10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Effects of Surprisal and Boundary Strength on Phrase-final Lengthening

Bistra Andreeva, Bernd Möbius, James Whang

Language Science and Technology, Saarland University, Saarbrücken, Germany
[andreeva, moebius, jwhang]@lst.uni-saarland.de

Abstract

This study examines the influence of prosodic structure (pitch accents and boundary strength) and information density (ID) on phrase-final syllable duration. Phrase-final syllable durations and following pause durations were measured for a subset of a German radio-news corpus (DIRNDL), consisting of about 5 hours of manually annotated speech. The prosodic annotation is in accordance with the autosegmental intonation model and includes labels for pitch accents and boundary tones. We treated pause duration as a quantitative proxy for boundary strength. ID was calculated as the surprisal of the syllable trigram of the preceding context, based on language models trained on the DeWaC corpus. We found a significant positive correlation between surprisal and phrase-final syllable duration. Syllable duration was statistically modeled as a function of surprisal and prosodic factors (pitch accent and boundary strength) in linear mixed effects models. The results revealed an interaction of surprisal and boundary strength with respect to phrase-final syllable duration. Syllables with high surprisal values are longer before stronger boundaries, whereas low-surprisal syllables are longer before weaker boundaries. This modulation of pre-boundary syllable duration is observed above and beyond the well-established phrase-final lengthening effect.

Index Terms: information theory, surprisal, phonetic encoding, boundary strength, phrase-final syllable duration

1. Introduction

Language provides not only the expressiveness needed to communicate but also offers speakers a multitude of choices regarding how they may encode their messages – from the duration of segments and syllables, to the choice of words, structuring of syntactic elements, and arranging sentences in discourse. There is a growing body of evidence suggesting that speakers and listeners have access to probability distributions over linguistic units (e.g., [1, 2]). It has been demonstrated that phonetic structures increase in duration and distinctiveness when they are difficult to predict from context compared to easily predictable structures (e.g., [3, 4] for American English; [5, 6] for American English, Czech, Finnish, French, German, and Polish). Moreover, contextual predictability impacts on word [7, 8] and segment duration [9, 10]. Vowels are also strengthened in their spectral features when they are difficult to predict from their context compared to easily predictable vowels [4]. Closely related languages, such as German [11] and Dutch [10], also seem to show the same positive relationship between vowel dispersion and predictability. These examples, among many others, illustrate that speakers' choices and listeners' preferences are affected by the occurrence probability and frequency of how such units are realized in a variety of contexts.

To account for such findings, Aylett and Turk have postulated the Smooth Signal Redundancy (SSR) hypothesis [3, 4], which in its strong version posits that the inverse relationship between information-theoretic factors (language redundancy) and acoustic realization of phonetic structures (acoustic redundancy) is moderated through prosodic structure. Since lengthening at the end of intonation phrases (e.g., [12, 13, 14]) compromises any simple relationship between redundancy and prosodic structure, the authors have formulated the weak version of SSR as well, which accepts that another major factor, viz. pre-boundary lengthening, modifies the relationship between prosodic structure, phonetic encoding, and language redundancy.

In an information-theoretic account, there are several measures to quantify the amount of information conveyed in a message [15]. One of them is surprisal. Surprisal captures the intuition that linguistic expressions that are highly predictable in a given context convey less information than those that are surprising. Surprisal is defined as the contextual predictability of a unit and can be used as a measure of the amount of information that is conveyed by that unit in terms of bits, using Equation (1) where S stands for surprisal and P for probability:

$$S(unit_i) = -log_2 P(unit_i|Context) \quad (1)$$

Production and perception studies have provided evidence that boundaries of the same prosodic category can have different strengths, operationalized as lengthening of the phrase final coda consonant [16], duration of whole constituents [17], fundamental frequency (F0) declination and boundary duration [18], acoustic duration, F0 movements and sandhi phenomena [19], and the duration of articulatory constriction [20]. We hypothesize that this variability and optionality is governed by properties of the information density profile across intonation phrase (IP) boundaries.

The present study is concerned with the analysis of the relationship between prosodic structure (in terms of phrasal accent and IP boundary strength), contextual predictability (defined as surprisal), and IP final syllable duration in German read speech, and addresses the following questions:

- What is the effect of contextual predictability on IP final syllable duration?
- What is the effect of phrasal accent and boundary strength on IP final syllable duration?
- Does contextual predictability relate to IP final syllable duration independently of prosodic structure?

We hypothesize that:

1. IP final syllables which are difficult to predict from the context are longer in their overall duration, while easily predictable syllables are temporally reduced.
2. The modulation between surprisal and phonetic encoding by means of prosody is not comprehensive, as posited by the SSR hypothesis in its strong version.
3. Surprisal shows effects in addition to, or in interaction with, prosody, as suggested by the findings in [5] and [6], which are compatible with a weak version of Aylett and Turk's SSR hypothesis.

 10.21437/SpeechProsody.2020-30

2. Material

2.1. Speech corpus

We used a subset of the German **D**iscourse **I**nformation **R**adio **N**ews **D**atabase for **L**inguistic analysis (DIRNDL) [21, 22]. The corpus consists of approximately five hours of read speech produced by nine news anchors (5m, 4f). It was automatically segmented into words, syllables, and phonemes using forced alignment [23]. The aligner has acoustic models for silent intervals, silence-breathing-silence sequences, and plosives following a silent interval. It is admittedly debatable whether the duration assigned to the closure phase, and thus to the entire plosive, is realistic, but it is at least consistent given the lack of articulatory data. Pitch accents and prosodic boundaries were manually labeled by one student assistant according to GToBI(S) [24]. Word level annotations were mapped to syllable-based prosodic labels using Festival [25].

2.2. Language modeling corpus

Language models (LMs) are based on the DeWaC corpus, which was preprocessed and normalized using German Festival [25]. DeWaC is a web-crawled corpus containing 1.5 billion running words and about 8 million types in a diverse range of genres from newspaper articles to chat messages. Text-internal criteria consist of removal of web-specific structures, such as HTML structures or long lists. DeWaC was split into a training (80 %) and a test corpus (20 %). The training corpus was used to train syllable-based trigram LMs using the SRILM toolkit [26]. We trained LMs including word boundaries. The default LM in SRILM calculates the conditional probability of a linguistic unit occurring with a preceding context. Witten-Bell smoothing was applied to all models.

3. Method

3.1. Speech data preprocessing

Each data point in our analysis is the last syllable before a full intonational phrase (IP) boundary corresponding to a ToBI break index of 4. We analysed 1425 IP final syllables in total. For these syllables as well as for the pauses occurring right after them, the automatic segmentation provided in the DIRNDL corpus was manually verified, and corrected if necessary, by the first author. Syllable and pause durations were determined on the basis of the segmentation. Since previous work has indicated that perceived boundary strength is positively correlated with pause duration [27, 28, 29, 30, 31, 32], we treated the duration of the pauses after the final syllables as a quantitative proxy for boundary strength. The pause duration values were binned into two groups: (a) pause duration longer than 500 ms (for strong boundaries) and (b) pause duration shorter than 500 ms (for weak boundaries). Preliminary investigations with different bin ranges pause durations showed no advantage over the range above.

In order to identify effects on the IP final syllable duration that were due to corpus-specific frequency distributions, word frequencies of the DIRNDL corpus were included as a control factor, binned into two categories for low and high frequency (the low-high frequency cutoff was 10 occurrences per word and speaker).

3.2. Language model

The ID measure used in this study was surprisal, which is frequently used in psycholinguistic studies. The trigram surprisal model is given in Equation (2).

$$S(syllable_i) = -log_2 P(syllable_i | syllable_{i-1}, syllable_{i-2}) \quad (2)$$

We acknowledge that word-based contextual predictability is an important meaning-based measure. However, since we are interested in the effects of surprisal and boundary strength on the duration of the IP final syllable, trigram surprisal values for the preceding syllabic context were obtained from the output of the LMs. Recall that word boundaries are represented in the LMs. Surprisal was log-transformed due to positive skewness.

3.3. Prosodic model

The prosodic model that accounted for variability in the duration of the IP final syllables contained information about prosodic prominence in terms of pitch accents (accented vs. unaccented), pitch accent type (L*H, H*L, !H*L) and boundary strength (strong vs. weak). Due to practical limitations, distinctions between primary and secondary stressed syllables were not taken into account.

4. Results

4.1. Descriptive statistics

The majority of the analyzed syllables was unaccented (n = 1116; total 1425). On average, unaccented syllables were shorter (M = 259 ms; SD = 92 ms) than accented syllables (M = 383 ms; SD = 104 ms), which is expected. Syllable duration was longest for !H*L accents (M = 408 ms; SD = 112 ms), shorter for H*L (M = 384 ms; SD = 115 ms) and L*H (M = 365 ms; SD = 83 ms), and even shorter for unaccented syllables (M = 259 ms; SD = 92 ms). There were two levels of prosodic boundary strength – strong and weak. Duration was shorter for syllables preceding a strong IP boundary (M = 279 ms; SD = 119.00) and longer for syllables before a weak IP boundary (M = 292 ms; SD = 101 ms). Syllables in high surprisal contexts were longer (M= 375 ms; SD = 111 ms) than syllables in low surprisal contexts (M= 236 ms; SD = 69 ms).

4.2. Linear mixed-effects model

We calculated Pearson's r correlations between IP final syllable durations and the corresponding surprisal values. Syllable duration and surprisal were significantly correlated (r = 0.48; t(1423) = 21.19; $p < .001$). The correlation is shown in Figure 1.

Following the results of the correlation analysis we calculated LMMs using the lme4 (1.1-12) [33] and lmerTest packages (2.0-33) [34]. The backward model selection method with maximal random structure was applied to identify the model that had the best fit for the data [35]. We included random intercepts for the random effects, and random slopes for all fixed effects. In the case of convergence errors we reduced the maximal random structure stepwise. First, we removed random slopes, and then, if necessary, random intercepts. Significance of fixed effects was evaluated by performing maximum likelihood t-tests using Satterthwaite approximations to degrees of freedom.

A collinearity analysis was performed to identify potential dependencies between the factors. Trigram surprisal and word frequency (based on word occurrences in the DIRNDL corpus) were only weakly correlated (r = 0.15). Trigram surprisal and accent showed a weak positive correlation (r = 0.36), with accented syllables showing higher trigram surprisal values than unaccented syllables. Boundary strength and accent were slightly negatively collinear (r = -0.11). This implies that unaccented syllables tend to occur at stronger IP boundaries. Gender and accent, on the other hand, are slightly positively collinear (r = 0.14), indicating that male speakers realize more unaccented IP final syllables than female speakers.

As a result of the collinearity analysis we entered as predictors trigram SURPRISAL of the preceding syllabic context, ACCENT with 4 factor levels (!H*L, H*L, L*H, unaccented), BOUNDARY STRENGTH with 2 factor levels (strong, weak) as well as their interactions, and included the control variables DIRNDL WORD FREQUENCY with 2 factor levels (high, low) and GENDER with 2 factor levels (male, female). The continuous dependent variable SYLLABLE DURATION was log-transformed due to positive skewness. All categorical variables were treatment coded. The maximal random structure included random intercepts for SPEAKER, SYLLABLE IDENTITY (which reflects the segmental make-up of the syllable), and WORD IDENTITY, as well as random slopes for all fixed effects. Because of convergence errors the model was simplified in a backward selection procedure. The random structure was reduced removing random slopes as described above. When the model converged, the predictors DIRNDL WORD FREQUENCY and GENDER as well as the interactions ACCENT*BOUNDARY STRENGTH and ACCENT*SURPRISAL did not explain variance in the data and were therefore removed from the model structure. Stepwise simplification resulted in a final model with random intercepts for SPEAKER, SYLLABLE IDENTITY and WORD IDENTITY. The final model structure is given in Equation (3).

$$\begin{aligned} SyllableDuration \sim\ & Surprisal + BoundaryStrength + \\ & Accent + Surprisal * BoundaryStrength + \\ & (1|SyllableIdentity) + (1|WordIdentity) + \\ & (1|Speaker) \end{aligned} \quad (3)$$

The coefficients, t-test, and p-values are presented in Table 1. All factors were significant in explaining variability in the duration of IP final syllables. As expected, ACCENT and trigram SURPRISAL significantly lengthen the IP final syllable duration (Figure 1). Moreover, we found a significant effect of accent type on IP final syllable duration (Figure 2). Syllable duration is the longest for !H*L accents and significantly shorter for H*L and L*H. Unaccented syllables have the shortest duration. Furthermore, duration is negatively affected by BOUNDARY STRENGTH. IP final syllables were found to be shorter before strong boundaries than before weak boundaries (Figure 3).

Regarding the interaction between SURPRISAL and BOUNDARY STRENGTH, syllable duration becomes longer with increasing surprisal, but this increase in duration is significantly greater before strong than before weak boundaries (Figure 4).

The marginal pseudo-R^2 indicating how much variance is explained by the fixed factors showed that the baseline prosodic factors (phrasal ACCENT and BOUNDARY STRENGTH explain 3.4% of the variance in the IP final syllable duration. Note that whereas this amount of variance explained may seem small, it is expected: when language redundancy is strong, the listener only needs a weak bottom-up 'checking' signal [36]. The explained variance increases by 7.7% when SURPRISAL is included in the additive model, and by a further 1% when SURPRISAL interacts with BOUNDARY STRENGTH. The conditional pseudo-R^2 for the variance explained by both fixed and random effects equaled 68% in the final model.

Table 1: *IP final syllable duration in German: regression coefficients, standard error (SE), and statistical output of LMM analysis including syllable trigram surprisal based on the preceding context.*

Terms	Coeff.	SE	t-value	p-value
Surprisal	0.04	0.01	6.15	<.001
Accent(no)	0.09	0.01	-6.32	<.001
Accent(!H*L)	0.04	0.02	2.50	<.05
Accent(H*L)	0.03	0.02	1.72	=.085
Boundary(str/wk)	-0.04	0.09	-4.47	<.001
Boundary*Surprisal	0.01	0.01	2.58	<.01

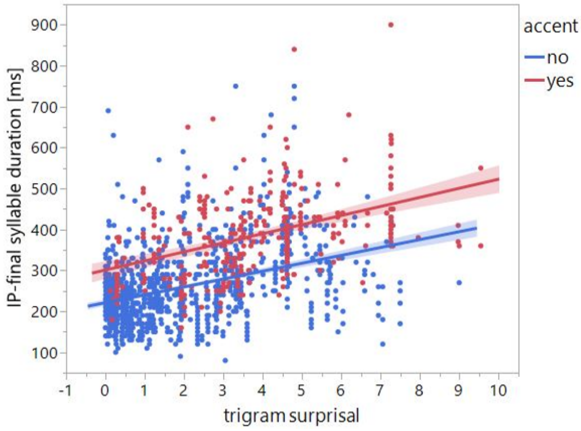

Figure 1: *The effect of trigram surprisal and accent on IP final syllable duration.*

5. Discussion and conclusion

This study investigated the impact of prosodic structure and information-theoretic factors on IP final syllable duration in German read speech. Following previous work in psycholinguistics and computational linguistics, predictability was defined as trigram surprisal and measured at the syllable level based on the preceding context. The prosodic model used in this study consists of two levels of prominence (accented, unaccented), three levels of accent type (!H*l, H*L, L*H) and two levels of boundary strength (strong, weak).

With regard to contextual predictability, our results confirm the first hypothesis that surprisal is predictive of IP final syllable duration. We found that syllables that are difficult to predict from context have an increased duration compared to easily predictable syllables. These findings are in accordance with previous studies on duration and predictability [3, 37, 7, 38, 8, 5] and are an important addition to the studies on syllable duration in the context of information theory. This is because they are based on IP final syllables only, which to the best of our knowledge have not yet been investigated in previous research. Phrasal accent type and boundary strength were found to be predictive of IP final syllable duration as well. As expected, syllables carrying phrasal accent are longer than unaccented syllables. Moreover, syllables occurring before strong boundaries were found to be shorter compared to syllables before weak boundaries.

With respect to our second hypothesis, phrasal accent and surprisal were found to have independent effects on syllable duration. Surprisal had a larger overall effect size than prosodic

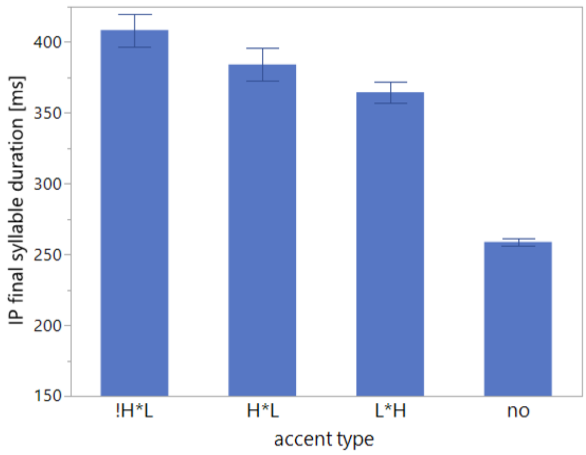

Figure 2: *The effect of accent type on IP final syllable duration.*

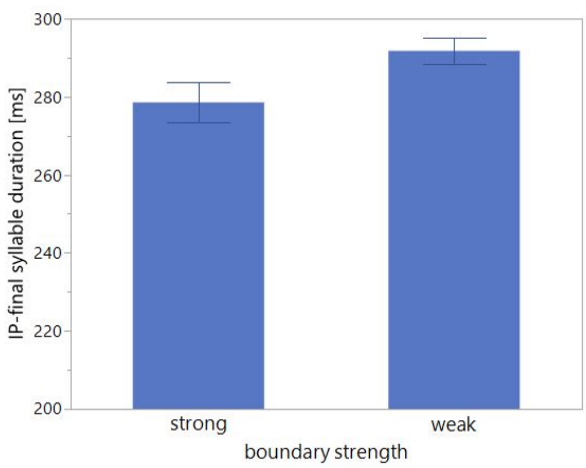

Figure 3: *The effect of boundary strength on IP final syllable duration.*

factors, indicating that the modulation between surprisal and phonetic encoding by means of prosody is, indeed, not comprehensive. This suggests that the prosodic structure (at least the phrasal accent) is not implementing much of the significant effects of redundancy, supporting the weak version of the Smooth Signal Redundancy hypothesis. In other words, smoothing is being carried out either by direct reference to redundancy factors or by boundary strength, which, among other things, causes an increase in duration of phrase-final syllables.

As for our third hypothesis, we also found an interaction between boundary strength and surprisal, indicating that these factors complement each other in explaining syllable duration variability: Syllable durations become longer with increasing surprisal, but this increase in duration is significantly greater before strong than before weak boundaries. This modulation of pre-boundary syllable duration by surprisal is observed above and beyond the well-established phrase-final lengthening effect.

To conclude, the findings in this study are generally compatible with the weak version of the Smooth Signal Redundancy hypothesis [3, 4]. They highlight the complex interactions

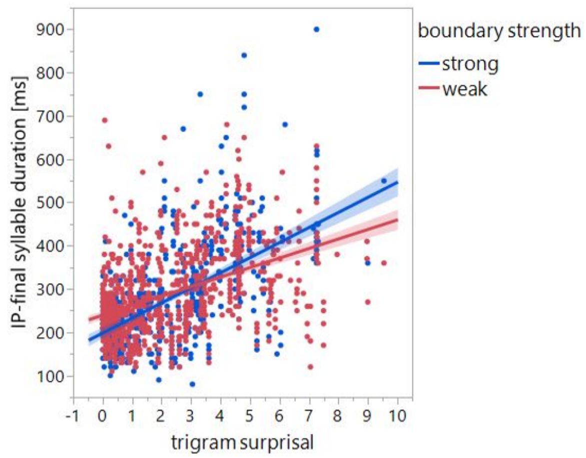

Figure 4: *The effect of surprisal and boundary strength on IP final syllable duration.*

in spoken language between segmental, suprasegmental, and information-theoretic factors related to predictability in context.

6. Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 232722074 – SFB 1102 ‘Information Density and Linguistic Encoding’ at Saarland University. We thank Raphael Baier, Kirstin Kolmorgen, and Anna Konobelkina for technical support, and three anonymous reviewers for constructive criticism.

7. References

[1] J. B. Pierrehumbert, “Exemplar dynamics: Word frequency, lenition and contrast,” in Frequency and the Emergence of Linguistic Structure, J. Bybee and P. Hopper, Eds. Benjamins, Amsterdam, 2001, pp. 137–157.

[2] T. F. Jaeger, “Redundancy and reduction: Speakers manage syntactic information density,” Cognitive Psychology, vol. 61, no. 1, pp. 23–62, 2010.

[3] M. Aylett and A. 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, pp. 31–56, 2004.

[4] ——, “Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei,” Journal of the Acoustical Society of America, vol. 119, no. 1, pp. 30–48, 2006.

[5] Z. Malisz, E. Brandt, B. Möbius, Y. Oh, and B. Andreeva, “Dimensions of segmental variability: interaction of prosody and surprisal in six languages,” Frontiers in Communication, vol. 3, pp. 1–18, 2018.

[6] E. Brandt, Information density and phonetic structure: Explaining segmental variability, 2019, PhD dissertation, Saarland University.

[7] S. Gahl, Y. Yao, and K. Johnson, “Why reduce? Phonological neighborhood density and phonetic reduction in spontaneous speech,” Journal of Memory and Language, vol. 66, no. 4, pp. 789–806, 2012.

[8] H. Tily, S. Gahl, I. Arnon, N. Snider, A. Kothari, and J. Bresnan, “Syntactic probabilities affect pronunciation variation in spontaneous speech,” Language and Cognition, vol. 1, no. 2, pp. 147–165, 2009.

[9] U. Cohen Priva, “Informativity affects consonant duration and deletion rates,” Laboratory Phonology, vol. 6, no. 2, pp. 243–278, 2015.

[10] R. J. J. H. van Son and J. P. H. van Santen, "Duration and spectral balance of intervocalic consonants: A case for efficient communication," Speech Communication, vol. 47, pp. 100–123, 2005.

[11] 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), 2016, pp. 350–354.

[12] J. H. Gaitenby, "The elastic word." Haskins report, Status Report 2:3.1–3.12., 1965.

[13] D. Klatt, "Vowel lengthening is syntactically determined in connected discourse," Journal of Phonetics, vol. 3, no. 3, pp. 129–140, 2006.

[14] P. Price, M. Ostendorf, S. Shattuck-Hufnagel, and C. Fong, "The use of prosody in syntactic disambiguation," The Journal of the Acoustical Society of America, vol. 90, no. 2-4, pp. 2956–2970, 1991.

[15] J. Hale, "Information-theoretical complexity metrics," Language and Linguistics Compass, pp. 1–16, 2016.

[16] C. Wightman, S. Shattuck-Hufnagel, M. Ostendorf, and P. Price, "Segmental durations in the vicinity of prosodic phrase boundaries," Journal of the Acoustical Society of America, vol. 92, pp. 1707–1717, 1992.

[17] M. Wagner, "Prosody and recursion," Ph.D. dissertation, Massachusetts Institute of Technology, 2005.

[18] D. Ladd, "Declination "reset" and the hierarchical organization of utterances," The Journal of the Acoustical society of America, vol. 84, no. 4, pp. 530–544, 1988.

[19] S. Frota, Prosody and focus in European Portuguese. Phonological phrasing and intonation, 01 2000.

[20] J. Krivokapić and D. Byrd, "Prosodic boundary strength: An articulatory and perceptual study," Journal of phonetics, vol. 40, pp. 430–442, 05 2012.

[21] K. Eckart, A. Riester, and K. Schweitzer, "A discourse information radio news database for linguistic analysis," in Linked Data in Linguistics: Representing and Connecting Language Data and Language Metadata, C. Chiarcos, S. Nordhoff, and S. Hellmann, Eds. Springer, 2012, pp. 65–75.

[22] A. Björkelund, Ö. Çetinooglu, A. Falenska, R. Farkas, T. Mueller, W. Seeker, and Z. Szántó, "Introducing the IMS-Wroclaw-Szeged-CIS entry at the SPMRL 2014 Shared Task: Reranking and Morpho-syntax meet Unlabeled Data," in Proceedings of the First Joint Workshop on Statistical Parsing of Morphologically Rich Languages and Syntactic Analysis of Non-Canonical Languages. Dublin, Ireland: Dublin City University, August 2014, pp. 97–102. [Online]. Available: http://www.aclweb.org/anthology/W14-6110

[23] S. Rapp, "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" (Moscow, Russia), 1995.

[24] J. Mayer, "Transcription of German intonation—the Stuttgart system," Institute of Natural Language Processing, University of Stuttgart, Tech. Rep., 1995.

[25] G. Möhler, A. Schweitzer, M. Breitenbücher, and M. Barbisch, "IMS German Festival (Version: 1.2-os)," 2000, University of Stuttgart: Institut für Maschinelle Sprachverarbeitung (IMS), retrieved 2020-01-02.

[26] A. Stolcke, "SRILM – an extensible language modeling toolkit," Proceedings of Interspeech, vol. 2, no. Denver, Colorado, pp. 901–904, 2002.

[27] G. A. Castellucci and D. Goldenberg, "Gradient prosodic boundary perception and recursion in syntactic disambiguation," The Journal of the Acoustical Society of America, vol. 135, no. 4, pp. 2197–2197, 2014.

[28] J.-Y. Choi, "Pause length and speech rate as durational cues for prosody markers," The Journal of the Acoustical Society of America, vol. 114, no. 4, pp. 2395–2395, 2003.

[29] P. Hansson, Prosodic Phrasing in Spontaneous Swedish, 2003, PhD dissertation, In Travaux de l'Institut de Linguistique de Lund 43.

[30] M. Horne, E. Strangert, and M. Heldner, "Prosodic boundary strength in swedish: Final lengthening and silent interval duration," in Proc. XIIIth ICPhS, 1995, pp. 170–173.

[31] A. A. Sanderman and R. Collier, "Prosodic phrasing at the sentence level," in Producing speech: Contemporary Issues. For Katherine Safford Harris., F. Bell-Berti and L. Raphael, Eds. American Institute of Physics, New York, 1995, pp. 321—-332.

[32] M. Swerts, "Prosodic features at discourse boundaries of different strength." The Journal of the Acoustical Society of America, vol. 101, pp. 514—-521, 1997.

[33] D. Bates, M. Maechler, B. Bolker, S. Walker, R. H. Bojesen Christensen, N. Singmann, and B. Dai, Package 'lme4': Linear mixed-effects models using Eigen and S4, February 20, 2015. [Online]. Available: http://cran.r-project.org/web/packages/lme4/lme4.pdf

[34] A. Kuznetsova, P. Bruun Brockhoff, and R. Haubo Bojesen Christensen, lmerTest: Tests in Linear Mixed Effects Models, 2016, r package version 2.0-33. [Online]. Available: https://CRAN.R-project.org/package=lmerTest

[35] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maximal," Journal of Memory and Language, vol. 68, no. 3, pp. 255–278, 2013.

[36] M. P. Aylett, "Stochastic suprasegmentals – relationships between redundancy, prosodic structure and care of articulation in spontaneous speech," Ph.D. dissertation, Department of Linguistics, University of Edinburgh, 2000.

[37] A. Bell, J. Brenier, M. Gregory, C. Girand, and D. Jurafsky, "Predictability effects on durations of content and function words in conversational English," Journal of Memory and Language, vol. 60, no. 1, pp. 92–111, 2009.

[38] D. Jurafsky, A. Bell, M. Gregory, and W. D. Raymond, "Probabilistic relations between words: Evidence from reduction in lexical production," in Frequency and the Emergence of Linguistic Structure, J. Bybee and P. Hopper, Eds. Amsterdam: Benjamins, 2001, pp. 229–254.