Frequency of occurrence effects on pitch accent realisation

Katrin Schweitzer[1], *Michael Walsh*[1], *Bernd Möbius*[1,2], *Hinrich Schütze*[1]

[1]Institute for Natural Language Processing, University of Stuttgart
[2] Institute of Communication Sciences, University of Bonn

{katrin.schweitzer|michael.walsh}@ims.uni-stuttgart.de
hs999@ifnlp.org, moebius@ifk.uni-bonn.de

Abstract

This paper presents the results of a corpus study which examines the impact of frequency of occurrence of accented words on the realisation of pitch accents. In particular, statistical analyses explore this influence on pitch accent range and alignment. The results indicate a significant effect of frequency of occurrence on the relative height of L*H and H*L pitch accents and an also significant but more subtle effect on the alignment of L*H accents.

Index Terms: exemplar theory, pitch accents, frequency effects

1. Introduction

In recent years, a growing body of evidence has emerged which has established the effects that frequent usage has on linguistic units across a variety of linguistic domains [1, 2].

In the domain of prosody, frequency of occurrence has been found to affect the variability of syllable duration [3, 4] and pitch accents [5]. Although frequency effects have been found which influence pitch accent shape [5] the influence of word frequency on pitch accent realisations has yet to be examined. The research presented in this paper specifically targets this question. In particular the following questions are addressed:

1. Is there an effect of the frequency of occurrence of a word-accent pair on the realisations of the accent?
2. If so, what is the impact of frequency on these realisations?
3. What are the implications of such an effect for theories of lexical storage and post-lexical accenting?

In examining the realisation of pitch accent tokens, a parametric intonation model, known as PaIntE [6], is employed to extract meaningful parameters from pitch accented words of varying frequency in a speech corpus. ANOVAs are carried out on these individual parameters to determine if frequency of occurrence influences their behaviour. It is important to note that these ANOVAs incorporate additional shape-influencing factors, e.g. syllable onset size.

The results in this study are examined from an exemplar-theoretic perspective (see section 2). According to this framework all perceived linguistic exemplars (segments, syllables, words etc.) are stored and offer the speaker a resource which can be drawn upon for both production and perception. Effects of frequency therefore are important as they offer the speaker more production and categorisation options, and in the current context could potentially indicate pitch-accent storage in the lexicon.

The structure of this paper is as follows: Section 2 provides a short introduction to Exemplar Theory and motivates the study. The parametric intonation model [6] which captures pitch accent shape is then outlined in section 3. Section 4 introduces the methodology which is used in the analyses. Results are presented in section 5, and conclusions and opportunities for future research are discussed in section 6.

2. Exemplar Theory

Exemplar Theory is concerned with the idea that the acquisition of language is significantly facilitated by repeated exposure to concrete language input, and it has successfully accounted for a number of language phenomena, including diachronic language change and frequency of occurrence effects [2], the emergence of grammatical knowledge [7], syllable duration variability [3, 4], entrenchment and lenition [1], among others. Central to Exemplar Theory are the notions of exemplar storage, frequency of occurrence, recency of occurrence, and similarity. There is an increasing body of evidence which indicates that significant storage of language input exemplars, rich in detail, takes place in memory [8, 9, 10]. These stored exemplars are then employed in the categorisation of new input percepts. Similarly, production is facilitated by accessing these stored exemplars as production targets. Computational models of the exemplar memory also argue that it is in a constant state of flux with new inputs updating it and old unused exemplars gradually fading away [1].

Up to now, little exemplar-theoretic research has examined pitch accent prosody (but see [11] for memory-based prediction of pitch accents and prosodic boundaries and [5] for evidence of frequency effects on within-type pitch accent variability) and to the authors' knowledge this paper represents the first attempt, from a usage-based perspective, to examine the relationship between the frequency with which words and pitch accents combine and the effect this frequency has on pitch accent realisation. Given the considerable weight of evidence for the influence of frequency of occurrence effects in a variety of other linguistic domains it seems reasonable to explore such effects on pitch accent. In particular, if pitch accent production is autonomous and independent of the lexicon, in keeping with autosegmental theories of intonation, then an examination of pitch accent on the basis of lexical frequency should be unlikely to yield frequency effects. However, an exemplar-theoretic account would suggest that perceived lexical items could be stored with the accompanying perceived pitch accent, in which case pitch accents might not always operate in an autosegmental fashion.

The search for possible frequency of occurrence effects takes place with respect to pitch accent shapes captured by the parametric intonation model discussed next.

26 – 30 September 2010, Makuhari, Chiba, Japan

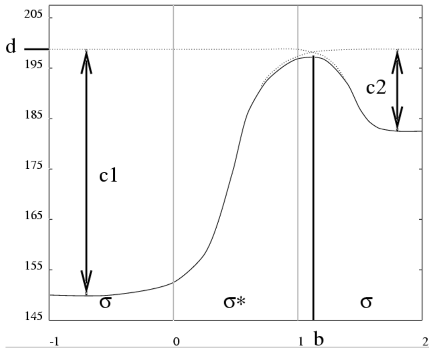

Figure 1: The PaIntE model function is the sum of a rising and a falling sigmoid with a fixed time delay. The time axis is normalised to the syllables' lengths. The parameters are calculated over the span of the accented syllable ($\sigma*$) and its immediate neighbours. Parameter *b* locates the peak of the accent, parameters *c1* and *c2* model the ranges of the rising and falling slopes, respectively, *d* corresponds to the actual height of the peak and parameters *a1* and *a2* (not displayed) denote the "amplitude-normalised" steepness of the rising and falling slopes [13].

3. The Parametric Representation of Intonation Events - PaIntE

The model approximates stretches of F_0 by employing a phonetically motivated model function [6]. This function operates on a three-syllable-window and consists of the sum of two sigmoids (rising and falling) with a fixed time delay which is selected so that the peak does not fall below 96% of the function's range.

The function employs six parameters reflecting the shape of the accent (cf. figure 1). In earlier versions of PaIntE, the approximation of the F_0-contour for a given accent was carried out on the accented and post–accented syllables. However, for some accents (e.g. H$*$L) the beginning of the rise is likely to start at the preaccented syllable. In the current version of PaIntE the window used for the approximation of the F_0-contour for H$*$L accents has been extended to the preaccented syllable, so that the parameters are calculated over the span of the accented syllable and its immediate neighbours (unless it is followed by a boundary tone which causes the window to end at the end of the accented syllable).

Three of the PaIntE parameters were employed in the analyses: parameters *c1* and *c2* to determine the accents' ranges and parameter *b* to analyse the accents' temporal alignment (which has been shown to be crucial in the description of an accent's shape [12]).

4. Methodology

The speech data used in this study consist of a day of radio news broadcasts from a German radio station (Deutschlandfunk, 26/04/2007). It comprises 2 hrs 43 mins of read speech by 4 professional speakers (2 female, 2 male). The corpus was labelled prosodically according to GToBI(S) [14].

For each pitch accent, six parameters reflecting its shape were extracted using a parametric intonation model (PaIntE, [6], see section 3). All tokens, for each of the PaIntE-parameters, which were identified as outliers, were removed. Outliers were defined such that the upper 2.5 percentile as well as the lower 2.5 percentile of the data were excluded.

Given that the data processing includes several steps (manual labelling, F_0-approximation, F_0-smoothing, PaIntE parametrisation) where each step increases the number of potential error cases, only pitch accent tokens that were clearly reasonable examples of either L*H or H*L were extracted. These examples consist of tokens where either both sigmoids of the PaIntE function were used or where single accent-appropriate sigmoids were used (e.g. rising sigmoid for L*H accents in cases where the fall is not realised within the three syllable window). Since nuclear and prenuclear accents can differ in their shape [15], only nuclear accents were extracted.

For each word type, the frequency of the combination of this type with an accent type (e.g. "Merkel – H*L"), referred to as *accented word* in the following, was calculated. The analysed dataset comprises 1098 L*H tokens (578 accented word types) and 580 H*L tokens (385 types).

To normalise for speaker differences the PaIntE parameters were z-scored, on an accent-type basis, for each speaker separately. A z-score represents the number of standard deviations the value is away from the mean value for that dimension. The z-score is given by:

$$z - score_{dim} = \frac{value_{dim} - mean_{dim}}{sdev_{dim}} \quad (1)$$

Hence, each of the z-scored values indicates if the PaIntE parameter value of the accent in question is increased or decreased with respect to the typical parameter values of the individual speaker. For instance, a z-score value of 1 means that the parameter value is one standard deviation higher than the average value of the speaker in question. In this way values from different speakers are made comparable.

Factors, other than accented word frequency, that are likely to influence pitch accent shape were extracted by processing the speech data with the Edinburgh Speech Tools [16]. The factors were: number of segments in the accented syllable's onset and coda (0, 1, 2, or more than 2 segments) and the position of the accented syllable in the word, which distinguishes not only between different positions of the syllable in polysyllabic words, but also between mono and polysyllabic words.

For each of the pitch-accent sets (L*H and H*L) 4-way ANOVAs were performed for each of the z-scored PaIntE parameters, where the parameter in question was the dependent variable and the above mentioned factors and accented word frequency were incorporated as predictors. The results of the statistical analyses are given in section 5.

In those cases where accented word frequency was found to be a significant factor, 3-way ANOVAs were performed with accented word frequency removed from the set of predictors. The residuals of these ANOVAs, i.e. the data which cannot be explained by the three factors, were then plotted against accented word frequency in the figures below to establish how the frequency of occurrence of accented words affects the data.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
onsetsize	3	6.66	2.22	2.7315	0.043242	*
codasize	2	4.60	2.30	2.8274	0.060081	.
sylpostype	3	3.44	1.15	1.4084	0.239455	
accwordfreq	1	13.45	13.45	16.5477	<0.001	***
Residuals	518	421.14	0.81			

Table 1: ANOVA results (main effects) for H$*$L dependent variable *c2*.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
onsetsize	3	3.69	1.23	1.2050	0.306734	
codasize	2	0.18	0.09	0.0868	0.916855	
sylpostype	3	0.80	0.27	0.2630	0.852098	
accwordfreq	1	4.09	4.09	4.0163	0.045324	*
Residuals	1024	1043.97	1.02			

Table 2: ANOVA results (main effects) for L∗H dependent variable *c1*.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
onsetsize	3	22.12	7.37	11.0537	<0.001	***
codasize	2	55.14	27.57	41.3276	<0.001	***
sylpostype	3	321.36	107.12	160.5742	<0.001	***
accwordfreq	1	2.71	2.71	4.0572	0.044246	*
Residuals	1024	683.12	0.67			

Table 3: ANOVA results (main effects) for L∗H dependent variable *b*.

5. Results

The results of ANOVAs which yielded significant effects for accented word frequency are presented in the tables above. It is important to note that only main effects are reported, for reasons of brevity (some interactions were also found involving frequency). For H∗L accented words table 1 indicates a clear effect of accented word frequency (accwordfreq) on pitch-accent realisation of parameter *c2* ($p< 0.001$). This parameter corresponds to the range of the fall in H∗L accents. Figure 2 plots the residuals of a 3-way ANOVA of the H∗L *c2* data, with accented word frequency removed as predictor, against accented word frequency (a regression line has been fitted). The figure illustrates greater z-score values for the *c2* parameter as accented word frequency increases. This corresponds to a greater range of the fall.

For L∗H accented words table 2 demonstrates a significant main effect of accented word frequency on parameter *c1* (range of the rise, $p<0.05$). Figure 3 (again plotting residuals against accented word frequency) indicates a subtle increase in the rise of the accent with increasing frequency. Table 3 presents a significant effect of accented word frequency on the *b* parameter ($p<0.05$). This parameter corresponds to the temporal alignment of the peak relative to the accented syllable. Figure 4 plots the residuals of a 3-way ANOVA of the L∗H *b* data, with accented word frequency removed as predictor, across accented word frequency. In this case, as accented word frequency increases, the peak is realised later (but see discussion below).

6. Discussion and conclusion

At the outset, this paper sought to establish the impact word frequency might have on the realisation of two frequently occurring pitch accent types, and what consequences might ensue from such an impact for theories of lexical storage and post-lexical accenting. Teasing apart the effect of accented word frequency from other, more obvious, influential factors is tricky, however the ANOVA results above indicate an effect of frequency of occurrence on the realisation of pitch accent ranges, and the temporal positioning of L∗H peaks. It is important to note however that while accented word frequency affects L∗H peak location (visible when plotted against the residuals), this effect is not present in the non-residual data. In other words while frequency of occurrence aligns the peak later, the other factors reign it in.

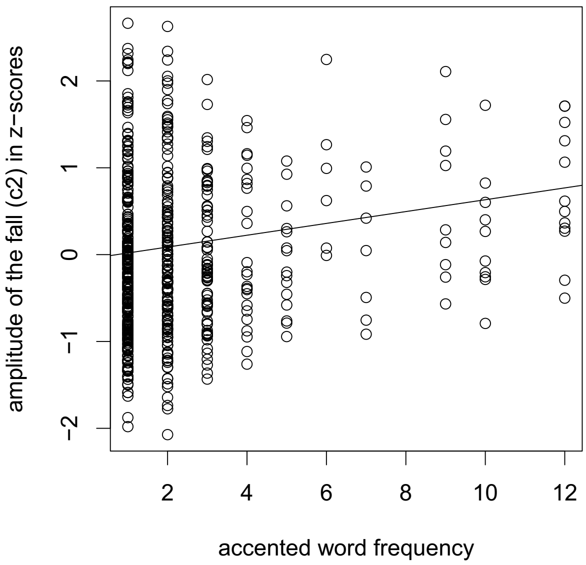

Figure 2: Residual-values for z-scored parameter *c2* for nuclear H*L accents plotted against the frequency with which the respective word occurs with a nuclear H*L accent.

With regard to pitch accent range, the rise tends to be larger with increasing accented word frequency in nuclear L*H accents, and the fall is greater in nuclear H*L accents. Thus, for both accent types, the realisation of an accent becomes more prominent as the frequency of the accented word increases.

From the perspective of theories of lexical and prosodic production these findings indicate a relationship between accented word frequency and pitch accent realisation. Within an exemplar-theoretic framework, the bias towards more pronounced accents with increasing frequency can be explained in terms of a production-perception loop: during speech production a speaker selects a stored exemplar as a production target. Assuming that pitch accents can be stored with the word, the speaker would select an exemplar that matches not only the intended word but also the intended pitch accent. The main purpose of pitch-accentuation is to make a word more prominent, the result of which is to draw the listener's attention to the word and make it more noteworthy. Moreover, the most prominent tokens are likely to have the greatest activation in the speakers memory as they are more marked. Consequently, it seems plausible that the exemplar that is most prominent is chosen as a production target. During production, imprecision inherent in the production process will yield a pitch-accented word which is either more, or less, prominent than the production target itself. This new production will be perceived and stored as a new exemplar, and, if more prominent, will be more likely to be selected as a production target than the previous target in subsequent productions (and possibly be produced more prominently again), and will be less likely to be selected for production if less prominent. Ultimately, this will yield more prominent productions and increase prominence within the cloud of accented

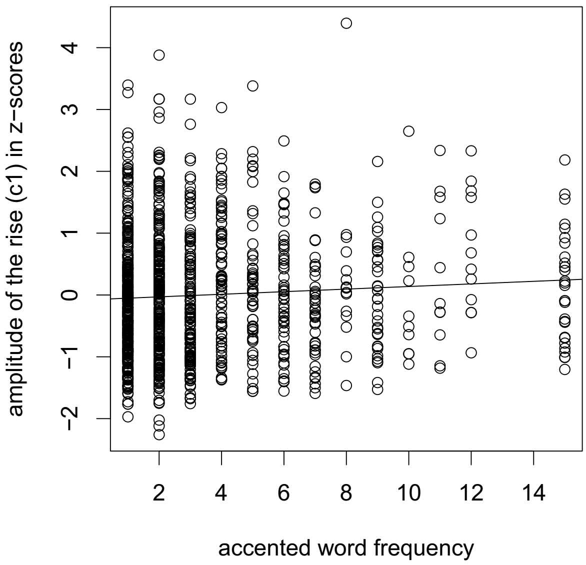

Figure 3: Residual-values for z-scored parameter *c1* for nuclear L*H accents plotted against the frequency with which the respective word occurs with a nuclear L*H accent.

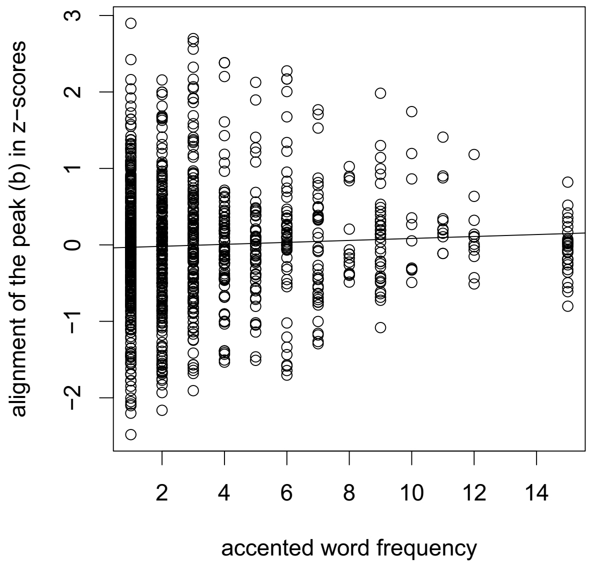

Figure 4: Residual-values for z-scored parameter *b* for nuclear L*H accents plotted against the frequency with which the respective word occurs with a nuclear L*H accent.

words as a whole, as the less prominent exemplars will fade from memory due to lack of activation. This behaviour will become entrenched over time, preventing excessive peak prominence [1].

Of course there are instances where high frequency accented words are produced less prominently than mean prominence. This is, in part, because idiosyncratic productions always occur and are accumulated over time. Discourse, tonal and segmental information (e.g. information status, distance from prosodic boundaries, vowel duration), is also likely to play a role, and requires further study. Nevertheless, there is a significant inclination towards prominence as the frequency of the combination of word type with pitch accent type increases.

7. References

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

[2] J. Bybee, "From usage to grammar: The mind's response to repetition," *Language*, vol. 84, pp. 529–551, 2006.

[3] A. Schweitzer and B. Möbius, "Exemplar-based production of prosody: Evidence from segment and syllable durations," in *Speech Prosody 2004 (Nara, Japan)*, 2004, pp. 459–462.

[4] M. Walsh, H. Schütze, B. Möbius, and A. Schweitzer, "An exemplar-theoretic account of syllable frequency effects," in *Proceedings of the International Congress of Phonetic Sciences (Saarbrücken)*, 2007, pp. 481–484.

[5] K. Schweitzer, M. Walsh, B. Möbius, A. Riester, A. Schweitzer, and H. Schütze, "Frequency Matters: Pitch accents and Information Status," in *Proceedings of EACL-09*, Athens, Greece, 2009.

[6] G. Möhler and A. Conkie, "Parametric modeling of intonation using vector quantization," in *Proceedings of the Third International Workshop on Speech Synthesis (Jenolan Caves, Australia)*, 1998, pp. 311–316.

[7] K. Abbot-Smith and M. Tomasello, "Exemplar-learning and schematization in a usage-based account of syntactic acquisition," *The Linguistic Review*, vol. 23, no. 3, pp. 275–290, 2006.

[8] K. Johnson, "Speech perception without speaker normalization: An exemplar model," in *Talker Variability in Speech Processing*, K. Johnson and J. W. Mullennix, Eds. San Diego: Academic Press, 1997, pp. 145–165.

[9] K. Croot and K. Rastle, "Is there a syllabary containing stored articulatory plans for speech production in English?" in *Proceedings of the 10th Australian International Conference on Speech Science and Technology (Sydney)*, 2004, pp. 376–381.

[10] S. P. Whiteside and R. A. Varley, "Dual-route phonetic encoding: Some acoustic evidence," in *Proceedings of the 5th International Conference on Spoken Language Processing (Sydney)*, vol. 7, 1998, pp. 3155–3158.

[11] E. Marsi, M. Reynaert, A. van den Bosch, W. Daelemans, and V. Hoste, "Learning to predict pitch accents and prosodic boundaries in dutch," in *Proceedings of the ACL-2003 Conference (Sapporo, Japan)*, 2003, pp. 489–496.

[12] J. P. H. van Santen and B. Möbius, "A quantitative model of F0 generation and alignment," in *Intonation—Analysis, Modelling and Technology*, A. Botinis, Ed. Dordrecht: Kluwer, 2000, pp. 269–288.

[13] G. Möhler, "Improvements of the PaIntE model for F_0 parametrization," Institute of Natural Language Processing, University of Stuttgart, Tech. Rep., 2001, draft version.

[14] J. Mayer, "Transcribing German Intonation – The Stuttgart System," Universität Stuttgart, Tech. Rep., 1995, http://www.ims.uni-stuttgart.de/phonetik/joerg/labman/STGTsystem.html.

[15] D. Mücke, M. Grice, J. Becker, A. Hermes, and S. Baumann, "Articulatory and acoustic correlates of prenuclear and nuclear accents," in *Proceedings of Speech Prosody 2006 (Dresden)*, 2006, pp. 297–300.

[16] P. Taylor, R. Caley, A. W. Black, and S. King, "Edinburgh speech tools library," [http://www.cstr.ed.ac.uk/projects/speech_tools/], 1999, system Documentation Edition 1.2, for 1.2.0 15th June 1999.