FREQUENCY EFFECTS AND PROSODIC BOUNDARY STRENGTH

Tina Bögel[1] and Alice Turk[2]

[1]University of Konstanz; [2]University of Edinburgh
Tina.Boegel@uni-konstanz.de; a.turk@ed.ac.uk

ABSTRACT

It has long been recognized that spoken language does not necessarily provide reliable cues to indicate word boundaries in all contexts. [19] hypothesized that prosodic boundary structure is planned in order to achieve smooth signal redundancy ([1], [2]). On this view, speakers manipulate the strength of boundaries between words in order to make the recognition of each word in an utterance equally likely. Prosodic boundary strength is assumed to inversely relate to language redundancy, i.e., the likelihood of recognition on the basis of non-acoustic information, including the likelihood of the syntactic structure, lexical word frequency and word bigram frequency.

This paper explores the impact of syntactic, word-bigram and word frequency measures on the placement and strength of prosodic word and phrase boundaries in English.

Keywords: frequency effects, prosodic boundary strength, duration, smooth signal redundancy

1. INTRODUCTION

Following [14], the *Smooth Signal Redundancy Hypothesis* [1] assumes that the recognition likelihood of linguistic items is spread evenly throughout the utterance to ensure robust and efficient communication between speaker and listener (see also [12]). The identity of these linguistic items is assumed to be signalled on two levels: a) via language redundancy, i.e., the likelihood of recognition via lexical, syntactic, pragmatic, and semantic factors, and b) via acoustic redundancy, i.e., the recognition likelihood based on acoustic salience. These two levels are assumed to inversely correlate with each other: If language redundancy is high, the acoustic saliency of the produced section of speech is low, and vice versa (see also [13], [5], [16], [15], a.o.).

In experiments on phrase-medial syllable durations and vowel quality [1], [2], and [3] showed that smooth signal redundancy is achieved through an inverse relationship between language redundancy and acoustic redundancy, i.e.: Speakers speak with high acoustic saliency during unpredictable (infrequent) sections of speech, but lower acoustic saliency if the immediate context is predictable (frequent). [2], for example, found that word frequency and the likelihood of a particular syllable based on previous information in form of previous syllables and givenness correlated inversely with syllable duration.

Similarly, [10] showed that highly frequent function words and function words with a high probability given previous and following material are more likely to be acoustically reduced, where increased likelihood correlated inversely with function word duration. [5] found a significant effect on word duration given the *following* word, and [16] also noted an effect of bigram frequency, but also of repetition, on stem and suffix duration. All of these findings are consistent with the hypothesis that language redundancy correlates inversely with acoustic saliency in the form of duration.

Based on evidence of shared effects of language redundancy and prosodic prominence on syllable duration and vowel quality [1], [2] and [3] proposed that smooth signal redundancy is controlled via prosodic prominence structure. [19] additionally proposed that prosodic boundary structure plays a similar role. On this view, stronger prosodic boundaries (e.g., longer duration of the preceding rhyme, pause (if any), and following onset, resulting in an overall longer boundary-related interval) are expected to occur where language redundancy is low. This view provides a unified explanation for effects of syntax, semantics and utterance length on prosodic boundary occurrence and strength, and is supported by findings in [21] and [8] (a.o.) but has not yet been tested experimentally.

By means of a production experiment, this paper investigates the relationship between boundaries and four measures of language redundancy: a) syntactic frequency, b) lexical (word) frequency, c) bigram (verb-adjective and adjective-noun) frequencies, and d) the ratio between these two bigram frequencies. In addition to the placement of intonational boundaries, we also examine the impact of the respective language redundancy measures on duration.

2. METHOD

2.1. Materials

The material included 58 target sentences where each sentence was part of a set of a minimal quadruplet or a minimal doublet. Each member of a set included a verb-adjective-noun sequence, where the verb and noun were each either highly frequent (f) or infrequent (i). Measures of lexical frequency were obtained via WebCelex's Cobuild frequency [4] (raw numbers, Verbs: f $>$ 2000, i $<$ 200; Nouns: f $>$ 3000, i $<$ 100). Each frequent verb was matched with an infrequent verb with an identical rhyme (e.g., [eə] in *wear (f)* and *pare (i)*), and each frequent noun was matched with an infrequent noun with an identical onset (e.g., [fɑ] in *farmers (f)* and *farthings (i)*. The adjective was the same for all members of a set.

Table 1: Quadruplet combining frequent (f) and infrequent (i) (partly homophonous) verbs and nouns.

Freq.	Matched quadruplet
ff:	*Whatever you* ***wear***$_{\mathbf{f}}$ ***thin farmers***$_{\mathbf{f}}$ *can't pay for it*
fi:	*Whatever you* ***wear***$_{\mathbf{f}}$ ***thin farthings***$_{\mathbf{i}}$ *can't pay for it*
if:	*Whatever you* ***pare***$_{\mathbf{i}}$ ***thin farmers***$_{\mathbf{f}}$ *can't pay for it*
ii:	*Whatever you* ***pare***$_{\mathbf{i}}$ ***thin farthings***$_{\mathbf{i}}$ *can't pay for it*

Each sentence was syntactically ambiguous in that the adjective could either be combined with the preceding verb (left attachment, resultative function) or the following noun (right attachment, modifying function). The resulting differences in syntactic phrasing are disambiguated via prosodic boundary placement (c.f. [17], [11]). With reference to the ICE-GB corpus [9] and the Brown corpus [7] we determined that adjectives occur much more often in a modifying function than in a resultative construction (across both corporas: modifying ca. 77%; resultative ca. 4%; others: ca. 19%). Table 2 shows the syntactic and expected prosodic grouping of the V-A-N sequence and the related syntactic frequency (f/i). Prosodic boundaries are indicated by the %-sign. In order to allow for unbiased boundary placement, no indicator of phrasing was given in the experiment.

Table 2: Variable adjective association resulting in syntactic ambiguity

Syntax	Sentence
[V] [A N] (f)	*Whatever you* ***wear % thin farmers*** *can't ...*
[V A] [N] (i)	*Whatever you* ***wear thin % farmers*** *can't ...*

In addition to lexical and syntactic frequency, a third measure of language redundancy was obtained by calculating the word bigram frequency for a) the verb-adjective (VA) sequence, and b) the adjective-noun (AN) sequence via Google. These bigram frequencies were furthermore used to calculate the ratio between the two bigram frequency values by dividing the VA-bigram-frequency by the AN-bigram-frequency. Since the bigram frequencies and the ratios differed considerably in size (e.g., from 9 to 45M for the VA frequencies), we normalized the data by dividing it into three parts, drawing cut-off points at the fourth and the sixth of 10-quantiles. All data below 40% was categorized as 'low frequency' and all data above 60% was categorized as 'high frequency' (with 20% of the data being used as buffer to avoid adjacent measurements for 'high' and 'low'). The ratio was normalized in the same way to indicate whether the VA bigram frequency was considerably smaller or larger in comparison to the AN bigram frequency.

2.2. Participants

23 participants took part in the study (age mean: 23.4, 14 females, all of them students at the University of Edinburgh) and received a small payment. All were analysed for intonational phrase boundary placement. In addition, a subgroup of 10 participants were used for a more fine-grained durational analysis of boundary strength (age mean: 25.7, 5 male).

2.3. Procedure

Recordings were made at the University of Edinburgh in a sound-treated studio using a high quality microphone. Recordings were digitized with a sampling rate of 44.1 Kz and a bit depth of 16. All target sentences (without comma) were presented in two repetitions and in randomised order to each participant, resulting in a total of 2668 sentences (116/speaker) for the boundary placement analyses, of which 356 were also used for the durational analysis of boundary strength.

2.4. Data analysis

2.4.1. Syntactic choice - placement of prosodic boundary

First, each utterance was judged by an expert for the location of the intonational phrase boundary (ToBI break index: 4 [18]) produced within the V-A-N sequence; i.e., which syntactic structure the speaker assumed. A second listener judged a subset (40%) of the data, and 100% agreement was obtained.

Figure 1: Part of the abstract annotation scheme used for the durational analysis

Verb end		**Adjective start**		**Adjective end**		**Noun start**	
`V-Rh`	rhyme	`A-On`	onset	`A-Rh`	rhyme	`N-On`	onset
`V-Co`	coda	`A-Cl`	closure	`A-Co/Co1/Co2`	coda/coda part 1/2	`N-Cl`	closure
`V-ORh`	with part of onset			`A-ORh`	with part of onset		
`R-V-...`	with onset release			`A-Nu`	nucleus, not coda		
				`R-A-...`	with onset release		

Intermediate (`IM1` **and** `IM2`)		***Comment:***
`...-R`	release	*Might include aspiration*
`...-P`	pause	*Missing pause (P) is only indicated if there is an intonational phrase boundary*
`...-RP`	release and pause	*Both -P/-RP are only indicated if there is no closure following*

In a second step we determined the extent to which frequency effects had an impact on the speaker's choice of the boundary position by means of an independent directional Wilcoxon rank sum test.

2.4.2. Duration and language redundancy

A further goal of this study was to determine the impact of lexical and syntactic frequency on boundary strength as measured by the duration of a boundary-related interval. In order to be able to compare boundary strength as measured durationally under identical conditions, consistency across repetitions and sets was required. That is, we only analysed those sentences where the choice of syntactic structure was identical across a complete set of four sentences and both repetitions. While consistency across repetitions was quite high (mean: 82% with lowest: 58% and highest: 93%), consistency across all members of a quadruplet was lower (356 sentences across the subset of 10 speakers).

These 356 utterances were hand-coded in Praat [6] for *duration* and *repetition* (1 or 2). Since we wanted to measure the impact of language redundancy on the strength of prosodic boundaries, we annotated intervals that spanned parts of the VA and AN sequence that are known to be most affected by boundary-related lengthening [22], under the constraints of segmentation reliability [20]. That is, an interval that included the coda or rhyme of the first word in the VA or AN sequence, the pause (if any), and the onset consonant constriction of the second word. For example, for the VA sequence *cropped flat*, we annotated the interval from the release of the /k/ in *cropped* to the release of the /f/ in *flat*.

In order to determine which parts of a particular interval are affected most by frequency effects, we furthermore annotated the three parts of each interval separately. In doing so, we used an abstract annotation scheme (Figure 1), which allowed us to explicitly mark the segmental composition of a sub-interval and enabled us to include the sub-intervals for statistical analysis. The resulting duration intervals were extracted automatically and analysed with respect to the lexical, syntactic, and bigram frequencies, and the bigram ratio, as discussed above. For the statistical analysis of the durational measurements, we used a linear mixed effects regression model (lmer), with the respective frequency effects as fixed factors and subjects and items as random factors (adjustment of intercepts; a model with random slopes did not converge).

2.5. Results

The following results are calculated on the basis of repetition 1 (1314 sentences for the placement of the prosodic boundary, 178 for the durational analysis).

2.5.1. Syntactic choice

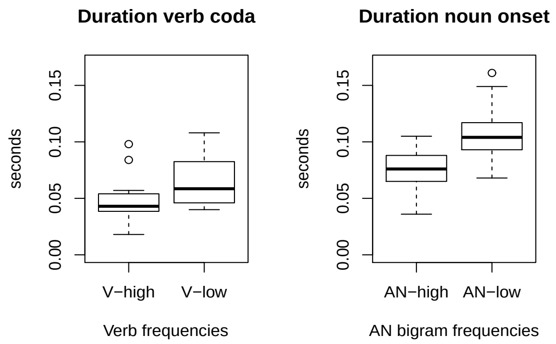
Contrary to the expectations based on syntactic frequency, speaker's choice between [V]%[AN] and [VA]%[N] was almost equally distributed (V%AN=55%, VA%N=45%). A Wilcoxon rank sum test showed that speakers prefer the infrequent VA%N structure in the presence of a highly frequent verb ($p < 0.001$, r=0.19) or a high VA bigram frequency ($p < 0.001$, r=0.22). The frequent structure V%AN is preferred if the noun is highly frequent ($p < 0.05$, r=0.07) or if there is a high AN bigram frequency ($p < 0.001$, r=0.09).

2.5.2. Duration results

We calculated the duration for the whole V-A and A-N intervals and the individual sub-intervals for both syntactic/prosodic boundary placement patterns: a) [V] [A N] $\leftrightarrow$ V%AN, where the prosodic boundary is placed after the verb, and b) [V A] [N] $\leftrightarrow$ VA%N, where the boundary is placed between the adjective and the noun.

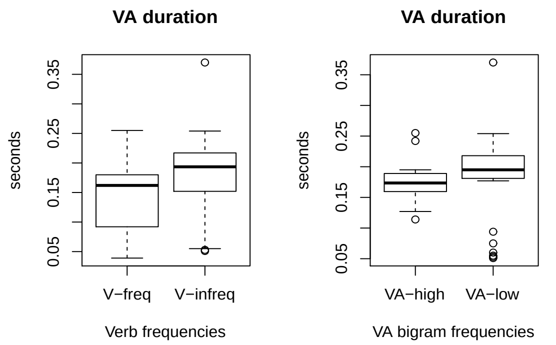
For V%AN, the overall interval analysis did not return any significant results. For the sub-intervals, a linear mixed effects regression model showed that the duration of the verb coda increased by 15 ms (β=0.015) if the verb was infrequent (SE=0.006, t=2.5, $p < 0.05$) and that the duration of the noun onset increased by 10 ms if the AN bigram frequency was low (β=0.01, SE=0.004, t=2.3, $p < 0.05$)

Figure 2: Durational measurements of the verb coda and the noun onset (V%AN)



For VA%N, with the boundary after the adjective, the duration of the overall VA interval increased with lower verb frequency (β=0.023, SE= 0.009, t=2.65, $p < 0.05$) and lower VA bigram frequency (β=0.029, SE=0.01, t=2.96, $p < 0.01$).

Figure 3: Durational measurements of the VA interval (pattern VA%N)



When looking at the sub-intervals of the VA interval, the same effect of lexical verb frequency (β=0.023, SE= 0.008, t=2.8, $p < 0.05$) and VA bigram frequency (β=0.023, SE=0.008, t=2.8, $p < 0.01$) was found with the verb coda, but not with the onset of the adjective. Furthermore, both intervals, the VA interval and the sub-interval of the verb coda were significantly shorter if the VA bigram frequency was higher than the AN bigram frequency (VA-interval: β=-0.019, SE= 0.009, t=-2, p=0.05; verb coda: β=-0.029, SE= 0.007, t=-3.97, $p < 0.001$). The opposite effect was found for the noun onset following the phrase boundary; duration increased when the AN frequency was low (in comparison to the VA frequency) (β=0.018, SE=0.005, t=3.27, $p < 0.01$).

3. DISCUSSION AND CONCLUSION

Results are consistent with the assumptions made by the *Smooth Signal Redundancy Hypothesis* in that an inverse relationship between language redundancy and durational measurements influences the prosodic boundary related intervals. Lexical verb frequency affected the verb coda in V%AN and the VA interval in VA%N. AN bigram frequency affected the AN interval in V%AN, the VA bigram frequency affected the interval in VA%N. The VA%N pattern furthermore showed an interaction between the two bigram frequencies, in that the duration of the VA interval decreased if the VA bigram frequency was high and the following AN bigram frequency was low, which reduces the chance of the adjective being wrongly grouped with the noun by the listener.

In contrast to corpus-based investigations, speakers' choice of intonational phrase boundary placement was relatively evenly distributed across both possibilities. This may be an artefact of our experimental situation in which resultative meanings were suggested by the linear word order in our materials (in contrast to real world situations, where resultative contexts occur less often). With respect to the durational measurements, the two patterns showed clear differences in that the results for the corpus-infrequent pattern VA%N were generally more reliable and included an interaction between bigram frequencies, in that frequent VA bigrams were significantly shorter if the following AN frequency was low.

We conclude that frequency effects on different levels influence the strength of the intonational phrase boundaries and intonational-phrase medial word boundaries as measured by the durations of the boundary-related intervals. These effects were more reliable for the VA%N structures (infrequent syntax), and appeared to be driven more by language redundancy relating to the verb, rather than the noun.

4. ACKNOWLEDGEMENTS

We would like to thank the Zukunftskolleg of the University of Konstanz for a small grant that supported the research presented in this paper, and Marisa Flecha-Garcia for assistance with participant recruitment and data collection.

5. REFERENCES

[1] Aylett, M. 2000. *Stochastic Suprasegmentals: Relationships between Redundancy, Prosodic Structure and Care of Articulation in Spontaneous Speech*. PhD thesis University of Edinburgh.

[2] Aylett, M., Turk, A. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech* 47(1), 31–56.

[3] Aylett, M., Turk, A. 2006. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America* 119(5), 3048–3058.

[4] Baayen, R., Piepenbrock, R., Gulikers, L. 2001. *WebCelex*. Online resource: http://celex.mpi.nl/. Max Planck Institute for Psycholinguistics.

[5] Bell, A., Brenier, J. M., Gregory, M., Girand, C., Jurafsky, D. 2009. Predictability effects on durations of content and function words in conversational english. *Journal of Memory and Language* 60(1), 92–111.

[6] Boersma, P., Weenink, D. 2013. Praat: doing phonetics by computer [computer program, Version 5.3.56]. available at http://www.praat.org/ [retrieved 15.09.2013].

[7] Francis, W. N., Kučera, H. 1964. A standard corpus of present-day edited American English, for use with digital computers. Technical report Department of Linguistics, Brown University, Providence, Rhode Island, US.

[8] Gahl, S., Garnsey, S. 2004. Knowledgef of grammar, knowledge of usage. Syntactic probabilities affect pronunciation variation. *Language* 80, 748–775.

[9] *ICE-GB corpus*, 1998. ˙Online resource: http://ice-corpora.net/ice/icegb.htm.

[10] Jurafsky, D., Bell, A., Gregory, M., Raymond, W. 2001. Probabilistic relations between words: Evidence from reduction in lexical production. In: Bybee, J., Hopper, P., (eds), *Frequency and the Emergence of Linguistic Structure*. Amsterdam: John Benjamins 229–254.

[11] Lehiste, I., Olive, J. P., Streeter, L. A. 1976. Role of duration in disambiguating syntactically ambiguous sentences. *The Journal of the Acoustical Society of America* 60, 1199–1202.

[12] Levy, R., Jaeger, T. F. 2007. Speakers optimize information density through syntactic reduction. In: Schlökopf, B., Platt, J., Hoffman, T., (eds), *Advances in Neural Information Processing Systems (NIPS)* volume 19. Cambridge, MA: MIT Press 849–856.

[13] Lieberman, P. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6, 172–187.

[14] Lindblom, B. 1990. Explaining phonetic variation: A sketch of the H&H theory. In: Hardcastle, W. J., Marchal, A., (eds), *Speech Production and Speech Modelling* volume 55. Dordrecht: Kluwer Academic Publishers 403–439.

[15] Pluymaekers, M., Ernestus, M., Baayen, R. H. 2005. Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica* 62, 146–159.

[16] Pluymaekers, M., Ernestus, M., Baayen, R. H. 2005. Lexical frequency and acoustic reduction in spoken dutch. *Journal of the Acoustical Society of America* 118, 2561–2569.

[17] Price, P., Ostendorf, M., Shattuck-Hufnagel, S., Fong, C. 1991. The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America* 90(6), 2956–2970.

[18] Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J. 1992. Tobi: A standard for labeling english prosody. *Proceedings of the 1992 International Conference on Spoken Language Processing* Banff.

[19] Turk, A. 2010. Does prosodic constituency signal relative predictability? A smooth signal redundancy hypothesis. *Laboratory Phonology* 227–262.

[20] Turk, A., Nakai, S., Sugahara, M. 2006. Acoustic segment durations in prosodic research: a practical guide. In: Sudhoff, S., Lenertová, D., Meyer, R., Pappert, S., Augurzky, P., Mleinek, I., Richter, N., Schliesser, J., (eds), *Methods in Empirical Prosody Research*. Berlin, New York: De Gruyter 1–28.

[21] Watson, D. G., Breen, M., Gibson, E. 2006. The role of syntactic obligatoriness in the production of intonational boundaries. *Journal of Experimental Psychology: Learning, Memory and Cognition* 32, 1045–1056.

[22] Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., Price, P. J. 1992. Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America* 91(3), 1707–1717.