Prosodic Marking of Information Status in Task-Oriented Dialogues

Christine T. Röhr, Tabea Thies, Stefan Baumann, Martine Grice

I*f*l Phonetik, University of Cologne, Germany

{christine.roehr; tabea.thies; stefan.baumann; martine.grice}@uni-koeln.de

Abstract

In the present paper we investigate the effect of information status on accent placement and accent types used in semi-spontaneous speech. As an elicitation method we use a 'spot-the-difference' task which provides natural (dialogue) but still controlled (task-oriented) speech data. The task has been shown to be an ideal testbed for the relation between prosody and discourse meaning. However, it has not been used in a fine-grained study of information status yet. This is done in the present study by applying the *RefLex* annotation scheme, which differentiates between a referential and a lexical level of givenness. The semi-spontaneous speech data indicate a systematic but probabilistic relation between prosodic prominence and an item's level of givenness. That is, the correlation between increasing newness and increasing prominence is predominantly reflected in a more frequent use of nuclear pitch accents as well as a less frequent use of deaccentuation. Both the referential and lexical levels of givenness turn out to have an incremental effect on the degree of an item's prosodic prominence. Consequently, the *combined* degree of givenness of a referent (reflected by a combination of RefLex labels) indicates an overall prominence value of the item's prosodic realization.

Index Terms: production, semi-spontaneous speech, information status/givenness, RefLex, prosody, accent placement, pitch accent

1. Introduction

It is well known that in intonation languages like German the marking of information status (given-new dimension) is an important linguistic function of prosody.

In particular, Chafe [1] and Prince [2] have shown that it is not sufficient to distinguish only given and new information but to take at least a third intermediate class of givenness into account, sometimes described as accessible or inferable information. In the last three decades, various labelling schemes have been designed to enable annotations of more fine-grained differences in an item's information status (e.g. different types of accessible information). However, following Baumann & Riester [3] none of these schemes have proven detailed enough to capture and distinguish all sorts of informational distinctions which are necessary to explain even the most elementary intonational patterns. They argue that for an adequate analysis of an item's information status in spoken language two levels of givenness have to be investigated: a referential and a lexical level. Accordingly they developed a new, two-layered, type of annotation system for information status of referring (and non-referring) expressions (called *RefLex*), which, moreover, does not only distinguish given and new but also intermediate classes of givenness/novelty. Referential information status is assigned at the level of DP and PP, whereas lexical information status applies at the word level or modified NP level. In Table 2 an overview of the scheme – divided into a referential and a lexical level – is presented. The overview is a simplification (cf. [4]) of a more comprehensive account and describes only those labels which play a role in the present study. For the entire scheme, consult [3]. Detailed annotation guidelines are about to be published.

With regard to prosody, differences in an item's level of givenness have been shown to be marked by nuclear pitch accent placement (e.g. [5], [6], [7]) and/or pitch accent type (e.g. [8], [9], [10], [11], [12], [13], [14]): The less given an item, the higher the prosodic prominence produced.

Baumann & Riester [4] investigated the impact of the information status categories at a referential and a lexical level (as proposed in their RefLex scheme) on the prosodic realization. For read speech they generally confirmed the relationship between information status and prosody showing a stepwise increase in prosodic prominence from given to new items, predominantly ordered according to the information status at the lexical level. However, the results have been found to be less clear in spontaneous speech.

In order to find out which combinations of information status levels serve as triggers for which intonational categories further speech corpora have to be built/used and annotated according to the RefLex annotation scheme. The aim of the present study is to provide semi-spontaneous speech data by using a rather natural test setting with two interlocutors (task-oriented dialogues). We thus set up a 'spot-the-difference' task (resembling the 'diapix' task by [15]) which involves two similar but not identical pictures, and two participants who cannot see each other's picture. The two subjects have the task of working together to find ten differences between the pictures, involving either missing or replaced items. This task has the advantage to elicit balanced speech from each participant, i.e. there are no predetermined 'Giver' and 'Receiver' roles, as in typical Map Task dialogues. It also facilitates the elicitation of repeated mentions of segmentally controlled expressions in a variety of syntactic positions and utterance types.

For the present study we generally hypothesize that the 'newer' (or less given) an item is (both at the referential and the lexical level) the more it is made prosodically prominent by a speaker. More precisely, we assume relative differences in the intonational marking of information status, indicated by different distributions and/or probabilities of prosodic categories. The categories we are looking at are a) accent placement and b) accent type for nuclear and prenuclear pitch accents (categorized according to GToBI [16]) - assuming an increase in prominence from left to right:

a) no accent < (postnuclear accent <) prenuclear accent < nuclear accent (cf. [17])

b) no accent < low accent (L*) < falling accent (H+L*, H+!H*), high accent (!H*, H*) < rising accent (L*+H, L+H*) (cf. also [18])

The RefLex categories are claimed to express an increase in the level of an item's newness from left to right:

a) referential level: r-given < r-bridging < r-unused < r-new

b) lexical level: l-given < l-accessible < l-new

2. Method

2.1. Test Material

We designed two pictures for a spot-the-difference task (see Fig.1). Both pictures show a picnic-setting: A girl and a boy are arranged on a meadow nearby a tree and a wooden hut.

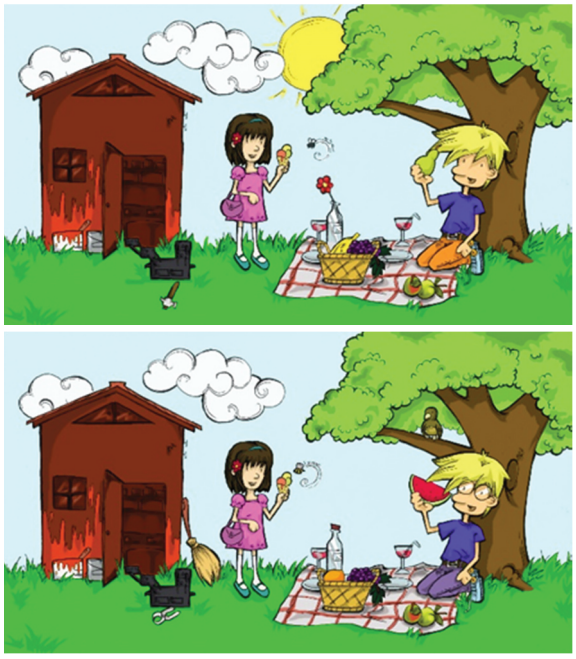

Figure 1: Pictures used for spot-the-difference task

Missing items	Replaced items		
Sonne sun	*Birne* pear	vs.	*Melone* melon
Blume flower	*Fliege* fly	vs.	*Biene* bee
Besen broom	*Hammer* hammer	vs.	*Zange* tongs
Brille glasses	*Banane* banana	vs.	*Orange* orange
Vogel bird	*orange* orange	vs.	*lila (Hose)* purple (trousers)

Table 1: Differences between the pictures of 'spot-the difference' task

Altogether the two pictures contain fifteen different target items/words that involve in total ten differences: On the one hand, there are five items that are only present in one of the two pictures (e.g. top picture: sun and flower) which are absent in the other picture. On the other hand, there are items in five particular positions that are in the other picture replaced by other items (e.g. the item in the boy's hand, top picture: pear vs. bottom picture: melon). A list of the ten differences or rather 'missing' and 'replaced' target items is given in Table 1.

The items on the pictures needed to be familiar and easily identifiable. Therefore, the target items were chosen from the *LEMO* database [19] which contains a set of 260 pictures that are standardized with regard to name agreement, image agreement, familiarity, and visual complexity.

2.2. Experimental setup

The experiment took place at the I*f*l Phonetik of the University Cologne and was composed of three parts: a priming phase, a practice section and the main experiment.

In the priming phase we familiarized the subjects with the target items that are shown in the pictures of the main experiment. This was necessary in order to guarantee an easy recognition of the images of the items in question and a uniform naming by all subjects. For the priming we used in total 40 images from the LEMO database (including the 15 target items of the experiment). The priming was conducted separately for each subject. We presented the priming elements successively on a computer screen and the subject's task was to read out loud the name of a depicted item. Subjects were instructed to only use those denotations in the main experiment for the particular items.

In the practice section we familiarized the subjects with the task. They were seated opposite to each other in a sound attenuated room. However, we placed a partition wall between the subjects so that they could not see each other. Both subjects received the same pair of two similar pictures involving five differences. To get an idea of the task of the main experiment they needed to discuss the five differences between the two pictures.

In contrast to the practice section in the following main experiment each subject only received one of the two pictures (see Fig.1). Since the subjects were not able to see each other's picture their task was to work together in order to discover the ten differences between their pictures. The main experiment was over as soon as the subjects identified all differences. The conversations between the subjects during the main part of the experiment were recorded acoustically using a headset condensator microphone for each subject.

2.3. Subjects

We recorded 12 native speakers of Standard German (six female and six male) in six dialogues, arranged in two female pairs, two male pairs and two mixed-gender pairs. They grew up in North Rhine-Westphalia or Rhineland-Palatinate and were aged between 19 and 29 years (mean = 22.8, SD = 2.7).

2.4. Analysis

In a first step, we produced transcripts of the dialogues. In a second step, we annotated the information status of nominal and also adjectival expressions in the conversations according to the RefLex annotation scheme (see Table 2). That is, at a referential level labels were applied to DPs, PPs and APs. The information status of the words within these phrases were separately labelled at a lexical level. Information that

expresses a contrast was additionally marked with a '(c)' attached to the lexical label.

In a third step we segmented and annotated the acoustic data in *Praat* [20]. At four segmental levels we a) annotated every spoken word, b) determined their part of speech (except for verbal expressions), c) classified the type of phrase that has been labelled at the referential level and d) marked the primary stressed syllable of all nouns, adjectives and verbs. The part of speech labels refer to the guidelines of the *Stuttgart-Tübingen-TagSet* [21]. At the phrasal level we distinguished noun phrases, prepositional phrases, adjective phrases, adverbial phrases and pronominal phrases.

Furthermore, we analyzed the prosodic realization of all sentences at two different levels. On a level of accent placement we marked for every word whether it was realized with no accent (coded as 0), with a postnuclear accent (1), a prenuclear accent (2) or a with a nuclear accent (in subordinate clauses coded as 3 and in main clauses coded as 4). On a tonal level we marked the positions of realized pitch accents and boundary tones and categorized their tonal configuration according to GToBI. In a last step the RefLex annotations were transferred to Praat.

In this paper we will present a descriptive analysis of the prosodic marking of the annotated RefLex categories. The results presented here are based on pooled GToBI and RefLex categories, even though we used the more fine-grained categories during the annotation process. That is, we basically distinguish between low (L*), falling (H+L*, H+!H*), high (!H*, H*) and rising (L*+H, L+H*) pitch accent types. RefLex categories were pooled according to the simplified overview given in Table 2.

Referential level (indicated by 'r-')	
r-given	coreferring anaphor that is present (immediately or displaced) in previous discourse context or contained in the text-external context (-sit)
r-bridging	non-coreferring anaphor dependent on previously introduced scenario
r-unused	globally unique discourse-new (non-anaphoric) entity which is generally known (-known) or identifiable from its own linguistic description (-unknown)
r-new	indefinite non-unique discourse-new entity
Lexical level (indicated by 'l-')	
l-given	markable is a repetition (-same), synonym (-syn), hypernym (-super) or holonym (-whole) of a previous expression
l-accessible	markable has an identical word stem (-stem) or is a hyponym (-sub) or meronym (-part) of a previous expression
l-new	markable is not related to another expression within the last five intonation phrases or clauses

Table 2: Simplified overview of annotation tags of the RefLex annotation scheme (cf. [3], [4])

3. Results

As an overall result, we found that the examined RefLex categories had an effect on the prosodic marking.

The distribution of accent *placement* (on all RefLex-annotated words) both as a function of the referential level and the lexical level of givenness shows that a word is more likely to get accented the less given it is. Figure 2 indicates that a decrease in referential and lexical givenness is reflected by a clear increase in the use of nuclear (and tendentially also prenuclear) accents. Accordingly, the number of words that are not accented progressively decreases. The data reveal similar results for a separate analysis of nouns and adjectives. However, the effect of information status on nouns applies more clearly to nuclear accents, while for adjectives it primarily shows differences in the distribution of prenuclear accents (due to structural reasons).

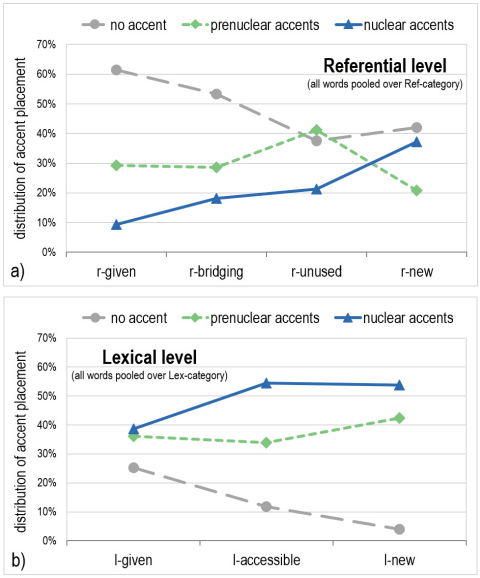

Figure 2: Relative distribution of (prenuclear and nuclear) accents on all RefLex-annotated words ordered according to their assumed level of givenness on a) a referential and b) a lexical level

In order to investigate the interaction between referential and lexical categories of givenness with regard to the prosodic marking we further analyzed the effect of combined RefLex categories.

Results for accent *placement* (see Fig.3) show that both levels have an impact on the resulting degree of prosodic prominence. Thus, if an item is both referentially and lexically *new*, there is a high probability that the item is marked by a nuclear accent (which is highly prominent in general). However, if an item is referentially *given* but lexically *new*, it is most consistently marked by a prenuclear accent, which may be considered secondary in its degree of prominence (see [17]). That is, the resulting prominence value of the prosodic realization seems to reflect the combined degree of givenness of an item, represented by the two levels of information status.

For the distribution of accent *types* (and their inherent level of prominence, see [18]) it is the *lexical* level that turns out to be decisive (see Fig.4). This is reflected by a reverse distribution of rising and high accents, i.e. rising accents (assumed to be most prominent) become more frequent, whereas high accents become less frequent with increasing newness on the lexical level.

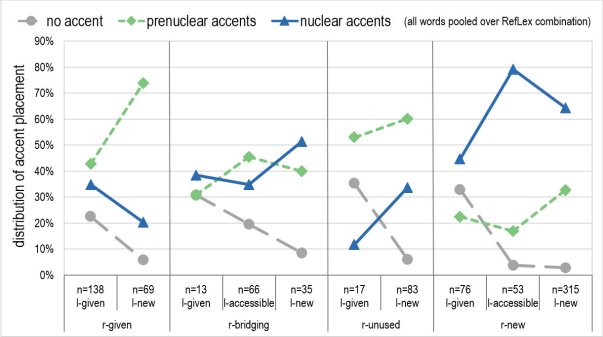

Figure 3: Relative distribution of (prenuclear and nuclear) accents on all RefLex-annotated words ordered according to their level of givenness (RefLex combination)

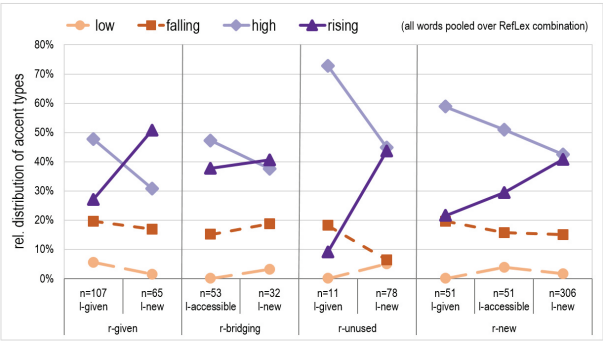

Figure 4: Relative distribution of accent types (pre-nuclear and nuclear accents pooled) on RefLex-annotated words ordered according to their level of givenness (RefLex combination)

4. Conclusions

Although we investigate (semi-)spontaneous speech, a style that has often been claimed to be less clear-cut than read speech (see e.g. [4]), our study provides clear evidence for the relation between levels of givenness and their prosodic realization.

The data suggest that accent placement is a more decisive prosodic marker of information status than accent type. Thus, the hypotheses on the distribution of accents in terms of their placement could be confirmed, showing a general tendency of a stepwise increase in prosodic prominence from given to new expressions *at both levels (referential and lexical)*. As a consequence, the *combined* degree of givenness of an item (reflected by combined RefLex labels) results in an overall prominence value of the prosodic realization.

Furthermore, the RefLex scheme has been shown to be a promising tool for the investigation of the relation between an item's information status and its prosodic realization. Our study has confirmed the relevance of both a referential and a lexical level of givenness.

Finally, the spot-the-difference task has turned out to be a useful source of task-oriented, (semi-)spontaneous speech for the examination of information status and prosody.

5. References

[1] W. Chafe, "Givenness, Contrastiveness, Definiteness, Subjects, Topics and Point of View," in *Subject and Topic*, C. Li, Ed. New York: Academic Press, 1976, pp. 25-56.

[2] E. F. Prince, "Toward a Taxonomy of Given-New Information," in *Radical Pragmatics*, P. Cole, Ed. New York: Academic Press, 1981, pp. 223-256.

[3] S. Baumann and A. Riester, "Referential and lexical givenness: Semantic, prosodic and cognitive aspects," in *Prosody and Meaning*, G. Elordieta and P. Prieto, Eds. Berlin & New York: Mouton De Gruyter, 2012, pp. 119-162.

[4] S. Baumann and A. Riester, "Coreference, Lexical Givenness and Prosody in German," in *Lingua* [Special Issue "Information Structure Triggers"], vol. 136, J. Hartmann, J. Radó, S. Winkler, Eds. 2013, pp.16-37.

[5] D. J. Allerton, "The Notion of 'Givenness' and its Relation to Presupposition and Theme," *Lingua*, vol. 44, pp.133-168, 1978.

[6] J. Terken and J. Hirschberg, "Deaccentuation of Words Representing 'Given' Information: Effects of Persistence of Grammatical Role and Surface Position," *Language and Speech*, vol. 37, pp. 125-145, 1994.

[7] C. Féry and F. Kügler, "Pitch accent scaling on given, new and focused constituents in German," *Journal of Phonetics*, vol. 36, no. 4, 680-703, 2008.

[8] J. B. Pierrehumbert and J. Hirschberg, "The Meaning of Intonational Contours in the Interpretation of Discourse," in *Intentions in Communication*, P. R. Cohen, J. Morgan and M. E. Pollack, Eds. Cambridge: MIT Press, 1990, pp. 271-311.

[9] K. Kohler, "Terminal Intonation Patterns in Single-Accent Utterances of German: Phonetics, Phonology and Semantics," *AIPUK*, vol. 25, pp. 115-185, 1991.

[10] S. Baumann, M. Grice, "The Intonation of Accessibility," *Journal of Pragmatics*, vol. 38, no. 10, pp. 1636-1657, 2006.

[11] A. Chen, E. den Os, E. and J. P. de Ruiter, "Pitch accent type matters for online processing of information status: Evidence from natural and synthetic speech," *The Linguistic Review*, vol. 24, pp. 317-344, 2007.

[12] P. Schumacher and S. Baumann, "Pitch accent type affects the N400 during referential processing," *NeuroReport*, vol. 21, no. 9, pp. 618-62, 2010.

[13] C. T. Röhr and S. Baumann, "Prosodic marking of information status in German," *Proceedings of the 5th International Conference on Speech Prosody*, vol. 100019, pp. 1-4, 2010.

[14] C. T. Röhr and S. Baumann, "Decoding information status by type and position of accent in German." *Proceedings of the ICPhS XVII*, 2011, pp. 1706-1709.

[15] A. R. Bradlow, R. E. Baker, A. Choi, M. Kim and K. J. van Engen. "The Wildcat Corpus of Native- and Foreign-Accented Englisch," *Journal of the Acoustical Society of America*, vol. 121, no. 5, pp. 2-3072, 2007.

[16] M. Grice, S. Baumann, and R. Benzmüller, "German Intonation in Autosegmental-Metrical Phonology," in *Prosodic Typology. The Phonology of Intonation and Phrasing*, S. Jun, Ed. Oxford: OUP, 2005, pp. 55-83.

[17] S. Baumann, "The Importance of Tonal Cues for Untrained Listeners in Judging Prominence," *Proceedings of the 10th ISSP*, pp. 21-24, 2014.

[18] S. Baumann and C. T. Röhr, "The perceptual prominence of pitch accent types in German," *Proceedings of the ICPhS XVIII*, vol. 298, pp. 1-5, 2015.

[19] J. G. Snodgrass and M. Vanderwart. "A standardised set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity," *Journal of Experimental Psychology: Human Learning and Memory*, vol. 6. 174-215, 1980.

[20] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 341-345, 2001.

[21] A. Schiller, S. Teufel, C. Stöckert and C. Thielen, "Vorläufige Guidelines für das Tagging deutscher Textcorpora mit STTS", 1995.