Comparing word and syllable prominence rated by naïve listeners

Denis Arnold[1], *Bernd Möbius* [2], *Petra Wagner* [3],

[1] Language and Speech Communication, University of Bonn, Germany
[2] Department of Computational Linguistics and Phonetics, Saarland University, Germany
[3] Faculty of Linguistics and Literature, University of Bielefeld, Germany

dar@sk.uni-bonn.de, moebius@coli.uni-saarland.de, petra.wagner@uni-bielefeld.de

Abstract

Prominence has been widely studied on the word level and the syllable level. An extensive study comparing the two approaches is missing in the literature. This study investigates how word and syllable prominence relate to each other in German. We find that perceptual ratings based on the word level are more extreme than those based on the syllable level. The correlations between word prominence and acoustic features are greater than the correlations between syllable prominence and acoustic features.

Index Terms: prosody, prominence, word, syllable, acoustic correlates

1. Introduction

The present study aims to investigate the differences between perceptual prominence ratings on the word level and the syllable level obtained from näive listeners. Following [1] we define perceptual prominence as a gradual difference in the perceptual salience of linguistic units like syllables or words compared to their neighbors. A broad body of work concerning the study of prominence has been carried out. Many aspects of prominence, like correlations with acoustics [2, 3], influence of linguistic knowledge and expectations on the perception of prominence [4, 5, 6, 7, 8], prediction of prominence [9], as well as automatic prominence labeling [10, 11] and the differences between machine labeled prominence and human labeled prominence [12] have been studied. The question what would be an optimal rating scale remains unanswered so far [13, 14]. The relation between perceptual prominence and acoustics is controversial since the different studies find different acoustic features to contribute more to perceptual prominence than the others. A lot of the differences in findings on perceptual prominence may be due to differences in the methods chosen. The findings from [4, 7, 9, 11, 13, 14] are based on word prominence, while the findings from [5, 6, 8, 10] are based on syllable prominence. Streefkerk [4] compared rating on words and syllables using a binary scale. She concluded that the ratings on word level seem to be more robust. In the present study we aimed to obtain prominence ratings on words and syllables from naïve listeners. We were looking for the differences emerging from rating the different units and had a closer look at the relation between prominence and the correlations with common acoustic features like duration, f0 and intensity.

2. Experiment

2.1. Method

One goal of the study was to explore the relation between the ratings on the word level and on the syllable level. Another goal was to find whether word or syllable prominence ratings result in a better correlation with acoustic features. 36 native speakers of German without any reported hearing impairment participated in our experiment. We had two groups rating exactly the same material with a 31-point scale that was also used in [6, 8]. The first group rated the prominence on the word level while the other group rated the material on the syllable level. We recorded 15 sentences in German spoken by a trained female speaker in a sound-treated studio and stored them as 16-bit, 44.1 kHz wave files. Two sentences contained only monosyllabic words. These sentences were used as control sentences. Because the units in these sentences were identical for rating on the word and the syllable level, we expect to find high correlations between the prominence ratings in these two sentences. If we do not find high correlations, we must expect that the two rating groups differ to a significant extent in their ability to rate perceptual prominence. Two sentences use one disyllabic word, while the other words where monosyllabic. We manipulated the position of the stressed syllable to capture a possible effect on the perception of word prominence. The other words where identical in both sentences to keep every linguistic aspect as constant as possible. The following sentences were used: (italic typesetting and underline indicates the stressed syllable) :

- In *Lon*don ist es echt schön. (It is really nice in London)
- In Ber*lin* ist es echt schön. (It is really nice in Berlin)

We constructed sentences in the same way for trisyllabic words and words containing four syllables. We also constructed two sentences of four monosyllabic words and two disyllabic words:

- Er fährt im *Ju*li nach Lu*zern*. (He will go to Luzern in July)
- Er fährt im Au*gust* nach *Zü*rich. (He will go to Zürich in August)

Following our definition of prominence we expected that the unit "nach" will be more prominent when the first sentence is rated on the syllable level than when the sentence is rated on the word level. We expected this effect to be smaller for the second sentence.

2.2. Rating Experiment

All subjects were randomly assigned to one of the two groups. The first group rated the perceptual prominence on the word
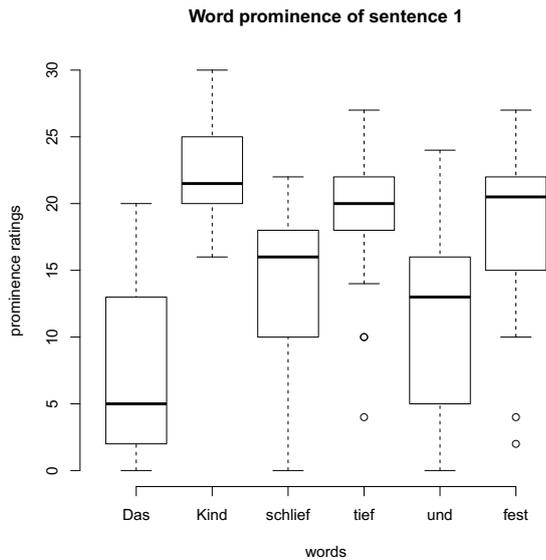
28 − 31 August 2011, Florence, Italy

Word prominence of sentence 1

Figure 1: *Word prominence rating for sentence 1. Das Kind schlief tief und fest - The child was sound asleep.*
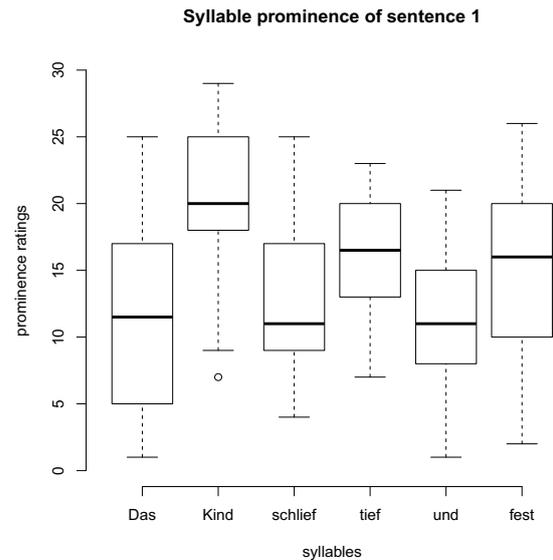
Syllable prominence of sentence 1

Figure 2: *Syllable prominence rating for sentence 1. Das Kind schlief tief und fest - The child was sound asleep*

level, the other group rated on the syllable level. The experiment was carried out by means of a software coded in Java that was also used in [8]. All instructions were presented on the computer screen. The stimuli were presented via headphones and the subjects were asked to judge the prominence of each linguistic unit (words for the first group, syllables for the second group) using sliders with a 31-point scale on the graphical user interface. The orthographic representation of each linguistic unit was shown above the corresponding slider. The subjects had the opportunity to listen to the signals again if they wished to, using a button on the graphical user interface. They had to rate all units of a sentence before they could proceed. The order for the presentation of the stimuli was randomized for each subject.

2.3. Analysis of acoustic features

All sentences where manually labeled on the word and the syllable level. The durations of the words and syllables where extracted from the label files and maximum f0 in the respective unit as well as the mean intensity of the unit where computed by means of scripts in praat [15].

3. Results

All statistical analyses were computed by means of the free statistics program and language R [16].

3.1. Correlation of the control sentences

We had two control sentences that contained monosyllabic words only. We found correlations of .94 between the ratings of word prominence and syllable prominence for both sentences. Thus we can assume that both groups were equally capable of completing the task. Figures 1 and 2 show the first sentence rated by the two groups. The group rating word prominence assigned more extreme values than the group rating syllable prominence.

3.2. Number of playbacks and time consumption

Rating on the syllable level should take longer than rating on the word level, since there are more units to be rated. Jensen and Tøndering [13] used number of playbacks and time consumption as indicators of rater effort. Table 1 shows the mean time consumption and number of playbacks as well as normalized values. For normalization we divided the number of playbacks and the time consumption by the number of linguistic units in the respective sentence. Both time consumption and number of playbacks are significantly lower for ratings obtained on the word level than on the syllable level (Table 1). After normalization only the difference in number of playbacks is significant.

3.3. Prominence ratings

There is no significant difference between the mean of the word prominence and the mean of the syllable prominence. We find more extreme results for prominence on the word level than for prominence on the syllable level on both ends of the scale (Figures 3 and 4). The word prominence of polysyllabic words is greater or equal to the maximum of the syllable prominence of the syllables the word contains, while the word prominence of monosyllabic words can be significantly smaller than syllable prominence because of changes in prominence of the immedi-

Table 1: *Mean times consumption and number of playbacks for ratings obtain on the word and on the syllable level. Normalized values have been divided by the number of linguistic units in the sentence.*

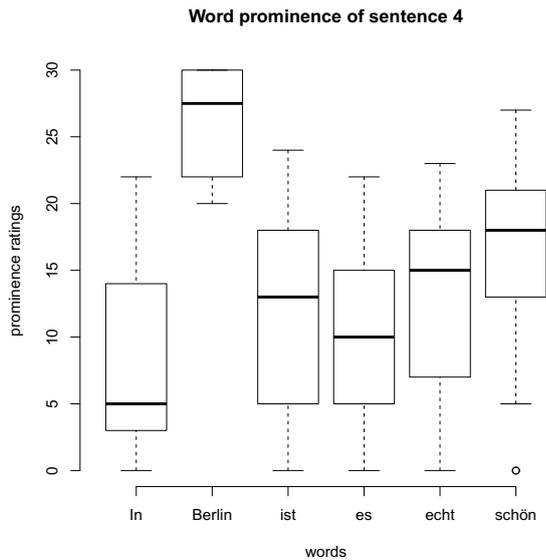|	Word	Syllable	Significant
Time [sec]	24.1	32.8	*
Playbacks	0.48	1.16	*
Time (normalized) [sec]	4.48	4.54	
Playbacks (normalized)	0.08	0.16	*

Word prominence of sentence 4

Figure 3: *Word prominence rating for sentence 4. In Berlin is es echt schön - It is really nice in Berlin.*
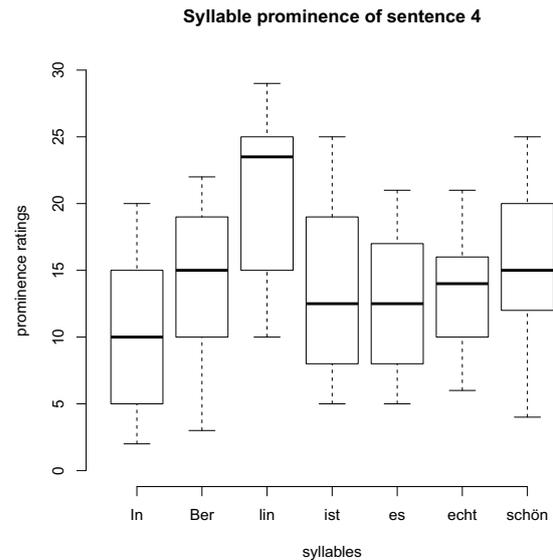
Syllable prominence of sentence 4

Figure 4: *Syllable prominence rating for sentence 4. In Berlin is es echt schön - It is really nice in Berlin.*

ate neighbor (Figures 5 and 6).

3.4. Acoustic correlates

We were interested in the question whether the correlations between prominence and acoustic features like duration, f0, and intensity depend on the type of linguistic unit on which the prominence is rated. We calculated Pearson's product-moment correlations between the prominence values and the acoustic features. We found stronger correlations between duration and prominence than between f0 and prominence, the latter being roughly on the same level as the correlations between intensity and prominence. The correlations between word prominence and its acoustic correlates are higher than for syllable prominence and its acoustic correlates (Table 2).

4. Discussion

The prominence ratings on the different linguistic levels were highly correlated for the two control sentences. This shows that both groups were equally capable of completing the task. Time consumption and number of playbacks differ significantly between the two groups. As expected, the subjects rating word prominence were faster and listened to the signal less often. After normalizing of both time consumption and number of playbacks with the number of units to be rated, the significant difference in time consumption disappeared. The difference in number of playbacks was still significant. Since time consumption

was almost equal after normalization, it seems that most of the time was consumed with moving the sliders to rate the stimuli. Jensen and Tøndering [13] used time consumption and number of playbacks as a measure for rater effort when evaluating different rating scales. When we employed this concept to our task, we found that the number of playbacks is a better measure for rater effort, which is the opposite of their findings. It needs further investigation to clarify whether any of these measures is a good predictor of rater effort.

Subjects assigned more extreme values to word prominence than to syllable prominence. The other way around, we can say subjects rating syllables had a stronger tendency toward the mean. It is questionable whether this was due to uncertainty in the usage of the concept of the syllable or whether this was an effect of the greater number of units in a sentence.

We computed correlations between prominence ratings and duration, f0, and intensity. Correlations between prominence and duration were higher than the correlations between prominence and f0 as well as the correlations between prominence and intensity. In all cases we found higher correlations between word prominence and acoustic features than for syllable prominence and acoustic features. The proportion between the correlations of prominence and intensity compared with the correlations between prominence and f0 did not change between word and syllable level. They had an equal value on the respective level. The situation was different for duration, where the correlation with prominence was equal to that between f0 and prominence and between intensity and prominence on the syllable level but the correlation between duration and prominence was higher on the word level. This may be an artifact of the stimulus design, since most words that were most prominent in their respective sentence were polysyllabic words.

5. Conclusion

The present study compared the rating of perceptual prominence on word and syllable level by naïve listeners. The re-
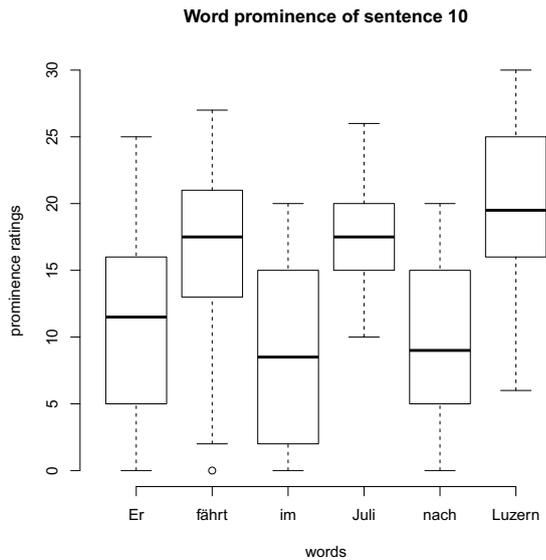
Table 2: *Correlation between prominence ratings and acoustic correlates. p < .001 for all correlations*

	Word prominence	Syllable prominence
Duration	.69	.41
Maximum f0	.54	.40
Intensity	.53	.39

Word prominence of sentence 10

Figure 5: *Word prominence rating for sentence 10. Er fährt im Juli nach Luzern - He will go to Luzern in July.*
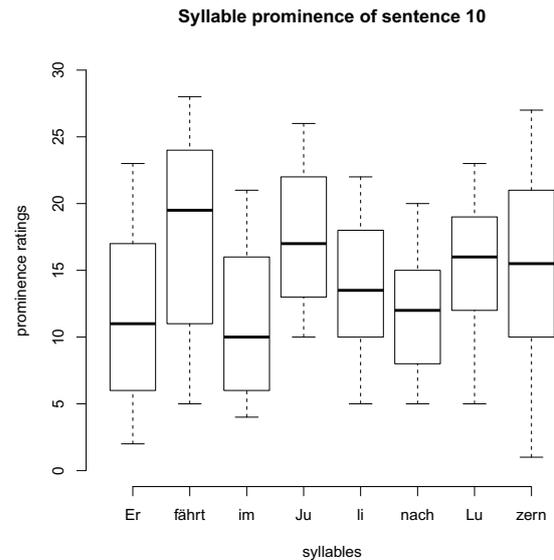
Syllable prominence of sentence 10

Figure 6: *Syllable prominence rating for sentence 10. Er fährt im Juli nach Luzern - He will go to Luzern in July.*

sults found with the two approaches differ a lot. We find that the contribution of acoustic features to perceptual prominence shifts when comparing word prominence and syllable prominence. Word prominence fits the acoustic data better than syllable prominence. It can be concluded that naïve listeners can handle the concept of word prominence more easily than the concept of syllable prominence. The relation between word prominence of polysyllabic words and the prominence of the syllables in the word is rather complex. For our data the word prominence of polysyllabic words is greater or equal to the maximum of the syllable prominence. Prominence of monosyllabic words can be significantly different from the syllable prominence of the same word due to differences in the word and syllable prominence of polysyllabic neighbors. This result shows that context is important in the perception of the prominence of a unit. There are significant differences between the ratings of word prominence and syllable prominence, resulting in significant differences in the correlations between rated perceptual prominence and the acoustic correlates of prominence. We conclude that results from studies using the concept of word prominence cannot be compared in a straightforward way to findings from studies using syllable prominence.

6. References

[1] Wagner, P., "Wahrnehmung und Vorhersage deutscher Betonungsmuster", Universität Bonn. PhD-Thesis, 2002. Online: http://hss.ulb.uni-bonn.de/2002/0054/0054.htm, accessed on 29 Mar 2011.

[2] Turk, A. E. and Sawusch, J. R., "The processing of duration and intensity cues to prominence", Journal of the Acoustical Society of America 99, 3782-3790, 1996.

[3] Kochanski, G., Grabe, E., Coleman, J., and Rosner, B., "Loudness predicts prominence: fundamental frequency lends little". Journal of the Acoustical Society of America 118, 1038-1054, 2005.

[4] Streefkerk, B., "Prominence - Acoustic and lexical/syntactic correlates", Utrecht: LOT, 2002.

[5] Eriksson, A., Grabe E. und Traunmüller, H., "Perception of syllable prominence by listeners with and without competence in the tested language". Proceedings Speech Prosody 2002, Aix-en-Provence, 275-278, 2002.

[6] Wagner, P., "Great Expectations - Introspective vs. Perceptual Prominence Ratings and their Acoustic Correlates", Proceedings of Interspeech 2005, Lisbon, 2381-2384, 2005.

[7] Cole, J., Mo, Y. and Hasegawa-Johnson, M., "Signal-based and expectation-based factors in the perception of prosodic prominence", Laboratory Phonology 2010, 1:2, 425-452, 2010.

[8] Arnold, D., Wagner, P. and Möbius, B., "The effect of priming on the correlations between prominence ratings and acoustic features", Proceedings of Speech Prosody 2010, Chicago, 2010.

[9] Widera, C., Portele, T. and Wolters, M., "Prediction of word prominence", Proceedings of Eurospeech 1997, 999-102, 1997.

[10] Tamburini, F. "Automatic prosodic prominence detection in speech using acoustic features: an unsupervised system", Proceedings of Eurospeech 2003, 129-132, 2003.

[11] Wang, D. and Narayanan, S.; , "An Acoustic Measure for Word Prominence in Spontaneous Speech." IEEE Transactions on Audio, Speech, and Language Processing, vol.15, no.2, 690-701, 2007

[12] Goldman, J.-P., Auchlin, A., Roekhaut, S., Simon, A. C., Avanzi M., "Prominence perception and accent detection in French. A corpus-based account", Proceedings of Speech Prosody 2010, Chicago, 2010.

[13] Jensen, C. and Tøndering, J. , "Choosing a Scale for Measuring Perceived Prominence", Proceedings of Interspeech 2005, Lisbon, 2385-2388, 2005.

[14] Grover, C., Heuft, B. und van Coile, B., "The reliability of labeling word prominence and prosodic boundary strength", Proceedings of the ESCA Workshop on Intonation, Athens, 165-168, 1997.

[15] Boersma. P. and Weenink, D., "Praat: doing phonetics by computer" (Version 5.1.31), Computer Program, 2010.

[16] R Development Core Team. "R: A language and environment for statistical computing.", R Foundation for Statistical Computing, Vienna, 2010.