WHAT IS COMPOUND STRESS?

Gero Kunter, Ingo Plag

Universität Siegen
kunter@anglistik.uni-siegen.de, plag@anglistik.uni-siegen.de

ABSTRACT

This paper investigates the implementation of stress in English noun-noun compounds. First, a perception experiment examines how listeners perceive prominence in compounds. After that, significant acoustic correlates of prominence are established. Finally, a cluster analysis is described that classifies compounds on the basis of their phonetic features and which is capable of separating different stress categories. The results demonstrate how gradient acoustic measurements and discrete phonological contrasts can be mapped onto each other.

Keywords: compounding, prominence, perception, acoustic features, classification.

1. INTRODUCTION[i]

Stress is the phonological category pertaining to prominence relations within linguistic units. But while the phonetic implementation has been analysed in single words [3] or on the sentence level [10], only little research has gone into compound stress. In English, compounding is a highly productive process to create new words such as *boarding schools*, *eviction notice* or *state colleges*. Noun-noun compounds like these are generally classified as receiving stress either on the first constituent or on the second [7]. There is, however, no empirical evidence that this is indeed the way compounds are perceived by listeners. In addition, it is largely unresolved how the different stress types are encoded acoustically. This paper will address these issues in turn, and will demonstrate how the phonetic, gradual implementation and the phonological, categorical theory of stress interface with each other.

2. PERCEPTION EXPERIMENT

The speech material that was used in the analysis was taken from the Boston University Radio Speech Corpus [8]. This corpus contains studio-quality recordings of 3 female and 4 male radio news speakers of American English. The transcripts were manually scanned for unambiguous noun-noun constructions, and from the resulting list of compounds, 15 different items were randomly chosen from each speaker, resulting in 105 unique compounds.

32 native speakers of American English participated in a perception experiment that used the data from the Boston corpus as stimuli. To reduce the influence of sentence intonation, the data were presented in contexts consisting of the preceding and following seven words. The participants listened to each stimulus in a randomized order, and indicated the prominence relation between the two members of the compound by moving a slider on a computer screen. The slider could be moved freely on the screen without a noticeable stepping or marks on the scale. Participants were instructed to use the slider to indicate which of the two constituents they perceived as more prominent, and to use the deviation of the slider from the center to indicate how clearly more prominent they perceived the respective constituent. A central position in which both members were perceived to have equal prominence was also allowed. The slider position was internally transformed to a scale ranging from 0 (leftmost position) to 999 (rightmost position).

For each item, the 32 prominence ratings were averaged to calculate the mean prominence rating for the respective compound. The standard deviations ranged from 116.7 to 266.9, with the average standard deviation being 176.3 across all items.

Figure 1: Density function of the distribution of prominence ratings.

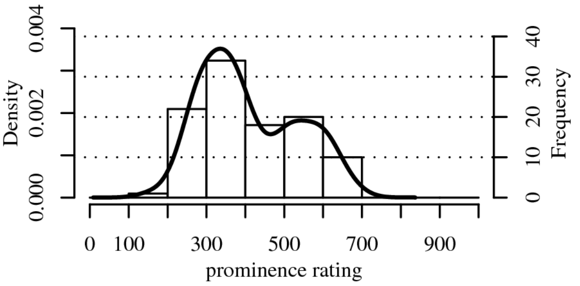


Fig. 1 shows the distribution of the mean prominence ratings. There are two important observations to be made. First, the prominence ratings in the perception experiment follow a bimodal distribution, with one peak at 335.1 and another at 547.9. These two regions were chosen most frequently by the participants to describe the prominence relation between first and second member. This suggests that there are, indeed, two possible prominence patterns in English compounds. In the first one, the left member is perceived as more prominent, and in the other one, this is the case for the right member. Given the respective peak positions, these two patterns correspond quite well to the phonological categories of leftward and rightward stress.

The second observation is that the range of the scale is not exploited symmetrically. The distance between the left peak and the scale center at 499 is 163.9 units, which is more than three times the difference between the scale center and the right peak (48.9 units). In addition to that, there is a medium positive correlation between the averaged prominence ratings for each item and the respective standard deviation (r=0.32, p<0.01), which indicates that the variability within the ratings increases with an increase in perceived right stress. The variability is lowest for items with a rating corresponding to leftward stress. In conclusion, listeners are more confident in their judgements for a compound with left prominence, but their ratings are less decisive if that is not the case. As a category, rightward stress seems to offer less clear perceptual cues than leftward stress. This raises the question which acoustic features contribute to the perceived prominence patterns in compounds. This is addressed in the next section.

3. ACOUSTIC MODEL

A linear regression model is one way of assessing which parameters are relevant for the perception of stress categories. In the present analysis, the prominence scale from the perception experiment is used as the response variable for the model. As above, a low estimated prominence rating corresponds to leftward prominence and a high rating to rightward prominence.

In English, pitch, intensity and duration are the standard parameters regarded as acoustic correlates of stress [6]. One of the few previous studies on the acoustics of compound stress showed that the implementation is similar: the difference between the stressed and the unstressed constituent in a compound is assumed to be signaled by a higher pitch, a higher intensity, and a longer duration of the more prominent member [2]. In the present analysis, each of these relative measures is expressed by measuring the values for the left and right constituent. Then, the difference is calculated. For example, the duration difference Δ_{dur} is obtained by $d_{left} - d_{right}$. It is expected to be clearly positive if the left constituent is more prominent than the right one. An increase of Δ_{dur} (longer left member) is expected to lead to an decrease in the estimated prominence rating (more prominent left member), and similarly for the intensity (Δ_{int}) and mean pitch differences (Δ_{pitch}, expressed as mean F_0 differences in semitones).

The spectral tilt T, the decrease in intensity of higher frequencies in the spectrum, has generally been shown to be larger for unstressed syllables than for stressed syllables within words [9], and to be larger for less prominent words than for more prominent words in phrases [10]. It is to be expected, then, that the prominence perception in compounds is correlated with the spectral tilt of the two constituents as well. A stressed constituent should have a less steep slope in the spectral balance than an unstressed one. We measured T as the difference between the mean intensities below and above 1000 Hz on the LTAS of each member.

Another measure that has been included as a predictor of prominence in this study is the mean absolute pitch slope S. This measure calculates the average of all absolute pitch changes between two adjacent time slices within a constituent, and thus represents the mean local variation in pitch. If a constituent has a rather steady pitch, S will be low, while if there is much pitch movement within the constituent, S will be higher.

Tab. 1 summarizes the acoustic measurements that were considered for inclusion in the linear regression model.[ii]

Table 1: Acoustic measurements considered for a regression analysis predicting prominence ratings.

(1)	mean pitch difference (in semitones)	Δ_{pitch}
(2)	mean intensity difference (in dB)	Δ_{int}
(3)	duration difference (in sec)	Δ_{dur}
(4)	spectral tilt (in dB)	T_{left}, T_{right}
(5)	mean absolute pitch slope (in semitones/sec)	S_{left}, S_{right}

Using the phonetic software *Praat* [1], these measures were taken for the 105 compounds from the perception experiment. The pertinent analysis intervals were determined by selecting the syllable with primary stress within the left and the right constituent for each compound. Acoustic measurements were then taken from the sonorant part of the rime of these two syllables.

To avoid overfitting the data, several linear regression models were constructed using all possible combinations of the well-established predictors from (1)–(3), and two of the four potential predictors from (4) and (5). Then, bootstrap validations were used to determine which of the predictors contributed significantly to the fit in each model. In result, only Δ_{pitch}, Δ_{int}, Δ_{dur}, T_{left}, and S_{right} were found to be significant predictors in all models, while T_{right} and S_{left} were not significant in any model. The data was then examined for points with undue leverage, which lead to the exclusion of one item from the model, as this item showed a highly unusual intonation pattern. There was no general effect of sentence intonation, operationalized as sentence positions (initial, final, other), on perception ratings ($F(2, 102) < 1$), though.

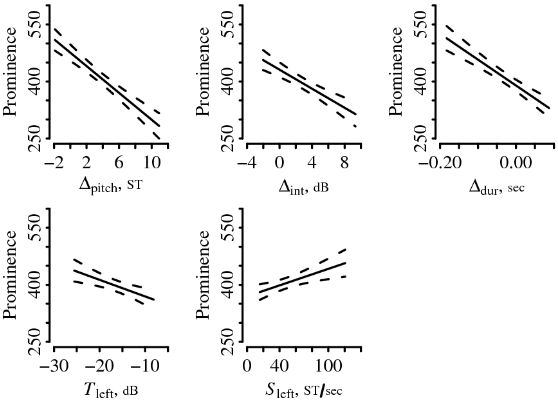
The coefficients *B* and the standardized coefficients β of the final regression model can be found in Tab. 2, and the partial effects are plotted in Fig. 2. The linear regression model accounts for 70.0% of the variance in the prominence ratings.

Table 2: Summary of regression analysis predicting perception rating ($N = 104$).

	B	Std. Err.	β
intercept	401.10	23.72	0.01 ***
Δ_{pitch}	-17.62	2.12	-0.63 ***
Δ_{int}	-12.54	2.23	-0.39 ***
Δ_{dur}	-679.39	92.52	-0.42 ***
T_{left}	-4.38	1.35	-0.19 **
S_{right}	0.73	0.23	0.21 **

$R^2 = 0.70$. ** $p < 0.01$, *** $p < 0.001$

Figure 2: Partial effects of a linear regression model predicting prominence ratings.



The steepness of the slopes for the partial effects as well as the absolute values of the standardized β coefficients reflect the strength of the respective predictor for the regression model. The most powerful predictor for prominence ratings is Δ_{pitch}, which reflects the notion that in English, stress, the phonological correspondence of prominence, is realized to a large extent by pitch accents [4]. The other significant correlates, in order of decreasing magnitude of influence on the predicted prominence ratings, are Δ_{dur}, Δ_{int}, S_{right}, and T_{left}.

The slopes for Δ_{pitch}, Δ_{int}, Δ_{dur}, and T_{left} are all negative; an increase in the respective measure leads to a decrease in the predicted prominence rating, which corresponds to a more prominent left member. This is in accordance with previous observations about prominence in compounds for Δ_{pitch}, Δ_{int}, and Δ_{dur}: if the left member is more prominent than the right member, it will have a higher pitch, a higher intensity and a longer duration than the right member.

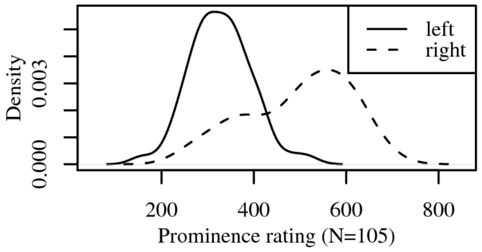
The spectral tilt measure turns out to be also a significant predictor for prominence ratings, and the direction of the influence again agrees with other findings on spectral tilt. A prominent left member has a flat spectral balance; with a more negative T_{left}, the prominence relation in the compound changes in the direction of the right member. It is noteworthy that there is no statistically significant effect of T_{right} on the prominence rating. Apparently, listeners do not use this acoustic feature for the detection of stress patterns. On the other hand, it is only for the right member that the mean absolute pitch slope *S* is a significant predictor. Here, a large pitch movement contributes to a more rightward prominence perception, while a steady pitch in the right member seems to signal leftward prominence. The statistical model does not show a similar influence of pitch slope in the left member.

4. AUTOMATIC CLASSIFICATION

The regression model in Tab. 2 can be used to estimate the prominence relation even for compounds that have not been rated by listeners; the estimate will be on the prominence scale ranging from 0 to 999. However, there are situations in which a categorical decision is desirable; these situations range from automated speech recognition applications to studies in theoretical phonology that analyze naturally spoken compounds. In this section, it is shown that such a classification task is possible using a hierarchical cluster analysis. This algorithm groups together the two items which have similar acoustic parameters. This group is, in turn, joined with the item or group that is again most similar, until all items are accounted for. The two uppermost branches of the resulting tree structure represent two groups of items that are as dissimilar as possible.

For the present analysis, the measures that were found to be significant phonetic correlates of prominence were used as clustering parameters for the data of 105 compounds as used above. Each measure was standardized to eliminate scale effects, and then weighted using the β coefficients from the linear regression model. The two head branches were taken to reflect two classes of compounds that are phonetically most distant. A *t*-test of the prominence ratings for the first group ($M_1 = 327.6$) and for the second group ($M_2 = 496.6$) shows that the stress patterns of the two clusters are, indeed, perceived in significantly different ways (t (103)=9.327, p<0.01). The effect size is very large (Cohen's d=1.834), corresponding to an overlap of less than 23 percent between the two groups. This is illustrated in Fig. 3, which shows the density functions for the first (solid line) and the second cluster (dashed line). The locations of the peaks closely resemble the plot in Fig. 1. Apparently, the cluster analysis is highly successful in separating the data into two categories that yield very different prominence responses.

Figure 3: Density functions of prominence ratings for left (solid) and right (dashed) cluster.



Like the listeners in the experiment, the automatic classification reaches a very good accuracy in detecting leftward stress: this cluster is well-defined and follows a normal distribution. The rightward stress category contains some misclassifications, though, as is visible in the negative skew of the density function. This is reminiscent of the observation that the variability in listener ratings increased with higher prominence ratings.

5. CONCLUSION

There are only few acoustic and perceptual studies of prominence and stress of English noun-noun compounds. In this paper, this gap is filled by examining three issues. First, it has been shown that listeners discriminate between two different prominence types in compounds. Left prominence is recognized almost unequivocally, while the perception of right prominence shows more variation. Second, differences in pitch, intensity, and duration, and to a lesser extent, the spectral tilt of the left member and the local pitch variability of the right member are phonetic correlates of compound stress and can be employed in a linear regression model to predict perception ratings with a high accuracy. Third, the stress category of a given compound can be determined on the basis of the acoustic signal by using a hierarchical cluster analysis. The resulting classification shows a similar performance as native listeners.

While phonetic correlates other than pitch are sometimes acknowledged as properties of phonological stress in English [4], the phenomenon is nonetheless often reduced to pitch accents alone. As shown above, the phonetic implementation is more complex, as compound stress is expressed by a number of acoustic correlates. There are also much more overlap and gradual differences between the stress categories than a description based on discrete phonological contrasts alone would predict. In this, the present paper contributes to the debate on the interface between phonology and phonetics (e.g. Kingston [5]) in that the results demonstrate how the understanding of phonological categories can profit from a thorough analysis of the phonetic material.

6. REFERENCES

[1] Boersma, P., Weenink, D. 2007. *Praat: doing phonetics by computer*. http://www.praat.org visited 30-Jan-07.

[2] Farnetani, E., Torsello, C. T., Cosi, P. 1988. "English compound versus non-compound noun phrases in discourse: an acoustic an perceptual study", *Language and Speech* 31(2), 157-180.

[3] Fry, D. B. 1958. "Experiments in the perception of stress", *Language and Speech* 1, 126-152.

[4] Gussenhoven, C. 2004. *The phonology of tone and intonation*. Cambridge: CUP.

[5] Kingston, J. 2007. The phonetics-phonology interface. In: de Lacy, P. (ed). *The Cambridge handbook of phonology*. Cambridge: CUP, 401-435.

[6] Lehiste, I. 1970. *Suprasegmentals*. Cambridge, Mass.: M.I.T. Press.

[7] Liberman, M., Sproat, R. 1992. The stress and structure of modified noun phrases in English. In: Sag, I. (ed). *Lexical Matters*. Chicago: University of Chicago Press, 131-182.

[8] Ostendorf, M., Price, P., Shattuck-Hufnagel, S. 1996. *Boston University Radio Speech Corpus*. Philadelphia: Linguistic Data Consortium.

[9] Sluijter, A. M. C., van Heuven, V. J., Pacilly, J. J. A. 1997. "Spectral balance as a cue in the perception of linguistic stress", *J. Acoust. Soc. Am.* 101(1), 503-513.

[10] Streefkerk, B. M., Pols, L. C. W., ten Bosch, L. F. M. 1999. "Acoustical features as predictors for prominence in read aloud Dutch sentences used in ANN's", *Proc. Eurospeech* Budapest, 551-554.

---

[i] This work was made possible by grant PL151/5-1 from the *Deutsche Forschungsgesellschaft* (DFG).

[ii] Jitter is another measure that might also be considered. As this reflects the variation of glottal pulse timing, it represents a concept similar to the mean average pitch slope, and correlates highly with it (r=0.76, p<0.01 for the right member). Because of this high correlation, jitter was not included to avoid collinearity in the data.