

Prominence and Coreference – On the Perceptual Relevance of F0 Movement, Duration and Intensity

Stefan Baumann¹, Anna Roth²

¹I/L-Phonetik, Universität zu Köln, Germany

²Institut für Linguistik, Goethe-Universität Frankfurt am Main, Germany

stefan.baumann@uni-koeln.de, a.roth.unifrankfurt@gmail.com

Abstract

We conducted a web-based experiment on German testing the perception of an element's prosodic prominence in relation to its status as a potential coreferent of an antecedent. Data were elicited by asking subjects to judge the probability of a coreference relation between a context noun (antecedent) and a target word (anaphor), whose lexically stressed syllable was manipulated as to the parameters F0 movement, duration and intensity. Results suggest a direct but inverse relationship between prominence and coreference judgements indicating that the likelihood of a coreference interpretation decreases with increasing prosodic prominence. F0 movement turned out to be the dominant cue for prominence – as the main trigger for the perception of pitch accents – with rises being perceived as more prominent than falls. In turn, lack of tonal movement probably led to perceived deaccentuation and thus favoured the evaluation of a target word as being coreferential with an antecedent. Duration was found to be a significant factor as well, while intensity did not prove to be relevant for the task given. Thus, the present study with its revised methodology adds new aspects to the debate of which parameters are crucial for prominence perception, directly linking it to the investigation of information structure.

Index Terms: prosody, coreference, prominence, perception, pitch accent, givenness

1. Introduction

A central aspect in the analysis of an utterance's information structure is the information status of its elements, defining whether an element can be regarded as Given, Accessible or New (e.g. [1]). In many accounts, Givenness is equated with coreference, i.e. referential identity (e.g. [2]) (although it can be shown that a lexical level of description is also relevant for a comprehensive account of information status; see [3]). In (1a), e.g., *a lasagne* introduces a New referent, while in (1b), *the tasteless stuff* is Given, since it stands in a coreference relation to the previously mentioned *lasagne*:

- (1) a. Yesterday, a friend of mine prepared a laSagne for me.
 b. I found it hard to enJOY the tasteless stuff. [4:16]

Prosodically, New referents are mostly marked by pitch accents in West Germanic languages (in the example, nuclear accents are indicated by capital letters, as in *lasagne* in (1a)), whereas coreferential anaphors are often deaccented (as *the tasteless stuff* in (1b)) or at least marked by a clearly attenuated, i.e. less prominent, prosody. Thus, there seems to be a more or less direct link between the information status of

an element and the prominence of its prosodic marking. We will base our study on this general assumption.

There has been a long-standing debate on which acoustic parameters are most relevant for the perception of post-lexical prominence, in particular change in fundamental frequency (F0) (perceptual level: pitch movement), longer duration (perceptual level: increased length), or higher intensity (perceptual level: increased loudness). Another relevant parameter is vowel quality, differentiating between full and reduced vowels. Previous perception studies presented quite different results, assigning the greatest importance for prominence judgements either to F0 variation (e.g. [5], [6]), duration (e.g. [7]), intensity (e.g. [8]), or to a combination of duration and intensity (a factor called 'total amplitude' by [9]). Furthermore, there is conflicting evidence for the question of whether rising or falling pitch accents are perceived as more prominent (cf. [4] and [10]). See also [11] for an overview of acoustic and perceptual correlates of prosodic prominence.

There are a number of recent empirical (perception or production) studies on prosody and information status in German but they either do not differentiate between various levels of Givenness (conflating coreference, semantic-pragmatic accessibility and lexical repetition, e.g. [12]) or they do not relate an element's information status directly to its degree or level of prosodic prominence ([13], but see [4]). In the experiment we report on in the present paper we explicitly ask subjects to judge the probability of a coreference relation between an antecedent and a potential anaphor, the latter being manipulated as to its degree of prosodic prominence. This elicitation technique allows us to analyze the relation between coreference and perceived prominence, expressed in the following hypotheses:

Hypothesis 1: Acoustic/prosodic prominence correlates with perceived (non-)coreference – here: the more prominence-lending parameters (e.g. longer duration, higher intensity) are present, the less likely is a referent to be perceived as coreferential with an antecedent.

Hypothesis 2: Prosodic parameters vary in their relevance for the perception of prominence/non-coreference: F0 movement > duration > intensity.

2. Method

2.1. Test material

Three female proper names were chosen as target words: *Tamara*, *Pamela* and *Simone*. All of them are trisyllabic in German, with lexical stress on the second syllable. The stressed syllables contain the bilabial nasal /m/ in the onset and one of the long vowels /a:/, /e:/, /o:/ in the syllable rhyme. We created three test sentences. These contain a time specification (weekday), the subject pronoun *ich* ('I'), one of the three proper names as an accusative object, and a predicate

composed of the auxiliary verb *haben* ('have') and a past participle of one of the transitive irregular verbs *getroffen* ('met'), *gesprachen* ('talked') or *gesehen* ('seen'):

- (2) Montag habe ich Tamara getroffen.
'On Monday I met Tamara.'
- (3) Dienstag habe ich Pamela gesprochen.
'On Tuesday I talked to Pamela.'
- (4) Freitag habe ich Simone gesehen.
'On Friday I saw Simone.'

The sentences were spoken by a female native German speaker and were recorded with an Edirol R-44 in a sound attenuated booth. The prosody of the recorded sentences was controlled for. In particular, three prosodic parameters of the stressed syllables in the target words were manipulated, namely F0 movement, duration and intensity.

F0 was varied to create rising, falling and level contours. For the rising condition, the stressed syllable starts at 190 Hz and rises to 240 Hz at the end of the syllable. For the falling condition, the F0 starts at 240 Hz at target syllable onset and falls down to 190 Hz at syllable offset. In the level condition, we assigned a value of 190 Hz throughout the whole target word. Figure 1 shows a *Praat* [14] screenshot for test sentence (2) with a rising pitch (solid line), a falling pitch (dotted line) and no pitch movement (dashed line) on the stressed syllable *ma* in the target word *Tamara*:

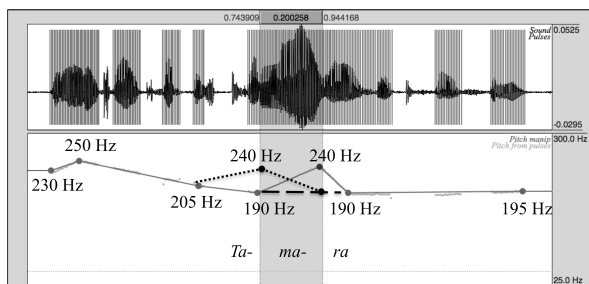


Figure 1: Screenshot of the sentence Montag habe ich Tamara getroffen ('On Monday I met Tamara') in the *Praat* manipulation window. The stressed syllable of *Tamara* is shaded; it contains an F0 rise (solid line; first condition) an F0 fall (dotted line; second condition) or no F0 movement (dashed line; third condition).

As to duration, two values were used: long syllables with 200 ms and short syllables with 150 ms. Similarly, we created two intensity levels: loud syllables with a maximum of 57 dB and soft syllables with a maximum of 47 dB. While F0 and duration were manipulated in *Praat*, we used the audio editor *Audacity* [15] to adjust the intensity of the target syllables. All combinations of the acoustic parameters were implemented in each sentence, resulting in 12 manipulated versions of each target sentence. Thus, 36 manipulated sentences for the perception experiment were generated in total. An overview of the manipulations is provided in Table 1.

The prosodic context for the target syllables was held constant throughout. That is, the weekday always carried a H* pitch accent (following [16]), with an F0 rise from 230 Hz to 250 Hz. It was crucial to create at least one pitch accent in the

sentence which turns into the nuclear accent in cases of largely attenuated prominence of the target word. In this condition, the sentences would have lacked a pitch accent altogether without the early accent in the phrase, which would have sounded highly unnatural. Furthermore, the pitch height of the personal pronoun (*ich* 'I') immediately preceding the target word constantly had a value of 205 Hz, and the intonation phrase always ended at 195 Hz (cf. Figure 1).

Table 1: Overview of manipulated prosodic parameters on target words.

	Stressed syllable of target word	
	F0 movement ('Tone')	rise
fall		240Hz – 190Hz
none		190Hz – 190Hz
Duration	long	200 ms
	short	150 ms
Intensity	loud	57 dB
	soft	47 dB

Duration and intensity values for the rest of the target sentence were also controlled. In particular, the first and third syllable of the target word invariably had a duration of 150 ms and 130 ms, respectively, while both syllables were set to a maximum intensity of 50 dB.

2.2. Procedure

We conducted an online perception experiment via an open URL, using the professional software package *SoSci Survey* [17]. All 36 manipulated sentences were pseudo-randomized and matched with a respective context question. We created 12 different context questions for each target sentence, in order to diversify the task for the subjects. All questions followed the same structure; an example is given in (5):

- (5) Hast du deine Cousine getroffen / gesprochen / gesehen?
'Have you met / talked to / seen your cousin?'

Twelve concrete, female nouns (as accusative objects) were used, which served as potential antecedents for the proper names in the target sentences: *Cousine* 'cousin', *Schwester* 'sister', *Nachbarin* 'neighbour', *Klassenkammeradin* 'classmate', *Tennispartnerin* 'tennis partner', *Arbeitskollegin* 'co-worker', *Mitbewohnerin* 'roommate', *Kommilitonin* 'fellow student', *Schulfreundin* 'schoolmate', *Teamkollegin* 'teammate', *Bekannte* 'acquaintance' and *Trainerin* 'trainer'.

The context questions were presented only visually, while the manipulated test sentences were presented auditorily. The participants controlled when to start a stimulus and were free to listen to it as many times as they chose. Afterwards, the subjects had to answer a question by rating the probability of a (non-)coreference relation between the proper name in the target sentence and the noun in the context question. An example of the questions is shown in (6):

- (6) Für wie wahrscheinlich halten Sie es, dass es sich bei der Cousine um Tamara handelt?
'How likely do you think it is that the cousin is Tamara?'

Judgements were given via a horizontal scroll bar whose poles were labelled *sehr wahrscheinlich* 'very likely' (left side) and *sehr unwahrscheinlich* 'very unlikely' (right side). Figure 2 gives an example of the setup.

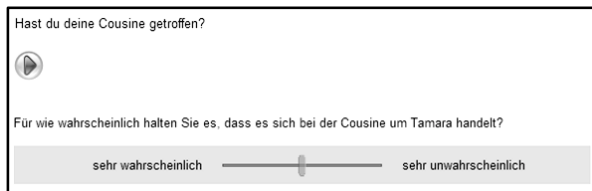


Figure 2: Screenshot of the response submission setup in the online questionnaire, including a context question ('Have you met your cousin?'), the button for playing the target sentence, and the rating task ('How likely do you think it is that the cousin is Tamara?') with the scroll bar.

The pole 'very likely' (= low values on the scale) corresponds to the judgement of coreference between the context noun (e.g. *Cousine*) and the proper name (e.g. *Tamara*). That is, the target word is perceived as Given information. In contrast, the pole 'very unlikely' (= high values on the scale) corresponds to the perception of a non-coreference relation between the context noun and the proper name. This means that the target word displays New information.

The experimental setup also included an instruction and a short practice section prior to the main experiment. The subjects were asked to conduct the experiment in a quiet environment and to wear headphones when listening to the stimuli. The entire procedure took approximately 10 to 15 minutes per subject.

2.3. Subjects and analysis

We collected judgements of 40 native speakers of German (32 female, 8 male), aged between 19 and 62 years. The mean age was 27 years. They grew up in nine different German Federal States. The subjects were no experts in the analysis of spoken language and did not report any hearing impairment.

The elicited judgements were encoded on an interval scale, illustrated as a horizontal continuous line in the experimental condition, ranging from 1% at the left pole ('very likely') to 100% at the right pole ('very unlikely'). For the statistical analysis, we used these percentage values as the dependent variable in a linear mixed model, with TONE (i.e. F0 movement), DURATION, INTENSITY and TEST WORD as fixed effects, and SUBJECT as a random effect.

3. Results and discussion

First of all, the subjects covered the whole range of the scale when judging the probability of coreference relations. This suggests that the stimuli were reasonably balanced for the type of task presented. Nevertheless, there is a slight bias towards the left pole of the scale (coreference judgements), maybe because a syntactically unmarked answer to a polar question

(i.e. lacking an explicit 'yes' or 'no') tends to be interpreted as a confirmation.

Results reveal a significant main effect for TONE ($p < 0.001$) in the coreference judgements. Figure 3 shows that the highest values were assigned to the target words if they carried a rising F0 movement, and that a falling contour still triggered clearly higher values than no F0 movement. That is, non-coreference (Newness) judgements correlate with tonal movement, and more so with rises (presumably more prominent; mean = 66.62%) than with falls (presumably less prominent; mean = 51.51%) whereas coreference (Givenness) judgements correlate with – the least prominent – lack of tonal movement (mean = 24.05%).

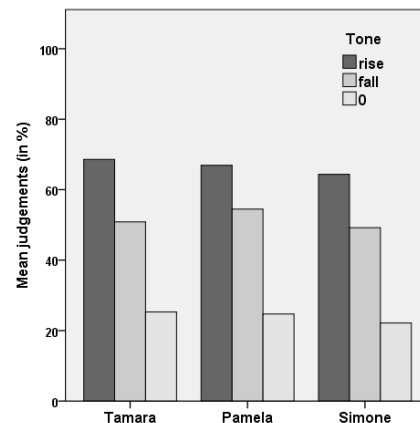


Figure 3: Bar plot of mean judgements on (non-)coreference-probability scale with respect to the factor TONE; '0' stands for lack of tonal movement; all subjects pooled.

The factor DURATION shows a significant main effect as well ($p < 0.01$), with longer syllables favouring the subjects' impression that a coreference relation between context noun and proper name is unlikely (means = 49.63% for long syllables, 45.16% for short syllables; see Figure 4). Thus again, an increase in prosodic prominence leads to a decrease in coreference judgements.

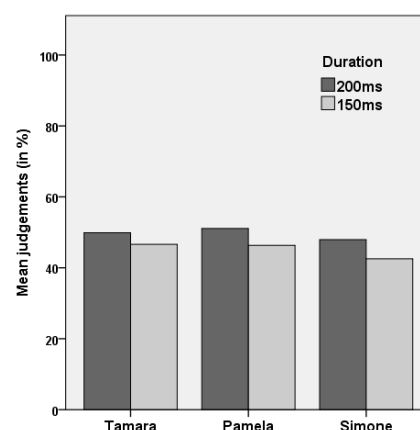


Figure 4: Bar plot of mean judgements on (non-)coreference-probability scale with respect to the factor DURATION; all subjects pooled.

No significant effect on coreference evaluations has been found for the parameter *INTENSITY* ($p=0.131$), although the judgements for two test words display the expected tendency, namely higher values (indicating lower probability of coreference) for test words carrying a louder stressed syllable (see Figure 5). In fact, the manipulation method for intensity may have been particularly prone to result in an unnatural outcome of this parameter (e.g. no adequate variation of spectral tilt) which may explain the somewhat inconsistent judgements.

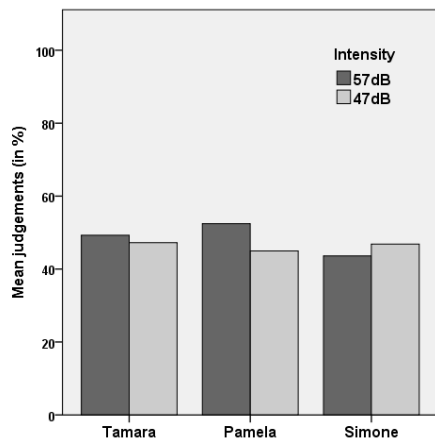


Figure 5: Bar plot of mean judgements on (non-)coreference-probability scale with respect to the factor *INTENSITY*; all subjects pooled.

The factor *TEST WORD* does not show a significantly distinct distribution over all three parameters either ($p=0.085$). Thus, the quality of the stressed vowel in the target word did not have an influence on the task.

Finally, no significant interactions were found between the fixed factors investigated. This result came as a surprise since an incremental effect of the prominence-lending parameters with respect to the (non-)coreference judgements would have been expected. That is, e.g., a target word with a rising F0 movement was not found to be interpreted as 'less coreferential' with an antecedent if it was *accompanied* by increased duration and intensity.

This notwithstanding, both hypotheses could generally be confirmed. As to Hypothesis 1, there are clear indications for an inverse but straightforward relation between perceived prosodic prominence and listeners' judgements of coreference. For example, if a target word such as *Tamara* in *Montag habe ich Tamara getroffen* ('On Monday I met Tamara') is realized in a prosodically attenuated manner, e.g. by lack of tonal movement or relatively short and soft syllables (in particular the lexically stressed syllable), the likelihood is high to interpret *Tamara* as the same referent as the cousin in the context sentence *Hast du deine Cousine getroffen?* ('Have you met your cousin?'). In contrast, a rising tone as well as increased duration on the lexically stressed syllable (*Tamara*) enhances the likelihood of the word to be interpreted as New information, which thus cannot be coreferential. However, as mentioned above, an additive effect of the parameters could not be found.

As to Hypothesis 2, the factor *tonal movement* is clearly dominant, while duration – which shows a significant effect as

well – and intensity are much weaker cues (contrary to the claims proposed e.g. by [7] or [8]). This primacy of F0 variation is an old assertion already made by Fry [5], who based his findings on perception experiments in English, and Bolinger [6], who closely linked prominence with pitch accents (at least for West Germanic languages). Although pitch accents are combinations of F0 movement and increased duration and intensity, an F0 change in the vicinity of a stressed syllable is the defining characteristic of a pitch accent.

Moreover, the present experiment adds to the findings of more recent studies on West Germanic languages by suggesting that the shape or *type* of pitch accent does not only play a crucial role for the evaluation of Given or New information but also for the perception of prosodic prominence – with low and falling accents being perceived as less prominent than high and rising accents ([4], [13], [18]). In particular, there is evidence that the most important factor is the question of whether the *on glide* to the accented syllable is rising or falling ([19]).

4. Conclusions

The present study relates the interpretation of information status to the perception of prosodic prominence, assuming a basic link between them. Methodologically, it proved to be possible to use a gradient probability measure based on distinctions at various levels of prosodic parameters to evaluate the intrinsically categorical distinction between coreference (i.e. referential identity) and lack of coreference.

Results suggest that tonal movement is the central cue to perceived prominence, since it is characteristic of a pitch accent. Nevertheless, it is generally agreed on today that non-tonal cues such as duration, intensity and vowel quality are highly relevant for indicating the presence of *rhythmic* prominences (e.g. [20]) as well, which, however, may not contribute to the marking of information status in the same way as pitch accents do.

5. References

- [1] Chafe, W., "Discourse, Consciousness, and Time", University of Chicago Press, Chicago/London, 1994.
- [2] Prince, E. F., "Toward a Taxonomy of Given-New Information", in P. Cole [Ed], *Radical Pragmatics*, 223-256, Academic Press, New York, 1981.
- [3] Baumann, S. and Riester, A., "Referential and Lexical Givenness: Semantic, Prosodic and Cognitive Aspects", in G. Elordieta and P. Prieto [Eds], *Prosody and Meaning*. Mouton De Gruyter, Berlin, New York, [Interface Explorations 25], 119-162, 2012.
- [4] Baumann, S. and Riester, A., "Coreference, Lexical Givenness and Prosody in German", in J. Hartmann, J. Radó and S. Winkler [Eds], [Special Issue "Information Structure Triggers"], *Lingua* 136:16-37, 2013.
- [5] Fry, D.B., "Experiments in the Perception of Stress", *Language and Speech* 1:126-152, 1958.
- [6] Bolinger, D., "A Theory of Pitch Accent in English", *Word* 14:109-149, 1958.
- [7] Cole, J., Mo, Y. and Hasegawa-Johnson, M., "Signal-based and expectation-based factors in the perception of prosodic prominence", *Laboratory Phonology* 1:425-452, 2010.
- [8] Kochanski, G., Grabe, E., Coleman, J. and Rosner, B., "Loudness Predicts Prominence; Fundamental Frequency Lends Little", *J. Acoustical Society of America* 11(2):1038-1054, 2005.
- [9] Beckman, M. E., "Stress and Non-Stress Accent", Foris, Dordrecht, 1986.

- [10] Hermes, D. and Rump, H., "Perception of prominence in speech intonation induced by rising and falling pitch movements", *Journal of the Acoustical Society of America* 90:97-102, 1994.
- [11] Terken, J., and Hermes, D., "The perception of prosodic prominence", in M. Horne [Ed] *Prosody: Theory and experiment*, Kluwer, Dordrecht, 89-127, 2000.
- [12] Féry, C. and Kügler, F., "Pitch accent scaling on given, new and focused constituents in German", *Journal of Phonetics* 36:680-703, 2008.
- [13] Röhr, C. and Baumann, S., "Decoding information status by type and position of accent in German", *Proceedings 17th ICPHS, Hong Kong, China, 1706-1709*, 2011.
- [14] Boersma, P. and Weenink, D., "Praat: doing phonetics by computer", Version 5.3.60, retrieved 4 March 2012 from <http://www.praat.org/>.
- [15] "Audacity" [Computer software audio editor], Version 2.0.5, retrieved 30 October 2013 from: <http://audacity.sourceforge.net/>.
- [16] Grice, M., Baumann, S. and Benz Müller, R., "German Intonation in Autosegmental-Metrical Phonology", in S.-A. Jun [Ed.] *Prosodic Typology. The Phonology of Intonation and Phrasing*. Oxford University Press, Oxford, 55-83, 2005.
- [17] Leiner, D. J., "SoSci Survey" (Version 2.3.05-i) [Computer Software]. Available from <https://www.sosicisurvey.de>, 2013.
- [18] Pierrehumbert, J. B. and Hirschberg, J., "The Meaning of Intonational Contours in the Interpretation of Discourse", in P.R. Cohen, J. Morgan, M.E. Pollack, [Eds] *Intentions in Communication*. MIT Press, Cambridge, 271-311, 1990.
- [19] Ritter, S. and Grice, M., "The Role of Tonal Onglides in German Nuclear Pitch Accents", Oral presentation at *Phonetics and Phonology in Iberia*, Lisbon, 26 June 2013.
- [20] Calhoun, S., "The centrality of metrical structure in signaling information structure: A probabilistic perspective", *Language* 86:1-42, 2010.