PROSODIC VARIABILITY IN LEXICAL SEQUENCES: INTONATION ENTRENCHES TOO

Katrin Schweitzer[a], *Michael Walsh*[a], *Sasha Calhoun*[b] & *Hinrich Schütze*[a]

[a]University of Stuttgart, Germany; [b]Victoria University of Wellington, New Zealand
Katrin.Schweitzer@ims.uni-stuttgart.de; sasha.calhoun@vuw.ac.nz

ABSTRACT

This paper investigates the relationship between the probability with which a word appears in a given lexical context and the prosodic variability of that word. Regression analyses indicate that the higher the probability of a word in its lexical context, the lower the variability in the pitch accent realisation of that word (if accented), and the lower the variability in the prosodic pattern around the word. These results imply that prosodic realisation can be subject to lexicalised entrenchment, extending previous findings within the Exemplar Theory framework on the phonetic properties of collocations, but contrary to the usual assumption that prosody is "post-lexical" in English.

Keywords: intonation, prosody, frequency, Exemplar Theory, collocations

1. INTRODUCTION

Much of language is made up of predictable lexical sequences, or collocations [4], e.g. *I don't know* or *make a mistake*: phrases where the component words occur together much more frequently than would be expected by chance, and where word substitution sounds odd to native speakers, e.g. *do a mistake* (e.g. [2]). Within the framework of Exemplar Theory, these frequent phrases are stored as 'exemplars' in memory [2]. Key features of Exemplar Theory are 1) the assumption that language percepts are stored in rich detail in memory as exemplars, 2) exemplars function as targets for subsequent productions and 3) these memory traces are sensitive to frequency and recency of usage. In an exemplar-theoretic account, the repeated occurrence of a phrase like *I don't know* leads, over time, to a fluent (i.e. entrenched) neuromotor routine which encodes the phonetic and lexical representations of the phrase together [2]. As the production of frequent exemplars becomes entrenched over time, they become less variable [2, 9]. To our knowledge, however, the *prosodic* properties of such frequent lexical sequences have not been investigated (but see [12] for effects of familiarity on prosodic phrasing).

It is usually assumed that prosody is assigned "post-lexically" in English. However, recent research has shown lexical effects on prosodic realisation related to relative frequency. For instance, [11] showed that the frequency of pitch accent type- word collocations in German affects the shape of the pitch accent, providing evidence for the lexicalised storage of different pitch accent types. [10] demonstrated that the greater the relative frequency of a word-pitch-accent type pair in English, the less variable the realisations of that accent. This again shows lexicalised storage of pitch accent types, as words that occur relatively often with a particular pitch accent type have less variable pitch accent realisation.

If the acoustic information stored with exemplars includes prosodic information (e.g. the pitch contour and/or the pattern of accents and phrasing), then lexical sequences that are highly probable are expected to exhibit less prosodic variation than less probable sequences, just as they exhibit less phonetic variation. Given this expectation, the following two hypotheses were tested on the Switchboard Corpus [5]: with increasing probability of a word in a lexical context

1. the variability of the pitch accent contour on the word decreases,
2. the variability of the prosodic patterns in which the word occurs decreases.

These hypotheses were investigated by looking at the prosodic variability of a word in relation to the probability of the word in its lexical context (one word on either side). Prosodic variability was measured by the variability of 1) the pitch accent contour *on* the word, and 2) the prosodic pattern *around* the word.

Below, section 2 introduces the dataset, section 3 describes the method for measuring the probability of words in their three-word context. Sections 4 and 5 set out the first and second experiments addressing the hypotheses above, section 6 offers some discussion and the outlook for future work.

2. DATA

The corpus used was Switchboard, a collection of spontaneous telephone conversations between

American English speakers [5]. 76 conversations, or around 6h of speech from 114 speakers, are annotated for pitch accent and prosodic boundary location using the Tones and Break Indices (ToBI) standards [1], see [3]. Tonal accent type is not marked.

The lexical frequencies (for words and trigrams) were extracted from the whole Switchboard corpus, along with the Callhome American English corpus [6], a smaller corpus of spontaneous telephone conversations. The combined corpus comprised just over 3M words. Trigrams containing fillers like "uh-hum", and incomplete words, were not included.

3. LEXICAL PROBABILITY

To calculate the probability of a word occurring in a certain lexical context, i.e. the probability P_{Lex} of the word w_i given its left neighbour l_i and its right neigbhour r_i, the trigram frequency (in the combined Switchboard and Callhome corpus) was divided by the number of other trigrams in which the word occurs in the same position.

$$(1) \quad P_{Lex}(l_i w_i r_i) = \frac{C(l_i w_i r_i)}{C(l w_i r)}$$

where *C* means a count. This probability measure accounts for the variability of lexical contexts in a token-based way. The greater the value of P_{Lex}, the smaller the remaining probability available for other contexts, i.e. less opportunity for diverse lexical contexts. For example, the word lot occurs 10059 time. Of these, it occurs 6631 times in the trigram a lot of yielding a probability value of

$$P_{Lex}(a\ lot\ of) = \frac{6631}{10059} \approx 0.66.$$

Table 1 lists the 10 most probable trigrams. Most would be considered collocations in English.

Table 1: List of the 10 most probable trigrams from the dataset and their probability in combined Switchboard and Callhome.

P_{lex}	trigram	P_{lex}	trigram
0.74	the rest of	0.49	a couple of
0.66	a lot of	0.48	to worry about
0.61	i grew up	0.43	a matter of
0.54	as far as	0.40	a nursing home
0.54	be able to	0.40	as soon as

4. EXPERIMENT 1: VARIABILITY OF PITCH ACCENT SHAPE

To create the pitch accent variability dataset for Exp. 1, trigrams that occured at least 4 times with a pitch accent on the middle word were extracted (95 trigram types). As the types vary greatly in their token frequency (from 4 to 55 tokens), 100 datasets were created where for each trigram type 4 tokens were randomly selected.

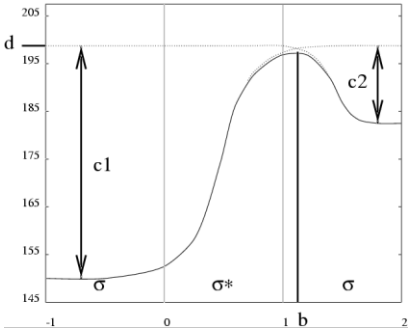
4.1. Determination of pitch accent shape

A parametric intonation model, PaIntE [8], was used to represent pitch accent shape using six linguistically meaningful parameters. The model approximates stretches of F_0 using a phonetically motivated model function.

This operates on a three-syllable-window centered on the accented syllable, if these syllables are in the same intonation phrase. Six parameters are used to describe the contour (see Figure 1): parameter *b* locates the peak of the accent within the three-syllable window, *c1* and *c2* model the ranges of the rising and falling slope of the accent's con- tour, d is the actual height of the peak and *a1* and *a2* (not displayed in the figure) denote the "amplitude-normalised" steepness of the rising and falling slope.

To normalise for speaker differences the PaIntE parameters were z-scored for each speaker.

Figure 1: The PaIntE model function over a 3-syllable-window. The accented syllable is starred.



4.2. Calculation of pitch accent variability

For the variability measure, each pitch accent was represented as a 6-dimensional vector (one dimension for each z-scored PaIntE parameter). For each trigram that occurred at least 4 times with an accent on the middle word, the pairwise Euclidean distance between the respective pitch accent tokens in the6-dimensional space was calculated: first, the Euclidean distance *d* between each pair of PaIntE vector tokens *x* and *y* of the same trigram type was calculated according to formula (2). Since there are 6 PaIntE parameters, $n = 6$.

$$(2) \quad d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

The distances between all pairs were then summed and an average calculated. This average distance of all pairwise comparisons gives a measure

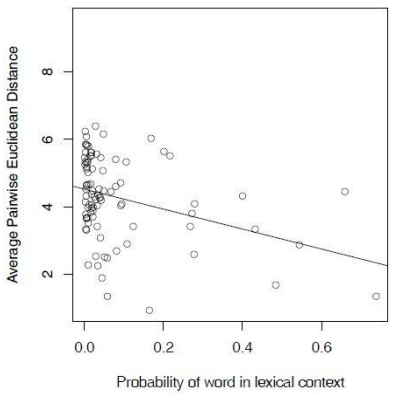
of the variability of the trigram type: tokens with more similar accent contours have a smaller average distance.

4.3. Results

For each of the randomised datasets of trigram tokens occurring at least 4 times with a pitch accented middle word, a regression model was fitted to predict pitch accent variability by the probability of the word in its lexical context. The model yielded a significant *p*-value ($\alpha = 0.05$) in 80% of the cases, and in 7 cases there was a tendency ($p < 0.08$). Figure 2 displays the result for a sample dataset: with increasing probability of the word in its lexical context, the average distance (i.e. the variability) of pitch accent tokens on this word decreases. That is, the more probable a word in its context, the more similar the realisations of pitch accent tokens on the trigram. This relationship (a negative correlation) was the same in all significant regression models.

As can be seen in the figure, the correlation is not a strong one ($Adj.R^2 \approx 0.11$), however the models yield significance in a vast majority of cases.

Figure 2: The probability of a word occurring in a certain lexical context plotted against pitch accent variability among pitch accented tokens of trigram-types. With increasing lexical probability, the variability in pitch accent shape decreases.



5. EXPERIMENT 2: VARIABILITY OF PROSODIC PATTERN

For the prosodic pattern dataset for Exp. 2, all middle words of trigrams that occured at least 4 times in the prosodically annotated part of Switchboard, were extracted (3705 tokens, 124 word types in 541 trigram types).

5.1. Calculation of prosodic pattern variability

The prosodic pattern of each extracted word was taken to be the sequence of any accent or phrase boundary (ToBI break index 3 or 4) on its left neighbour, the word itself and the right neighbour. For example, for the word *lot* in the trigram a lot of, if there was an accent on lot and a boundary after of, the prosodic pattern was encoded as *NoAcc-NoBound/Acc-NoBound/NoAcc-Bound*. To calculate the probability of a word occurring with a certain prosodic pattern, for each word w_i, the number of times the word occurred with the particular pattern, p_i, was divided by the number of times the word occurred with any prosodic pattern, *p*.

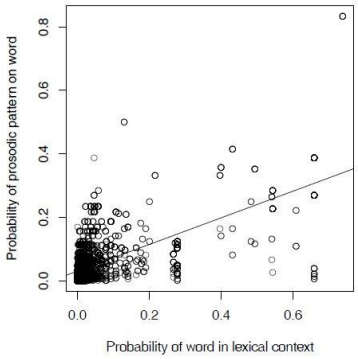
$$(3) \quad P_{pros}(w_i, p_i) = \frac{C(w_i, p_i)}{C(w_i, p)}$$

For instance, of the 100 tokens of lot in the prosodically annotated part of Switchboard, 48 were with the prosodic pattern P1 (*NoAcc-NoBound/Acc-NoBound/NoAcc-NoBound*) yielding a probability value of $P_{P1} = \frac{48}{100} = 0.48$. This measures the variability of the word's prosodic pattern in a token-based fashion: the greater the value of P_{pros}, the smaller the remaining probability for other prosodic patterns (apart from P1).

5.2. Results

A regression model was fitted to predict the probability of a word occurring with a certain prosodic pattern by the probability of the word in its lexical context, yielding a significant p-value of $p < 0.001$ ($Adj.R^2 = 0.49$). Figure 3 displays the result: with increasing probability of a word in its lexical context, the probability of it occurring with a certain prosodic pattern increases as well. Hence, the variability of prosodic patterns decreases with increasing probability of the word in its lexical context. Note that with the obvious outlier removed, the model is still significant ($Adj.R^2 \approx 0.48$).

Figure 3: The probability of a word occurring in a certain lexical context plotted against its probability of occurring in a certain prosodic pattern. With increasing lexical probability, the probability of a particular prosodic pattern around the word also increases.



6. DISCUSSION

Both experiments confirmed the hypotheses made based on exemplar-theoretic expectations. Exp. 1 confirmed that pitch accent contours on the accented middle word of trigrams are less variable with increasing probability of the word in the trigram context. Decreasing variability might come from a decrease in the number of accent types on the word or more similar realisations of the same pitch accent type. Since the dataset is not annotated for accent type, these different cases could not be distinguished. However, the conclusions are the same in both cases: higher lexical probability leads to lesser tonal variability. Exp. 2 demonstrated that words that are highly probable in their lexical context are also more likely to occur with a certain prosodic pattern. Together, these findings indicated that lexically probable sequences of words exhibit less prosodic variation than less probable sequences.

This relationship between the lexical level and prosody would not be expected within autosegmental-metrical theories of intonation (see [7]). In such theories, prosodic realisation is determined by a combination of 'top-down' syntactic, semantic and pragmatic factors (e.g. given/new status), and the phonological context (e.g. how close together accents are). While these factors are undoubtedly still relevant, these results show that in collocations, at least, stored information about the prosodic patterns and accent realisation of that collocation also plays a part.

Within an exemplar-theoretic framework, frequent collocations are expected to be stored as exemplars along with phonetic and contextual information. As is shown here, this includes the prosodic realisation in the form of pitch contours. In production, the exemplars containing a word serve as production targets. If most of these come from the same lexical context, then there should be less variability amongst them than if they come from a variety of contexts. Hence, productions of that word should show less variability. For example, 66% of lot tokens come from the collocation *a lot of*. As such collocations are realised with little variance (because of entrenchment), productions of *lot* overall should on average show less prosodic variablity. This was confirmed in these experiments where both the accent realisation of, and the prosodic pattern around, words like *lot* in *a lot of* were found to vary less.

The correlation between accent contour variability and lexical probability was much lower than for prosodic pattern variability. This could be because the entrenchment effect is weaker for accent realisation, or because the dataset in Exp. 1 was too small to illustrate the strength of the effect.

In conclusion, this study extends existing Exemplar Theory research, showing that intonation can be entrenched in collocations. However, there is still a great deal to be understood about how lexicalised storage interacts with 'top-down' information in the production of prosody.

7. ACKNOWLEDGEMENTS

This work was supported by the German Research Foundation DFG (SFB-732) and a British Academy Postdoctoral Fellowship.

8. REFERENCES

[1] Beckman, M., Hirschberg, J. 1999. The ToBI annotation conventions. *http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html*

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

[3] Calhoun, S., Carletta, J., Brenier, J., Mayo, N., Jurafsky, D., Steedman, M., Beaver, D. 2010. The NXT-format switchboard corpus: A rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation* 44(4), 387-419.

[4] Erman, B., Warren, B. 2000. The idiom principle and the open choice principle. *Text* 20(1), 29-62.

[5] Godfrey, J., Holliman, E., McDaniel, J. 1992. Switchboard: telephone speech corpus for research and development acoustics. *Proc. of ICASSP* 1, 517-520.

[6] Kingsbury, P., Strassel, S., McLemore, C., McIntyre, R. 1997. Callhome American English transcripts. *Linguistics Data Consortium*, Philadelphia, No. LDC97T14.

[7] Ladd, D.R. 2008. *Intonational Phonology*. Cambridge University Press, Cambridge, UK.

[8] Möhler, G., Conkie, A. 1998. Parametric modeling of intonation using vector quantization. *Proc. of International Workshop on Speech Synthesis* Jenolan Caves, Australia, 311-316.

[9] Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J., Hopper, P. (eds.), *Frequency and the Emergence of Linguistic Structure*, Benjamins, Amsterdam, 137-157.

[10] Schweitzer, K., Calhoun, S., Schütze, H., Schweitzer, A., Walsh, M. 2010a. Relative frequency affects pitch accent realisation: Evidence for exemplar storage of prosody. *Proc. of SST* Melbourne, Australia, 62-65.

[11] Schweitzer, K., Walsh, M., Möbius, B., Schütze, H. 2010b. Frequency of occurrence effects on pitch accent realisation. *Proc. of Interspeech 2010* Makuhari, Japan, 138-141.

[12] Vaissière, J., Michaud, A. 2006 Prosodic constituents in French: a data-driven approach. In Fònagy, I., Kawaguchi Y., Moriguchi, T. (eds.), *Prosody and Syntax*. Benjamins, 47-64.