Two-argument structures were more correctly identified, when a clash occurred (Table 2 b+d).

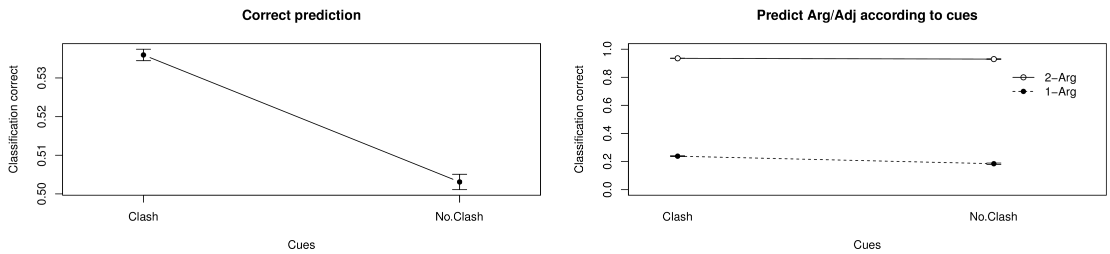

Figure 2

Percentage of correct status classifications by the naive discriminative learner model.

4.0 Discussion

4.1 Results from the production analysis

In our first experiment, we replicated previous findings from experimental studies. Generally, verbal arguments tend to be separated from each other by a prosodic phrase boundary. This result nicely corresponds to the present corpus data. Though there is a demand for a prosodic boundary in the argument condition or a prohibition of a boundary in the adjunct condition, alternative realisations are permitted as well. In our study, 80% of all arguments were realized by a prosodic phrase boundary, and 20% were realized without a boundary (see Augurzky, 2008; and Gollrad, 2008, for comparable results).

As for the effects of consonant clashes, the results are in accord with our initial predictions: the number of prosodic phrase boundaries significantly increases in the presence of a segment clash. This can be observed independently of the actual argument status of the construction. Moreover, our results replicated previous experiments showing that increasing constituent length generally increases the likelihood of a prosodic phrase boundary (e.g. Watson & Gibson, 2004). We included the frequency of segment combinations within words and across word boundaries in our analysis. Across word segment frequency was congruent with our hypothesis that frequent combinations would lead to a decrease in pronounced consonant clashes / boundaries (Munson, 2011). We interpret this as a learning effect due to frequent use, as has recently been shown for CVC articulation (Tomaschek et al., 2013). By contrast, frequent within-word-sequences resulted in a higher boundary probability, probably due to articulatory inhibition of across-word sequences by within-word sequences. Given the present data, we cannot answer why this might be the case.

4.2 Results from the comprehension modelling

In the second part of our study, we modelled how the *Clash* cue in combination with other cues would affect the classification of the speech signal either as a one-argument or a two-argument structure by means of the *naive discrimination learning classifier* (NDL, Baayen 2011). The NDL learning model was able to classify the two-argument structures almost perfectly whereas the classification of one-argument structures failed (Figure 2b). One possible explanation for this failure might be that the cue combinations did not consistently refer to the one-argument structure but pointed to both structure categories in a heterogenous way. These result predicts that subjects should have a big disadvantage in classifying one-argument structures correctly. Furthermore, the simulation shows that the consonant clash, though minimally, is used as a cue to predict which structure follows the phrase boundary. In this sense, we could show that segments are not only used to parse low-level structures, but also provide information about higher-level structures. This finding has to be investigated in future behavioral experiments.

5.0 Conclusions

We have shown that the speech signal is affected by the interplay of syntactic, prosodic and segmental information. Consonant clashes have both an effect on the production and the comprehension of argument structure. In addition, our results demonstrate that specific combinations of these information types can be used to predict the actual syntactic outcome prior to further disambiguating speech material e.g. at the semantic level. The present analyses offer a computational basis for future behavioural and neurolinguistic studies on the comprehension of syntactic ambiguities.

Acknowledgements

This study was funded by TüZLI, by the DFG and the Alexander-von-Humboldt Professur. We thank two anonymous reviewers for their helpful comments and Moritz Stiefel, Tanja George, Hanna Kicherer and for their help with labelling the phonological data set.

References

Augurzky, P. (2008). Prosodic phrasing in German sentence production: Optimal length vs. argument structure. In: Botinis, A. (Ed.): *Proceedings of the ISCA Tutorial and and Research Workshop on Experimental Linguistics* (pp. 33-36). University of Athens.

Baayen, R.H. (2011): Corpus linguistics and naive discriminative learning. *Brazilian Journal of Applied Linguistics*, 11, 295-328.

Breen, M., Gibson, E., & Watson, D. (2011). Intonational phrasing is dependent on the length of semantic dependency. *Language and Cognitive Processes*, 26 (10), 1532-1562.

Cholin, J., Schiller , N.O., & Levelt, J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, 50 (1), 47-61.

Crouch, D., Dalrymple, M., Kaplan, R., King, T., Maxwell, J., & Newman, P. (1993-2011, ms). XLE Documentation, Palo Alto Research Center. URL http://www2.parc.com/isl/groups/nltt/xle/doc/xle_toc.html

Byrd, D., & Choi, S. (2006). At the juncture of prosody, phonology, and phonetics—The interaction of phrasal and syllable structure in shaping the timing of consonant gestures. *Papers in Laboratory Phonology 10*. Mouton.

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics*, 33(2), 121-157.

Eckart, K., Riester, A., & Schweitzer, K. (2012). A Discourse Information Radio News Database for Linguistic Analyses. In: C. Chiarcos, S. Nordhoff &S. Hellmann (eds.) *Linked Data in Linguistics: Representing and Connecting Language Data and Language*.

Fodor, J.D. (2002). Psycholinguistics cannot escape prosody. Paper presented at Speech Prosody 2002, Aix-en-Provence, France, April 11-13.

Gee, P. & Grosjean, F. (1983). Performance Structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, 15, 411-458.

Gollrad, A., & Sommerfeld, E. (2008): Prosodic disambiguation of German case ambiguity. *Prosody '08 -Experimental and Theoretical Advances in Prosody*, Ithaca.

Jaeger, T. F., Furth, K., & Hilliard, C., (2012). Phonological overlap affects lexical selection during sentence production. *Journal of Experimental Psychology.* 38, 5, 1439 – 1449.

Keating, P. (2006), Phonetic Encoding of Prosodic Structure. In J. Harrington and M. Tabain (Eds.), *Speech production: Models, phonetic processes, and techniques* (pp. 167-186), Macquarie Monographs in Cognitive Science, Psychology Press, New York and Hove,.

Mayer, J. (1995). *Transcribing German Intonation - The Stuttgart System*. Universität Stuttgart, Technical Report.

Munson, Benjamin (2011): Phonological pattern frequency and speech production in adults and children. Journal of Speech, Language and Hearing Research. 44, p 778-792.

R_Development_Core_Team. (2010). R: A Language and Environment for Statistical Computing. edited by R. F. f. S. Computing (Vienna, Austria).

Rapp, S. (1995). Automatic phonemic transcription and linguistic annotation from known text with Hidden Markov models - An aligner for German. In *Proceedings of ELSNET Goes East and IMACS Workshop "Integration of Language and Speech in Academia and Industry"* (Moskau, Russland).

Rohrer, C. & Forst, M. (2006). Improving coverage and parsing quality of a large-scale LFG for German. In: *Proceedings of the 5th Conference on Language Resources and Evaluation* (LREC 2006), Genoa (Italy), 2006.

Schweitzer, K., A. Riester, M. Walsh, & G. Dogil (2009). Pitch Accents and Information Status in a German Radio News Corpus. *Proceedings of Interspeech 2009*, Brighton.

Selkirk, E. (2000). The interaction of constraints on prosodic phrasing. In M. Horne (Ed.), *Prosody: Theory and Experiment* (pp. 231-261). Dordrecht: Kluwer.

Shaoul, C., Tomaschek, F., (2013). A phonological database based on CELEX and N-gram frequencies from the SDEWAC corpus. Personal communication.

Taylor, P., Caley, R., Black, A.W., & King, S. (1999). Edinburgh speech tools library. [http://www.cstr.ed.ac.uk/projects/speech tools/]. System Documentation Edition 1.2, for 1.2.0 15th June 1999.

Tomaschek, F., M. Wieling, D. Arnold, H. Baayen (2013): Word frequency, vowel length and vowel quality in speech production. *In Proceedings of the Interspeech 2013, Lyon.*

Watson D, Gibson E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19 (6), 713–755.