

Annotating Information Status in Spontaneous Speech

Stefan Baumann¹ & Arndt Riester²

¹IfL Phonetik, University of Cologne, Germany

²IMS, University of Stuttgart, Germany

stefan.baumann@uni-koeln.de, arndt.riester@ims.uni-stuttgart.de

Abstract

We propose a fine-grained annotation scheme for the analysis of information status in spoken language, subdivided into a referential and a lexical level. The taxonomy is easy to handle and allows for a precise and comprehensive way of capturing the informational differences between nominal expressions in texts. First results on the prosodic marking of an item's information status in a small corpus of spontaneous monologues in German confirm the relevance of the two levels of description. There are significant effects of both levels of information status on (de)accentuation in isolation and in combination. However, results show more accents than postulated by our hypotheses.

Index Terms: prosody, information status, spontaneous speech, multi-layer annotation, referential givenness, lexical givenness, German

1. Introduction

The information structure of utterances can be described and analysed at several levels, e.g. in terms of focus and background, or the information status of discourse referents. Central to our (long-term) investigation is a constituent's degree of *informativeness*, derived from the interplay between various aspects of information structure, and its marking at different levels of linguistic description. In particular, we will relate the semantic and (morpho-)syntactic correlates of informativeness, which can be gained purely from the text, to prosodic aspects of the speech signal, so as to estimate the extent of each contribution to the encoding of information structure.

As a step towards this general aim, we developed a multi-layer labelling set for the annotation of a variety of data types. In the present paper, we report on a study of spontaneous monologues in German and restrict our linguistic analysis to the level of information status and its prosodic marking. Before we present our data including relevant labels for the annotation and start discussing first results, we will give some background on the notion of information status (or givenness) in the literature and its relation to prosody.

2. Theoretical Background

2.1. Information status or degrees of givenness

Information is generally conveyed via propositions, which express properties of and predicative relations between referents. These (non-propositional) denotata of individual type constituents, which are typically expressed in argument categories such as noun phrases (determiner phrases), pronouns or prepositional phrases, can be regarded to have a certain *information status* or degree of givenness.

We are starting out from the somewhat informal cognitive approach proposed by Chafe [5]. He defines givenness in

terms of the *activation cost* a speaker has to invest in order to transfer an idea from a previous state into an active state. If a referent is already active in the listener's consciousness at the time of the utterance, it is *given*, and if a referent becomes activated from a previously inactive state, it is *new*. Chafe additionally assumes a third, semi-active state. He calls a referent that becomes activated from this semi-active state *accessible* information.

In the approach by Prince [16], at least three different sources for the cognitive activation of a referent can be distinguished. First, a referent may be identifiable because it is stored in the memory of speaker and listener, either as a unique referent which is part of a public knowledge of the world (e.g. *the sun*) or as a more idiosyncratic piece of shared knowledge between the interlocutors (e.g. *John*). In [16], referents of this type are classified as 'unused' (i.e. identifiable but discourse-new). Second, a referent can be present in the text-internal world due to previous mention. This mention may be explicit, as in *John [...] John*, in which the anaphor may be called 'textually evoked', or it may be implicit, as in the case of bridging (see [6]), e.g. *restaurant ... waiter*. The latter type of anaphor has also been called 'inferrable' [16]. Third, a referent can be present in the text-external world, i.e. visible or otherwise salient in the speech setting. We may differentiate between immediate 'situationally evoked' referents (in particular the interlocutors themselves) and less salient objects within the environment which can be referred to by means of demonstratives (accompanied by pointing).

According to Chafe and Lambrecht [13], recently mentioned and immediate situationally evoked entities are given, whereas bridged/inferrable entities and 'demonstratives' are accessible; all other (including unused) entities are therefore new. A ternary taxonomy of information status (given – accessible – new; in varying terminology and definition) with their different modes and sources has been used as the starting point for most current models or annotation schemes that deal with the classification of discourse referents (e.g. [14], [1], [7]).

2.2. Linguistic marking of information status

In West Germanic languages, the information status of referents is marked at different linguistic levels. On the one hand, there are morpho-syntactic and lexico-semantic features such as (in)definiteness, the choice of lexical form (as a full noun phrase or pronoun) and sense relations between antecedents and anaphors. On the other hand, referents are marked by prosodic means, in particular (de)accentuation.

However, the degree of an item's activation is not simply marked by the presence or absence of accent. Additionally, both accent type and strength convey relevant meaning differences. As to accent type, it has been shown that high and/or medial peak accents are used to mark hearer-new information, whereas downstepped, low and/or early peak accents mark accessible information. This rough classification holds for English [15] and German ([12]; see also [1]). For accent

strength, different kinds of secondary prominence have been proposed in several studies on English and German as markers of semi-active (accessible) information. One of them is the ‘phrase accent’, defined by Grice et al. [8] as an edge tone with a secondary association to a stressed syllable.

In perception and production experiments on the prosody of a discourse referent’s information status in German ([1], [2], [18]), accessible information was found to be no uniform category. Rather, there are different types of more or less activated information, e.g. denoting different semantic relations, which demand different accentual patterns as linguistic markers. A rough ternary distinction between high accents for new information, low accents for accessible information and no accents for given information could be observed, suggesting a somewhat iconic use of pitch height in the marking of a referent’s information status. This finding is in line with the function of intonation attributed to the *Effort Code* [10]: the higher the pitch, the newer (and more ‘informative’) the discourse referent.

2.3. Two semantic levels of givenness and their influence on prosody

Most semantic approaches to givenness define the information status of an individual type anaphor in terms of coreference with an antecedent (e.g. [19], [20], [17]). The dialogue in (1) e.g., adopted from Büring [3:4], shows that the NP *the butcher* in A is given (indicated by lack of accent), since it is interpreted as coreferential with the previously mentioned *Dr. Cremer* (capitals indicate nuclear, small capitals prenuclear accents).

- (1) Did you see Dr. Cremer to get your root canal?
 A Don’t remind me. I’d like to STRANGLE the butcher.
 B # Don’t remind me. I’d like to STRANGLE the BUTcher.

The accent pattern in (1) B is inappropriate for an answer to the preceding question. In fact, it would only be appropriate if *Dr. Cremer* and *the butcher* were not coreferential. Nevertheless, the example in (2) (from [3:4]) clearly shows that givenness does not necessarily imply coreference.

- (2) Why do you study Italian?
 A I’m MARRIED to an Italian.
 B # I’m MARRIED to an ITALIAN.

In answer (2) A, the second mention of *Italian* is treated as given, although it is not coreferential with the first mention: *Italian* in the question denotes the language, *Italian* in A denotes a person. Thus, *Italian* in the question can be thought of as a homophonous ‘pseudo-antecedent’, whose mere mention is sufficient to render *Italian* given. In other words, it is the *lexical item* which determines the prosodic marking, not the *referent*. Note that an accent on *Italian*, as in B, would be unacceptable.

Moreover, an item’s cognitive activation does neither require coreference (as in (1)) nor literal previous mention (as in (2)). This is the case with bridging relations introduced in section 2.1 above and with non-identity anaphors of a subsumed, i.e. extensionally included, word, as in (3), where the anaphoric hypernym *string instruments* (the subsuming term) can be deaccented:

- (3) Bach wrote many pieces for viola.
 A He must have LOVED string instruments.

Example (3) is adopted from van Deemter [20:7], who calls this type of activation *concept-givenness*. Note that in the reversed case the anaphoric hyponym *viola* (the subsumed term) would not be concept-given, but at least accessible, and have to be accented, as in (4) A [20:7]. Deaccentuation of the hyponym, as in (4) B, would be inappropriate.

- (4) Bach wrote many pieces for string instruments.
 A He must have LOVED the viOLA.
 B # He must have LOVED the viola.

That is, some lexical relations, namely those implying a hierarchical structure like hyponymy or meronymy, are proposed to be asymmetrical as to their likelihood of being accented (see also [2] for an investigation of different *accent types* on items in a variety of semantic relations).

The examples presented in this section indicate that both the referential and lexical level have an impact on an item’s degree of givenness and, in turn, its accentuation. A distinction between the two levels can also be found in the system of *cohesion* within the framework of Systemic-Functional Linguistics (e.g. [11]). Cohesion describes the lexico-grammatical links between elements in a discourse. While reference operates at phrase level and creates links between elements from the situation (exophoric) or from the text (endophoric), lexical cohesion operates at word level and is achieved through the choice of lexical items [11:535]. We adopt the two levels of cohesion and the different domains involved in our annotation scheme.

2.4. Hypotheses

From the theoretical and empirical studies mentioned above we derive the hypotheses for an initial investigation of our spontaneous data. As a general hypothesis (Hypothesis I) we expect that (preliminarily following the terminology of [5] and [13] here) new items are marked by high (nuclear) pitch accents, accessible items by low accents, often in pre- and post-nuclear position (the latter classified as *phrase accents*), and given items by deaccentuation. We have to be aware, however, that the prosodic context may influence the actual realisation of certain accent types.

More specific hypotheses we want to test:

- II New referents encoded by new lexical items are marked by high accents.
- III Given referents encoded by new lexical items are deaccented (*butcher* example (1)).
- IV New referents encoded by given lexical items are deaccented (*Italian* example (2)).
- V Lexical anaphors which are not coreferential are deaccented if they are superordinate to the antecedent (*viola – string instruments* example (3)).
- VI Lexical anaphors which are not coreferential are marked by low accents if they are subordinate to the antecedent (*string instruments – viola* example (4)).

3. Data and Annotation

We analysed three spontaneous monologues, which are part of a larger corpus on German currently being compiled for an investigation of various types of spoken data. The monologues annotated so far were produced by three native speakers aged between 27 and 30, whose only instruction was to tell a story of their choice for no longer than five minutes. The three stories, which were digitally recorded in a quiet room, consist of 277 intonation phrases comprising 1619 words.

The labelling at various linguistic levels was done by two independent annotators using EMU software [4]. For the present paper, we concentrate on the consensus annotation of the information status of nominal categories at the referential and lexical levels, and their phonological marking in terms of position of accents (prenuclear, nuclear, phrase accents) as well as the type of pitch accents used (following GToBI [9]).

As to information status, the **referential level** applies to the domain of the phrase, i.e. to the full projection of an NP (DP) or PP. We propose the following labels to be relevant, accounting for the degree and source of a referent's givenness, and for the distance from its last mention.

<i>Definites</i>	
r-given	anaphor corefers with antecedent in previous discourse
r-given-sit	antecedent is immediately present in text-external context (in particular discourse participants)
r-given-displaced	coreferring antecedent does not occur in last 5 intonation phrases
r-environment	refers to item in text-external context (conversational environment), pointing required
r-bridging	anaphor can be inferred from non-coreferring antecedent
r-bridging-contained	the anchor is embedded in the anaphoric phrase
r-unused-known	item which is generally known, but not derivable from previous discourse
r-unused-unknown	item which is identifiable from its own linguistic description, but not derivable from previous discourse
<i>Definites or Indefinites</i>	
r-cataphor	item whose referent is established later on in the text
r-generic	abstract or generic item
<i>Indefinites</i>	
r-new	specific indefinite introducing a new referent

The **lexical level** applies to the word domain. The proposed labels include:

l-given	recurrence of same noun
l-given-syn	relation between nouns at the same hierarchical level
l-given-super	noun is lexically superordinate to previous noun (e.g. hypernym or holonym)
l-accessible	two related nouns whose hierarchical lexical relation is unclear
l-accessible-sub	noun is lexically subordinate to previous noun (e.g. hyponym or meronym)
l-new	noun not related to another noun within last 5 intonation phrases

Note that not every referent is annotated at the lexical level, as e.g. in the case of pronouns.

4. First Results and Discussion

Hypothesis I is partly confirmed. First, a referent's information status has a significant influence on *accent position* (chi square; $p < 0.001$). The distribution of accents on new referents (in this case the labels r-new and r-generic) clearly differs from given referents, the former displaying more nuclear accents (Fig.1; significance between r-given-displaced and r-new: $p < 0.05$, between r-given and r-new: $p < 0.001$). We do not find significant differences between the categories r-bridging,

r-unused and r-new, though. All categories significantly differ from r-given items ($p < 0.001$), since the vast majority of r-given referents is deaccented. We do not observe a significant difference in the distribution of prenuclear accents and phrase accents depending on information status.

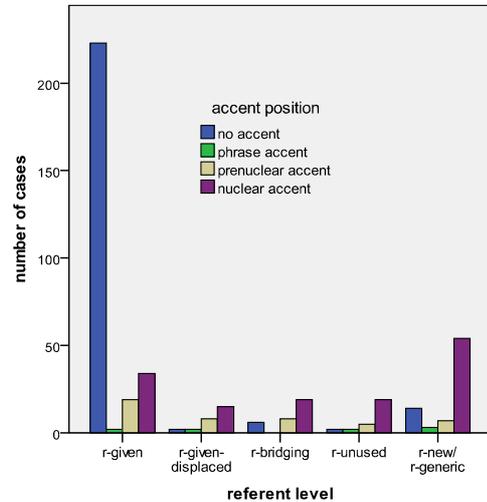


Figure 1: Distribution of accent position categories over various types of information status at the referential level.¹

Second, we do not find a significant influence of information status on *accent type*, apart from the already mentioned observation that r-given referents are mostly deaccented. Interestingly, for all accented items, L* accents appear to be the most frequent type. This can be explained by the large number of continuation rises in spontaneous monologues, realised by a combination of low nuclear accents and high boundary tones.

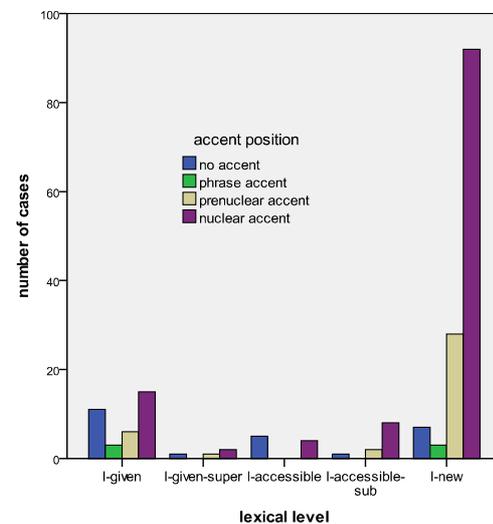


Figure 2: Distribution of accent position categories over various types of information status at the lexical level.²

As to the *lexical* level of information status, we also find significant effects on accent position and accent type ($p < 0.001$ for both). Fig.2 shows that l-new lexical items are significantly more often marked by prenuclear and nuclear pitch accents than l-accessible and l-given lexemes (difference l-new vs. all

¹ The label 'r-given' comprises the categories r-given and r-given-sit, 'r-bridging' comprises r-bridging and r-bridging-contained, and 'r-unused' r-unused-known and r-unused-unknown. The categories r-environment and r-cataphor are not listed due to their rareness.

² 'l-given-syn' is missing due to rareness of occurrence.

other categories: $p < 0.001$). Apart from the distribution of accents in words classified as l-given-super and l-accessible, the difference between all other label classes is significant. Despite the relatively large number of nuclear accents for given lexemes (partly due to contrast), there is an increase in the deaccented-accented ratio towards accentuation from l-given through l-accessible(-sub) to l-new words. As to accent type, we find that new words are almost equally often assigned H* accents (45%) and L* accents (41%). For all other lexical categories, L* is the most frequent accent type (if accented).

Hypothesis II is also partly confirmed, since new referents (e.g. r-new and r-generic) encoded by new lexical items are marked by H* accents in 41% of the cases, and even in 84% of the cases the referent receives a nuclear pitch accent.

Hypothesis III, according to which the combination of given referents and new lexical items leads to deaccentuation (*butcher* example (1)), is not confirmed. The speakers deaccent the items in question only in 2 out of 10 cases. Possible reasons for the larger number of accents are slow speech and short phrases (particularly in one speaker) as well as the occurrence of the items in predicative constructions (predications are accented in 93% of the cases). These seem to overrule deaccentuation of r-given constituents.

Our hypothesis IV postulating that new referents plus given lexical items are deaccented (*Italian* example (2)) can be partly confirmed. However, there are only 5 tokens in our data, 3 of which display deaccentuation.

The fifth hypothesis is not confirmed, since only 1 out of 5 lexically superordinate anaphors (*viola – string instruments* example (3)) lacks an accent. This outcome requires further investigation.

Finally, hypothesis VI is partly confirmed: 5 out of 10 lexically subordinate anaphors of the *string instruments–viola* type are marked by a L* accent. Fig.3 gives an example from our spontaneous data.

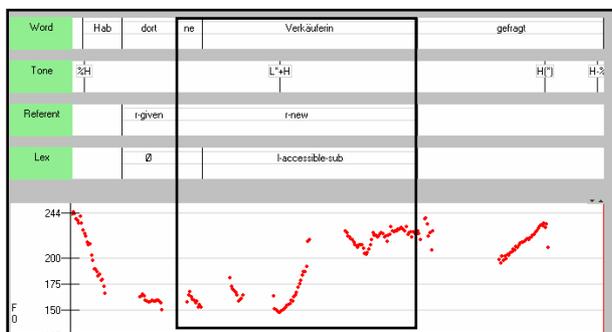


Figure 3: EMU screen shot of the utterance *Hab dort ne Verkäuferin gefragt* ('asked a shop assistant there'), with F0 contour and label tiers for words, GToBI tones and two levels of information status.

The context of the utterance is the sentence *bin in'n Laden reingegangen* ('went into a shop'), rendering the non-coreferential item *Verkäuferin* ('shop assistant') accessible information, since *Verkäuferin* is a meronym of (and thus subordinate to) *Laden* ('shop'). Nevertheless, *Verkäuferin* is a new (indefinite) referent and thus informative. This combination of new and activated information seems to be appropriately encoded by L*-type pitch accents (here: L*+H).

5. Conclusions

The tentative analysis of a small corpus of spontaneous data confirms the importance of both a referential and a lexical level for an investigation of information status in spoken lan-

guage. It could be shown that *given referents* are mostly deaccented, while *new lexical items* are mostly accented. However, we find more accents than postulated by our hypotheses. There are no clear results in terms of accent type, probably due to the kind of data investigated. That is, the large number of low accents stems from continuation rises which are characteristic of monologues. More detailed findings about the proposed categories will be gained from further analyses of the interplay between both levels of information status in combination with other layers of linguistic description.

6. References

- [1] Baumann, S., 2006. The Intonation of Givenness - Evidence from German. *Linguistische Arbeiten* 508. Tübingen: Niemeyer.
- [2] Baumann, S. and Grice, M., 2006. The Intonation of Accessibility. *Journal of Pragmatics* 38 (10), 1636-1657.
- [3] Büring, D., 2007. Intonation, Semantics and Information Structure. In: G. Ramchand & C. Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*.
- [4] Cassidy, S. and Harington, J., 2001. Multi-Level Annotation in the EMU Speech Database Management System. *Speech Communication* 33 (1-2), 61-78.
- [5] Chafe, W., 1994. *Discourse, Consciousness, and Time*. Chicago/London: University of Chicago Press.
- [6] Clark, H., 1977. Bridging. In: P.N. Johnson-Laird & P.C. Wason (eds.), *Thinking: Readings in Cognitive Science*. Cambridge: Cambridge University Press. 411-420.
- [7] Dipper, S., Götze, M. and Skopeteas, S. (eds.), 2007. *Information Structure in Cross-Linguistic Corpora: Annotation Guidelines for Phonology, Morphology, Syntax, Semantics, and Information Structure*. ISIS 7, Universität Potsdam.
- [8] Grice, M., Ladd, D.R. and Arvaniti, A., 2000. On the Place of Phrase Accents in Intonational Phonology. *Phonology* 17 (2), 143-185.
- [9] Grice, M., Baumann, S. and Benz Müller, R., 2005. German Intonation in Autosegmental-Metrical Phonology. In: Jun, S. (ed.), *Prosodic Typology. The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press. 55-83.
- [10] Gussenhoven, C., 2002. Intonation and Interpretation: Phonetics and Phonology. *Proceedings 1st Int. Conference on Speech Prosody*. Aix-en-Provence. 47-57.
- [11] Halliday, M.A.K. and Matthiessen, C., 2004. *An Introduction to Functional Grammar*. 3rd, revised edition. London: Edward Arnold.
- [12] Kohler, K., 1991. Terminal Intonation Patterns in Single-Accent Utterances of German: Phonetics, Phonology and Semantics. *AIPUK* 25. 115-185.
- [13] Lambrecht, K., 1994. *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.
- [14] Nissim M., Dingare, S., Carletta, J. and Steedman, M., 2004. An Annotation Scheme for Information Status in Dialogue. In *Proceedings of the 4th LREC*, Lisbon.
- [15] Pierrehumbert, J. and Hirschberg, J., 1990. The Meaning of Intonational Contours in the Interpretation of Discourse. In: Cohen, Philip R., Jerry Morgan & Martha E. Pollack, (eds.), *Intentions in Communication*. Cambridge: MIT Press. 271-311.
- [16] Prince, E., 1981. Toward a Taxonomy of Given-New Information. In: Peter Cole (ed.), *Radical Pragmatics*, New York: Academic Press. 223-256.
- [17] Riester, A., 2008. The Components of Focus and their Use in Annotating Information Structure. *AIMS* 14 (2), Universität Stuttgart.
- [18] Röhr, C. and Baumann, S., submitted to *Speech Prosody 2010*. Prosodic Marking of Information Status in German.
- [19] Schwarzschild, R., 1999. GIVENness, AvoidF and Other Constraints on the Placement of Accent. In: *Natural Language Semantics* 7, 141-177.
- [20] van Deemter, K., 1999. Contrastive Stress, Contrariety, and Focus. In: P. Bosch & R. van der Sandt (eds.), *Focus - Linguistic, Cognitive, and Computational Perspectives (Studies in Natural Language Processing)*. Cambridge: CUP, 3-17.