Information status marking in spontaneous vs. read speech in story-telling tasks – Evidence from intonation analysis using GToBI

CrossMark

Laura E. de Ruiter [a,b,*]

[a] Faculty of Linguistics and Literary Studies, Bielefeld University, 33615 Bielefeld, Germany
[b] Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands

ARTICLE INFO

ABSTRACT

Two studies investigated whether speaking mode influences the way German speakers mark the information status of discourse referents in nuclear position. In Study 1, speakers produced narrations spontaneously on the basis of picture stories in which the information status of referents (new, accessible and given) was systematically varied. In Study 2, speakers saw the same pictures, but this time accompanied by text to be read out. Clear differences were found depending on speaking mode: In spontaneous speech, speakers always accented new referents. They did not use different pitch accent types to differentiate between new and accessible referents, nor did they always deaccent given referents. In addition, speakers often made use of low pitch accents in combination with high boundary tones to indicate continuity. In contrast to this, read speech was characterized by low boundary tones, consistent deaccentuation of given referents and the use of H+L* and H+!H* accents, for both new and accessible referents. The results are discussed in terms of the function of intonational features in communication. It is argued that reading intonation is not comparable to intonation in spontaneous speech, and that this has important consequences also for our choice of methodology in child language acquisition research.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

In this article, I will report on an experiment about intonational differences between read speech and spontaneous speech in German that was triggered by a mismatch of what was reported in the literature and my own findings. As a child language researcher I was primarily interested in how German children use intonation to signal the information status of discourse referents, as research suggested that this is a rather complex mapping to acquire. Leaving aside the findings regarding the children, which are given in detail in De Ruiter (2014), an unexpected outcome was that the adult control group in that study did not do what would be expected on the basis of previous research. This led me to investigate the influence of speaking mode on intonation and consider the consequences this may have for language acquisition research. The goal of this article is thus two-fold: First, to present my findings as an empirical contribution to the topic of prosodic markings of "givenness", and in particular the role speaking mode plays in those markings. Second, to heighten the awareness of researchers of the impact the choice of methodology can have for the study of language acquisition.

In what follows, I will first give a brief account on the intonation model used here, give some background on the proposed relation between the information status of discourse referents and intonation, and summarize the state of knowledge regarding the prosodic differences between read and spontaneous speech. Then I will report the adult results of the acquisition study mentioned above (De Ruiter, 2014), which used a picture story-telling task (Study 1). In the next section, I will present the methodology and results of a second study that used the same picture stories, but this time accompanied by text (Study 2). In the General discussion, the differences between the two sets of results and potential reasons for this will be discussed. I will relate the findings to current issues in oral reading research, which support the assumption that reading intonation is an acquired cultural skill that does not necessarily reflect intonation patterns in natural speech. Finally, I will consider the implications for language acquisition studies on intonation.

* Correspondence address: Faculty of Linguistics and Literary Studies, Bielefeld University, 33615 Bielefeld, Germany. Tel.: +49 521 106 67194.
E-mail address: laura.deruiter@uni-bielefeld.de

1.1. Intonation

Intonation will be discussed within what is now known as the "autosegmental-metrical" (AM, a term coined by Ladd, 1996) theory of intonation (Pierrehumbert, 1980), and annotated using the German Tones and Break Indices (GToBI) system (Grice & Baumann, 2002). In the AM-framework, intonation patterns are described as sequences of high (H) and low tones (L). These tones are aligned with prominent (lexically stressed) syllables in the text and the edges of phrases. Tones that are associated with prominent syllables are called pitch accents, which can consist of one or two tones (connected with a "+" sign). The diacritic "*" indicates which tone bears the association with the lexically stressed syllable. Pitch accents can be prenuclear or nuclear. Intonational phrases contain at least one nuclear accent. Nuclear accents are the strongest accents in a phrase and occur towards the end of that phrase. All accents preceding the nuclear accent are prenuclear accents. Tones associated with phrase edges are referred to as boundary tones, symbolized with the "%" sign. Both pitch accents and boundary tones can be affected by two kinds of modifications: If a high tone (H) is considerably lower than a preceding high tone (without being a L*), it is said to be *downstepped*. If it is considerably higher in relation to the preceding high tone, it is said to be *upstepped*. Downstep is marked in transcriptions by inserting a "!" before the affected tone; upstep is indicated by a "^".

Some models (including the model of German intonation adopted here) posit the existence of a third type of tone – the phrase accent (e.g., Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980). It is marked with a "-", and always monotonal (i.e., H- or L-). Several functions have been attributed to the phrase accent: On the one hand it is said to control the pitch contour between the last pitch accent in the phrase and the boundary tone at the end of the phrase (Benzmüller & Grice, 1997: 9). On the other hand, it has also been assumed to be a demarcation marker of intermediate phrases (ip) (Beckman & Pierrehumbert, 1986). Thirdly, phrase accents have been taken to be instantiations of post-nuclear prominence (i.e., more comparable to pitch accents). Grice, Ladd, and Arvaniti (2000) have defined phrase accents as "edge tones with a secondary association to an ordinary tone-bearing unit" (p. 180). Where a phrase accent and a following boundary tone have the same pitch level (i.e., both H or both L), the German transcription system GToBI uses a simplified notation with only one tone (e.g., H-% instead of H-H%).

Pierrehumbert and Hirschberg (1990) assume that pitch accents "convey information about the status of the individual discourse referents" (p. 286), whereas the type of boundary tone indicates whether the current utterance should be interpreted together with the following utterance (high boundary tones), or not (low boundary tones; Pierrehumbert & Hirschberg, 1990: 305).

One comment is in order regarding two pitch accent types in the GToBI inventory, because they play a prominent role in the literature on information status marking in German, which will be discussed in the next sub-section: H+L* and H+!H*. Although two distinct accents are claimed in the GToBI literature (Grice & Baumann, 2002), production and perception experiments have up to this point failed to find support for a categorical distinction between these two (Grice, Baumann, & Jagdfeld, 2009; Rathcke & Harrington, 2010), and it has been suggested that both accents may actually be variants of one and the same type. There is still a debate as to whether this single category should be modelled as H+L* or as H+!H*. Even before the inception of GToBI, Kohler (1991) proposed an accent category "early peak", which would cover both accents. The notion of early peak accents has been taken up again in recent publications (e.g., Baumann & Riester, 2013; Röhr & Baumann, 2010) and will also be used in this article.

1.2. Information status and intonation

For West-Germanic languages like English, Dutch or German, it is commonly assumed that referents that are *new* to the discourse are typically pitch-accented (e.g., Chafe, 1974; Terken & Nooteboom, 1987). In contrast, *given* referents have been said to be realized without an accent, or to be "deaccented" (Ladd, 1980). However, it has also been observed that given referents tend to be accented when they are re-introduced into the discourse at a later point (Brown, 1983; Hirschberg, 1993; Terken, 1984). Other scholars argue that viewing the given/new distinction as a dichotomy is too simplistic (e.g., Ariel, 1990; Chafe, 1974; Gundel, Hedberg, & Zacharski, 1993). Chafe (1994) assumes that there is at least a third category between given and new, which he calls *accessible*. The precise nature of the intonational marking of accessible referents, however, is not agreed upon. Chafe hypothesizes that accessible information is treated in the same way as new information (Chafe, 1994, p. 75). Lambrecht (1994), on the other hand, assumes that speakers can choose whether to accent an accessible referent or not, depending on several other discourse factors. More recently, it has been suggested that the type of accent also plays a role in signalling a referent's information status, and that there may even be a distinct accent which is used to mark accessible referents. According to Pierrehumbert and Hirschberg (1990), H* pitch accents signal that a referent should be treated as new to the discourse ("added to the mutual belief space", p. 290), whereas L* accents convey that the referent is already part of the listener's discourse model. When the hearer should be able to access a referent either because it is already in the mutual belief space of the conversational partners or because it can somehow be inferred via pragmatic reasoning, a H+!H* accent (H+L* accent in Pierrehumbert and Hirschberg's notation) or a H+L* accent is used (p. 297ff.). For German, Baumann and Grice (2006) have proposed that the accessible status of a referent is signalled with the accent H+L*. New referents are said to be marked – as in English – with a H* accent, and given referents with deaccentuation.

What is the empirical evidence for these information status-pitch accent type mappings? Baumann and Grice' (2006) hypothesis about H+L* as a marker for accessibility was based on the results of a corpus study (Baumann, 2006) and an offline perception experiment (Baumann & Grice, 2006). The corpus used, the MULI-corpus (Baumann et al., 2004; Baumann et al., 2004), consisted of 22 short texts (250 sentences altogether), which were taken from the economics section of the German newspaper *Frankfurter Rundschau* and which were read by a single female speaker of standard German. Baumann analysed the information status

(new, given, or accessible) and the type of pitch accent of the last referring expression in each sentence, provided this last referring expression was a full noun phrase. An example is given in (1). Capital letters indicate the position of the nuclear accent.

(1) *Die vor einem Jahr eröffnete Fabrik geht gerade*
the before one year opened factory goes presently
zum ZWEIschicht-Betrieb über.
to the TWO-shift operating system over
status: New
accent: H*

'The factory that opened one year ago is presently changing to TWO-shift operation.'

Only low-ending utterances (i.e., those ending with a L-% boundary tone) were used, which left 106 sentences (and consequently referents) for the analysis. The analysis showed that more than 55% of the accessible referents were realized with a H+L* accent, suggesting that it is a suitable marker for accessible referents. In contrast, the accent H*, which is typically assumed to signal newness, was used in only 2 out of 61 cases. This suggested that H* is not a prototypical accent for accessible referents. Baumann and Grice (2006) found further evidence for the role of H+L* as a marker of accessibility using an offline perception experiment: Participants had to rate the appropriateness of different intonation contours in varying contexts. In the experiment, subjects read (silently) and simultaneously listened to 40 short paragraphs in which the target referents were embedded. The texts were intended to render the target referents accessible by establishing different relations between an antecedent and an anaphor (i.e., an expression that refers back – directly or via inference, as explained above – to an already established concept). These relations included repeated mention of the same word (textual displacement), but also other relations like synonymy (e.g., settee – sofa) and meronymy (e.g., finger – hand). Baumann and Grice assume (following Chafe, 1987) that in the case of textually displaced referents, a referent becomes demoted from being "new" (when it is first mentioned) to accessible "due to interference from other competing referents mentioned in the intervening discourse" (Baumann & Grice, 2006: 1643). In other cases like synonymy or meronymy, the referent has become "activated" via conceptual inference, so that for instance the mention of a part of a concept (finger) implies the entire concept as well (hand). An example of a textually displaced referent used in the experiment is provided in (2) below. The first mention of the referent (house) and its second mention three clauses later are underlined.

(2) *Die Wanderer kamen an einem alten <u>Haus</u> vorbei. Sie waren schon ziemlich erschöpft vom langen Weg, den sie zurückgelegt hatten. Daher sehnten sie sich nach einer kleinen Pause. Die Wanderer näherten sich dem <u>Haus</u>. Einer von ihnen klopfte an die Tür.*

'The hikers passed an old <u>house</u>. They were already quite exhausted from the long way they had come. Thus they were longing for a short break. The hikers came near the <u>house</u>. One of them knocked on the door.'

(From Baumann and Grice 2006: 1643)

Listeners were presented with three different versions of the target sentence, that is, the sentence in which the referent was mentioned again (accessible). In one version, the referent (e.g., house) was realized with a H* accent. In another version, it was realized with deaccentuation. In the third version, the referent was realized with a H+L* accent. There was always a sentence following the target sentence in each text in order to avoid a paragraph finality effect on the target word. Results showed that the accent H+L* was always preferred over H*, and in many cases also over deaccentuation. Based on the results of the corpus study (Baumann, 2006) and the perception study (Baumann & Grice, 2006), Baumann and Grice propose the following mapping between information status and accentuation type for German:

Röhr and Baumann (2010) conducted a reading experiment in German, in which participants read short passages that contained referents with different information statuses (new, given, textually accessible and inferentially accessible). Analysing the intonation of 1080 items, they found that new referents were mostly marked with H* and L+H* accents. As referents became 'more given', speakers were more likely to use L* and early peak (i.e., H+L* and H+!H*) accents, if they accented referents at all. It was, however, not the case that early peak accents were the most frequent realization for accessible referents; both deaccentuation and H* accents were used more often than early peak accents. Furthermore, speakers used early peak accents also for new referents in 20% of the cases. Thus while the results of this reading experiment lend general support to the hypothesis that there is a relationship between information status and pitch accent type, they do not exactly mirror the mapping in Fig. 1.

Using the read-out material from Röhr and Baumann (2010), Röhr and Baumann (2011) asked participants to listen to single words only (denoting referents) that had been cut out from the materials, and to rate the perceived givenness of the referent solely on the basis of the acoustics of that referent. The results showed "a stepwise decrease in the degree of perceived givenness from deaccentuation and prenuclear accents through low and early peak nuclear accents to high and rising nuclear accents" (p. 1706).

In summary: There is considerable evidence both from production and perception experiments that there is a relationship in German between the information status of a discourse referent and the pitch accent type used in the way proposed by Baumann (2006) and Baumann and Grice (2006), although there is some uncertainty surrounding the status of early peak accents.
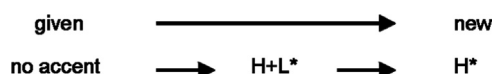

Fig. 1. Proposed mapping between information status and accentuation type in German (adapted from Baumann & Grice 2006: 1655).

1.3. Differences between read and spontaneous speech

Read and spontaneous speech differ from each other in several aspects. In cases in which read speech is elicited from standard written materials (i.e., not transcriptions of previously spontaneously uttered speech), it is usually syntactically more complex and contains more abstract lexical items (Biber, 1993). Prosodically, read speech is characterized by a lower articulation rate, decreased segmental reduction, fewer hesitations and a different pause structure (see e.g., Barik, 1977; Blaauw, 1994, 1995; Howell & Kadi-Hanifi, 1991; Laan, 1997; Levin, Schaffer, & Snow, 1982; Silverman, Blaauw, Spitz, & Pitrelli, 1992). In terms of intonation, read speech has been found to have a wider pitch range (e.g., Eskénazi, 1992), steeper slopes in $F0$ declination (Swerts & Heldner, 1996) and to contain fewer rising movements than spontaneous speech (Blaauw, 1995; Silverman et al., 1992).

None of these studies were concerned with aspects of information status marking, so it is unclear if speaking mode has an influence here. However, since the studies referred to above (Baumann & Grice, 2006; Pierrehumbert & Hirschberg, 1990; Röhr & Baumann, 2010) did not qualify their findings to be applicable to only read speech, it is here being assumed that the proposed markings should hold irrespective of speaking mode.

1.4. Hypotheses

The findings cited in Section 1.2 motivated the original investigation into the way German children use intonation to mark the information status of discourse referents (De Ruiter, 2014). However, the remainder of this article is concerned not with the child data, but with the data collected from an adult control group. In order to get a completely comparable control data set, an adult group performed the same task as the children: telling a story based on pictures. Based on the literature review presented above, the following results were expected:

- Hypothesis 1: New referents (in nuclear position) are marked with H* accents.
- Hypothesis 2: Textually accessible referents (in nuclear position) are marked with H+L* and H+!H* accents.
- Hypothesis 3: Given referents (in nuclear position) are deaccented.

2. Study 1: Marking information status in spontaneous speech

This study has been reported in De Ruiter (2014), but the details are provided here for the reader's convenience.

2.1. Methods

2.1.1. Materials

Four coloured picture stories were constructed to elicit the production of four different referents with three distinct information statuses each: new, given, and accessible. In the *new* condition, the target referent occurred for the first time. In the *given* condition, a referent was first introduced, and occurred again as target referent in the immediately following picture. In the *accessible* condition, a referent was introduced at the beginning of the story, and re-occurred only after four (in one story) or five (in three stories) other intervening pictures, respectively. Between the two occurrences of the target referent, two other referents, one animate and one inanimate, played a part in the plot of the story to ensure that the target referents' activation was reduced. Note that the manipulation is thus restricted to textually accessible referents, and does not include other inferentially accessible referents that have been investigated by Baumann and Grice (2006) and Röhr and Baumann (2010).

In the original studies on German that inspired the present investigation (Baumann, 2006; Baumann & Grice, 2006) as well as in Röhr and Baumann (2010), the target referents were always the last referent in the sentence. In order to make the results of the present study comparable with those studies, the stimuli were therefore also created in such a way as to make it most likely that referents were produced in sentence-final position, but not always paragraph-final position. In order to prevent participants from anticipating the end of a story (the stories were presented picture-by-picture, see Section 2.1.3) and producing discourse segment demarcation cues such as creaky voice (Hirschberg, Litman, Pierrehumbert, & Ward, 1987; Lehiste, 1979), the number of pictures in a story varied from five to eight. The target pictures showed simple transitive events (e.g., hugging, drawing), in which the target referents had non-agentive semantic roles (e.g., recipient, beneficiary). All four target referents were animals. The words (the names of the target referents) were disyllabic and contained mostly sonorant segments to facilitate pitch tracking: *Kamel* (/kaˈmeːl/ 'camel'), *Möwe* (/ˈmøːvə/ 'seagull'), *Biber* (/ˈbiːbɐ/ 'beaver') and *Biene* (/ˈbiːnə/ 'bee'). Note that Kamel – unlike the other three words – has a iambic stress pattern.[1]

The new and the accessible condition were elicited with the same story. This means that the introduction of the referent served as the target referent for the new condition and its re-introduction as target referent for the accessible condition. (Note that in order to obtain approximately the same number of items per condition, the referents that were initially introduced in the given condition were not additionally analysed as new referents.) An illustration of a new/accessible story is provided in Fig. 2.

---

[1] The choice of target words was determined by a pre-test with several animal names done with children for the study in De Ruiter (2014). These target words were best recognized by the children of the youngest age group. Due to this, the set is not balanced in terms of stress pattern.
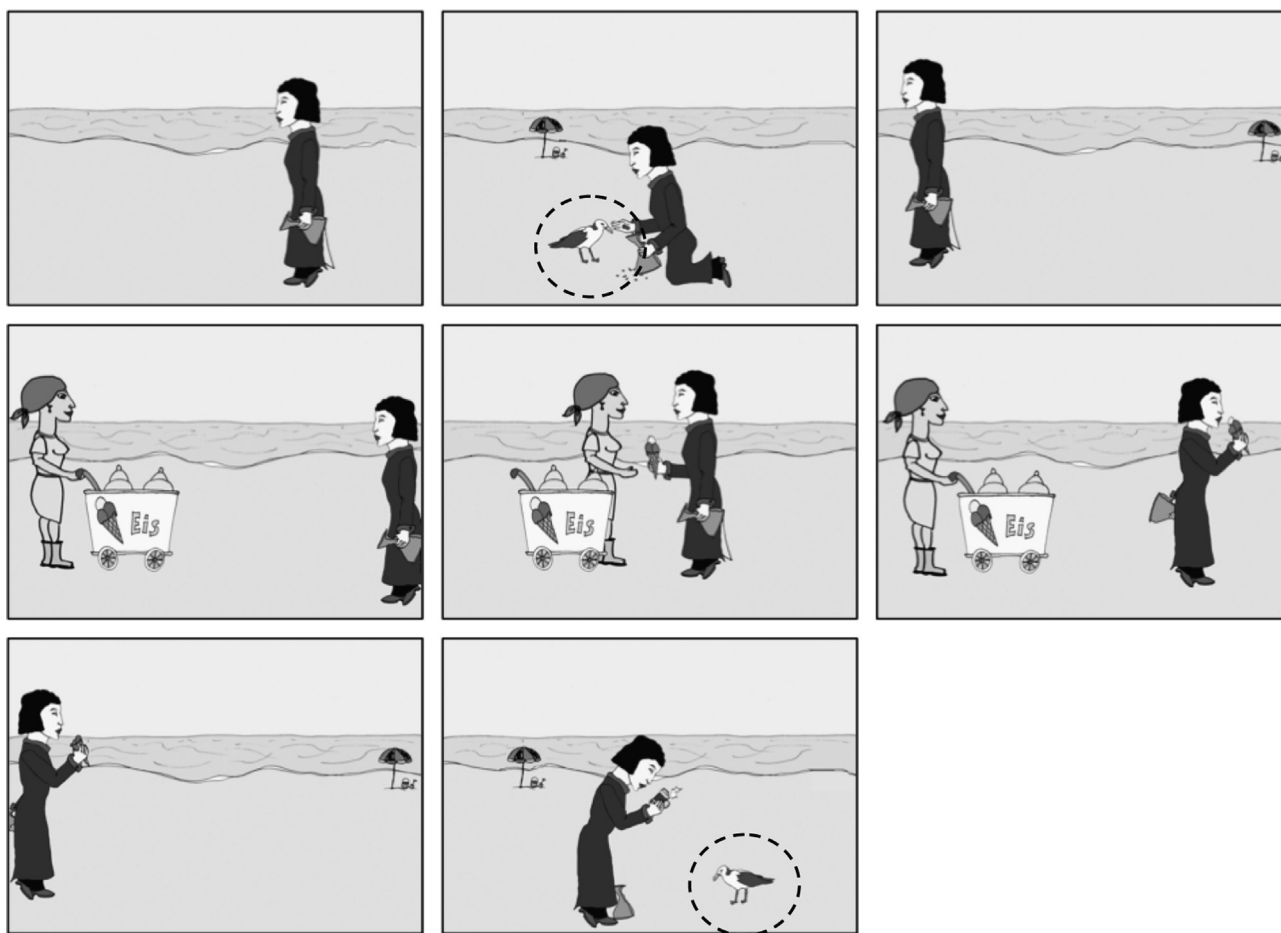
Fig. 2. Example of a picture story (order from top left to bottom right) for eliciting a referent with new and with accessible status. The target referent is the seagull (encircled) introduced in picture 2 (new condition), which occurs again in picture 8 (accessible condition).
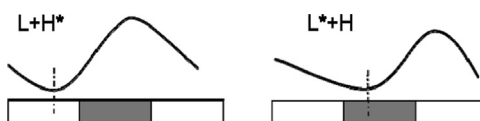

Fig. 3. Schematic representation of the location of the $F0$ minima in the accents L+H* and L*+H. The grey-shaded area indicates the lexically stressed syllable.

Referents in the given condition were presented in a separate story. Four additional stories (three pictures each) from another experiment served as fillers. In total, each participant told twelve stories. Four different lists were constructed by varying the order in which the stories appeared. Stories containing the same target referent (e.g., the new/accessible story containing *Kamel* and the given story containing *Kamel*) were always separated from each other by three other stories. Each time the story with the new and the accessible referent preceded the story with the given referent. Participants were randomly assigned to one of the four lists Fig. 3.

2.1.2. Speakers

Twenty-eight adult speakers (mean age 23 years, 5 males, 23 females) took part in the experiment. Most of them were undergraduate students of Psychology or Linguistics at the University of Potsdam. All participants were native speakers of German. They had no known speech or hearing impairments and no knowledge of the research purpose. The participants received course credit or a financial reimbursement.

2.1.3. Procedure

The participants sat in an anechoic chamber in the laboratory of the Linguistics Department of the University of Potsdam. The microphone was placed at about 25 cm distance from the speaker's mouth. The narrations were recorded using an ECM and a DAT recorder. The signals were digitized directly into the computer at a sampling rate of 44.1 kHz (16 bit format). The picture stories were presented one picture at a time on a computer screen using Microsoft PowerPoint. The screen was placed outside the recording booth and visible to the participants through a glass window in the recording booth.

The participants were instructed that they were going to see a number of short picture stories, which they should tell in such a way that another person listening to the recording would be able to retell the story without seeing the pictures. It was explained that the

stories were simple because they had originally been constructed for children. In order to make the situation more natural, the experimenter acted as listener, giving minimal feedback such as "uh huh" and asking a broad-focus question before the target picture came up ("And what happens then?"). This was done in order to make the results comparable to the experimental conditions with the children, where interaction and encouragement was needed. A warm-up story similar to the experimental items preceded the first experimental story. Each recording session lasted about 15–20 min.

2.1.4. Item selection

Two recordings from adult subjects had to be excluded because of technical problems. All other recordings were fully transcribed using ELAN, version 3.4.0 (Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006). Speakers did not always produce the intended target referent. They sometimes used different words (e.g., *Vogel* 'bird' instead of *Möwe* 'seagull') or pronominal forms. The speakers produced the target words in 84.5% of the cases (265 target words, 31 non-target words). Only target words were considered for the analysis.[2] The criteria for exclusions of items were strict in order to ensure maximal comparability both across items and with the studies reported in the Introduction. Items were excluded if they

- were realized as pronominal forms (i.e., personal and demonstrative pronouns instead of full NPs);
- were realized with a definite article in the case of new referents, or with an indefinite article in the case of accessible and given referents;
- were produced with creaky/breathy voice or laughter;
- contained disfluencies (restarts, filled pauses like *ähm*, 'uhm', etc.)[3];
- contained self-corrections of the target word (e.g., *eine Wespe, nee, Biene*, 'a wasp, no, bee');
- were referents in the new condition that had been realized with a possessive pronoun (e.g., *Und jetzt setzt er sich auf sein Kamel*, 'and now he is sitting on his camel'), because the use of the possessive pronoun suggests that the referent was considered accessible by the speaker (see Ariel, 1988).

In two cases speakers mentioned the target referent in the accessible condition before they were actually presented with the target picture (e.g.*, Er geht wahrscheinlich zurück zu seinem Kamel*, 'He is probably going back to his camel'). These were not included in the analysis, as in this case the activation of the target referent's representation was arguably higher than intended by the design. As mentioned in the Section 4.1.1, the stimuli were designed to elicit referents in phrase-final position. In about half of all utterances the speakers produced the word denoting the target referent indeed phrase-finally. In these cases, the target word was syntactically the final element in the sentence, and the end of the phrase was marked by a clearly perceptible pause. However, in the other half of the utterances, the target referent was followed by one or several words. These utterances were included in the analysis, if the target word was followed by either a verb particle (e.g., *aus* 'out', *mit* 'with') or a non-finite verb form (e.g., *gemalt* 'painted', *machen* 'to make'). In these cases, the target referent was still the last referring expression in the sentence, a criterion that was also used by Baumann (2006) for his analysis of the *Frankfurter Rundschau* corpus. This structure also matches the sentence structure used by Röhr and Baumann (2010). The final data set used in the present study contained 120 items.[4]

2.1.5. Prosodic annotation

The utterances were annotated and analysed using the speech-analysis tool Praat (Boersma & Weenink, 2013). Intonation of the target words, of the word following the target word (if there was one), and of the subsequent phrase boundary was labelled by a native speaker of German (the author) following the GToBI guidelines (Grice & Baumann, 2002; Grice, Baumann, & Benzmüller, 2005). Note that owing to the selection (see Section 2.1.4) only nuclear accents were analysed. Nuclear and prenuclear accents differ in their phonological status. Nuclear accents are typically easier to perceive and to categorize, while prenuclear accents are generally less prominent. What is more, prenuclear accents are often placed by speaker for purely rhythmic reasons (e.g., Féry & Kügler, 2008), so they may play a lesser role in the signalling of pragmatic aspects such as information status marking. As will become clear later, the speakers produced more accentuation types than the ones suggested, so that the entire GToBI set of pitch accents and boundary tones was required for annotation.

The phonological annotation of intonation is, however, not a trivial task. It is well known that even trained labellers experience difficulties in agreeing on the same labels for a given accent. This is shown by intertranscriber consistency studies that have been conducted both with the original English ToBI system (Pitrelli, Beckman, & Hirschberg, 1994; see also Syrdal & McGory, 2000) and the German GToBI system (Grice, Reyelt, Benzmüller, Mayer, & Batliner, 1996). While the agreement on the presence or absence of a pitch accent was reasonable (81% in the English study, 87% in the German study), agreement on the type of pitch accent was much lower (64% in the ToBI experiment and only 51% in the GToBI experiment). Considerable inconsistency is also reported by Braun (2006), who asked trained labellers to annotate topical referents in German sentences choosing only among the three accents H*, L+H*, and L*+H.

---

[2] Note that this differs from De Ruiter (2014) where also non-target words such as *Wespe* 'wasp' were included, because especially the young children in that experiment often did not produce the intended word.

[3] If there was a hesitation (silent pause) which did not have a disruptive effect on the $F0$ contour if the pause was cut out, the item was not excluded, following a convention that has been used in the Dutch ToDI System (Gussenhoven, 2005).

[4] Note that including items followed by verb particles means that these items differ from the other (phrase-final) items in that the boundary tone is realized on the verb particle or on the infinite verb rather than on the target word (as in phrase-final items). This should, however, not affect the accent type on the target word.

Table 1

Estimates for the position of the *F*0 minimum, lower and upper bounds and probability levels for L+H* and L*+H accents followed by a high boundary tone. *N*=34.

	Estimate	Lower bound	Upper bound	*p*
Intercept (L+H*)	1.589	−7.57	7.504	.9428
L*+H	27.461	19.77	37.455	.0001

Given that further analyses depend on the accuracy of the labels, it is crucial to obtain additional support for the correct application of the GToBI system. In the present study, an instrumental method was used to check the labelling. In the labelling guidelines for GToBI (Grice & Baumann, 2002), a number of distinguishing features are described to help identify the accent types. For instance, both the accents L+H* and L*+H are rising accents. They are said to differ, however, in the alignment of the tones with respect to the syllable structure. While in L+H* accents, the low tone (*F*0 minimum) is located before or at the beginning of the stressed syllable, and the rise occurs typically within the stressed syllable, and the high tone (*F*0 maximum) is reached in the stressed syllable. In L*+H, the low tone is supposed to be clearly within the stressed syllable. The rise begins only very late in the syllable, peaking only on the post-stressed syllable or even later.

From the defining features of the different pitch accent types given in the GToBI literature testable predictions can be derived regarding phonetic differences between certain pairs of accents, in particular those that are prone to confusion, as evidenced in intratranscriber consistency studies (Grice et al., 1996; Pitrelli et al., 1994). Subsequently it can be tested whether the accents that had been assigned different accent labels (e.g., L+H* and L*+H) do differ from each other in the predicted way (e.g., in the position of the *F*0 minimum with respect to the lexically stressed syllable). If this is the case, this can be seen as evidence that the accents have been labelled consistently.

This procedure of systematic comparisons of the pitch accent types along a number of phonetic parameters is described in detail in De Ruiter (2009). This study used the same corpus of narrations described above, but analysed the intonation of discourse referents in the German "Vorfeld". Below I will give an example of the analysis of one accent pair, L+H* and L*+H. There were 34 items labelled L+H* or L*+H, respectively. The details of the analyses of the other accentuation types can be found in the Supplementary material. The intonational phrase (IP) containing the target word was first segmented at the level of the syllable using visual information from a wide-band spectrogram. The position and value of local *F*0 maxima (max) and minima (min) were then determined manually with a domain that consisted of the lexically stressed syllable, the preceding syllable and the syllable following it. From these *F*0 landmarks, several derived measurements were made, including the relative position of the *F*0 maximum and the relative position of the *F*0 minimum. In both cases the position of the *F*0 extreme is expressed as percentage (%) into the stressed syllable. This means that if the stressed syllable had for instance a duration of 200 ms and the *F*0 maximum occurred 150 ms after the syllable onset, the relative position of the *F*0 would be said to be at 75%. In the case of the accent pair L+H* and L*+H, these measurements were used to test the prediction that the min occurs later in L*+H accents. The prediction was tested using multilevel regression modelling (the lme4 package) in R (R Development Core Team, 2008). I tested whether the pitch accent type is a significant predictor for the position of *F*0 min. In addition, subject (participant) and word (i.e., the produced item) were included as random factors, but excluded if the variance explained by them was effectively zero (smaller than .00001). If the explained variance was not zero, it was tested whether the random effect parameters are justified in the model using likelihood ratio tests (Baayen, 2008: 275ff.). The reported p-values were obtained by estimating the posterior probability of a Markov Chain Monte Carlo (MCMC) simulation with 10 000 runs.

The analysis showed that *F*0 minima in L+H* accents were aligned 27.5% earlier than in L*+H accents (1.5% into the accented syllable vs. 29.05%, *p*<.001). As can be seen in Table 1,[5] the ranges in which the minima in the two accent types occurred were non-overlapping.

Thus, while the *F*0 min in L+H* accents occurs either in the pre-accented syllable or at the very beginning of the accented syllable, it occurs in L*+H accents towards the middle of the accented syllable. This and the results of the other analyses (see Supplementary material) confirmed that the phonetic properties of the accent types differed significantly from each other in the way predicted by the GToBI literature.[6] The labels that had been assigned by the labeller (see Section 2.1.5) were therefore subsequently used to look at the distribution of pitch accent types across the three conditions, the information statuses new, accessible and given.

2.2. Results

Speakers used a variety of intonation patterns, far more than the three patterns that would be expected on the basis of previous research. Recall that the suggested mapping between information status and pitch accent type is that new referents are signalled with a H* accent, accessible referents are marked with the accent H+L*, and given referents are deaccented. However, results showed that – in addition to producing deaccented referents – speakers used seven different accent types. A second observation was that speakers produced not only falling contours (i.e., phrases that ended in a low boundary tone), but also rising contours

---

[5] The second line of the table gives the values for the reference level in the model, the intercept (here: items labelled L+H*). In order to obtain the values for the other levels (e.g., L*+H), their estimate needs to be added to the estimate of the intercept. The same procedure applies to the upper and lower bounds.

[6] One anonymous reviewer raised the concern that the procedure might bear the danger of circularity: The manual annotation uses information about the F0 minima and maxima (in addition to auditory information) to assign labels, which is then used as input for the statistical analysis. What is important to note here is that the statistical analysis is used as a safeguard against incorrect annotation. It is not circular to apply a rule for the assignment of a certain label and to check if the rule has been correctly applied.

Table 2

Frequency distribution (absolute and relative) of pitch accent types information status. The grey-shaded areas mark the most frequent realization in each condition.

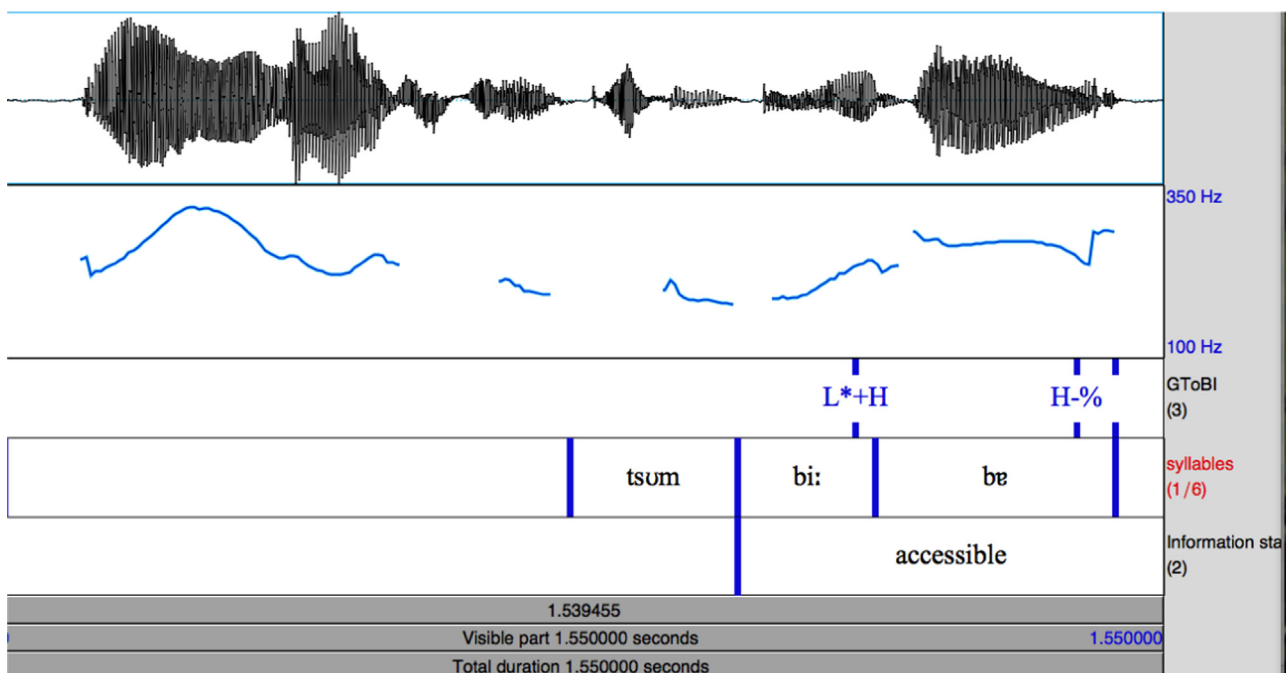|	Information status					
	New		Accessible		Given	
	N	%	N	%	N	%
deacc.	0	0	2	8.7	15	32.6
H*	11	21.6	3	13.0	11	23.9
L+H*	9	17.6	9	39.1	2	4.3
L*+H	13	25.5	6	26.1	1	2.2
H+L*	2	3.9	1	4.3	5	10.9
H+!H*	7	13.7	0	0	6	13.0
!H*	2	3.9	0	0	1	2.2
L*	7	13.7	2	8.7	5	10.9
Total	**51**	**100.0**	**23**	**100.0**	**46**	**100.0**

Fig. 4. F0 contour of the spontaneously produced utterance *Nun läuft er zum Biber* ('now he is walking towards the beaver'), in the accessible condition, realized with a L*+H accent followed by a high boundary tone.

(i.e., phrases that ended in a high boundary tone). The distribution of accentuation types (including deaccentuation) across the three conditions is given in Table 2.

New referents were always accented. Speakers realized them most often with a L*+H accent (25.5%), but H* (21.6%) and L+H* accents (17.6%) were also used. When realizing an accessible referent, speakers used L+H* accents most often, followed in frequency by L*+H accents. The most frequent realization for given referents was deaccentuation (32.6%), but still 67.4% of all given referents received an accent. Among those, H* accents were most frequent (23.9%). Fig. 4 shows an example of an utterance in the accessible condition using a L*+H accent.

As mentioned earlier, speakers produced target sentences with both low and high boundary tones: 45.8% (55) of all target sentences ended in a high or an upstepped boundary tone; 54.2% (65) ended in a low boundary tone.

3. Interim discussion

Speakers produced a range of different intonation patterns. Instead of only three expected accentuation types, there were eight. Recall that new referents should preferably be marked with a H* accent, accessible referents with the accent H+L* and given referents with deaccentuation (Baumann & Grice, 2006). Speakers in the current experiment always accented new referents, which is consistent with claims in the literature (e.g., Chafe, 1974; Lambrecht, 1994) and previous production studies (e.g., Baumann, 2006; Brown, 1983; Hirschberg, 1993; Terken & Hirschberg, 1994). However, no particular accent type emerged as the prototypical 'newness accent', contra hypothesis 1 formulated above. The accent H*, which is assumed to be a marker for newness (Baumann &

Grice, 2006; Pierrehumbert & Hirschberg, 1990) and the related L+H* were used by the speakers in about 40% of the cases, but about one quarter of new referents was realized with L*+H accents. This finding is at odds also with the suggestion that has been made (for English) that L* accents signal givenness (Pierrehumbert & Hirschberg, 1990). For accessible referents, speakers used L+H* accents most often, but not H+L* accents. H+L* accents were rare overall, and were used for less than 5% of all accessible referents. The picture does not change even if the accents H+!H* and H+L* are taken together as early peak accents. This result does thus not confirm Hypothesis 2. In line with previous findings (e.g., Baumann, 2006; Braun & Chen, 2010; Brown, 1983; Röhr & Baumann, 2010), given referents were deaccented often, but in more than two thirds of the cases given referents were still accented, contrary to Hypothesis 3. This finding corresponds with other reports that mere repetition of a referent does not necessarily lead to deaccentuation (Braun & Chen, 2010; Terken & Hirschberg, 1994).

The results of this study thus differ clearly from the mapping suggested by Baumann (2006) and Baumann and Grice (2006) and the empirical findings. What could be the reason for the differences between the suggested information status-intonation mapping and the distribution observed in the present experiment? One possible explanation is that the information statuses were not operationalized in the same way. In other words: It may be the case that for example accessible referents in the present experiment would not have been regarded as an accessible referent in the other studies. This is unlikely, since the referents in this study were designed to be very similar to textually accessible referents used in previous studies. Accessible referents in the present study were separated from their antecedent by at least four clauses, since all speakers produced at least one clause per picture (but often more). Consider the transcription of one story in (3), where the referent *Möwe* ('seagull') is mentioned first in the second sentence, and then again after the speaker talks about another action performed by the protagonist (walking along the beach and buying an ice-cream). The target referent is underlined.

(3) *Eine Frau steht am Strand. Sie hat eine Tüte mitgebracht, mit Brotkrumen, und füttert eine <u>Möwe</u>. Dann nimmt sie die die Tüte wieder zu ähm, zusammen, nee, und… verlässt den Strand. Ähm, nein, sie ist doch noch am Strand,… und geht auf einen… Eisstand zu, und kauft sich ein Eis. Sie nimmt das Eis, und geht wieder zurück. Sie… geht weiter spazieren am Strand, und sieht die <u>Möwe</u>.*

'There is a woman on the beach. She has brought a bag with her, with breadcrumbs, and is feeding <u>a seagull</u>. Then she takes the bag again to- uhm, together, no, and… leaves the beach. Uhm, no, she is in fact still at the beach and walks toward an ice cream seller and buys herself an ice cream. She takes the ice cream and walks back. She keeps walking along the beach, and sees the <u>seagull</u>.'

This context should yield the referent *Möwe* accessible, because it "has not been mentioned in the last two or three clauses and [is] thus only semi-active" (Baumann, 2006: 75). Furthermore in keeping with the structures used in the previous studies, the target referents were almost always (direct) objects and in phrase-final position. It is therefore unlikely that the discrepancies between these findings and those reported in the Introduction (Section 1.2) have to do with differences in the operationalization of information statuses. Another explanation may lie in the way in which the studies were set up. All previous studies analysed the intonation of read-out texts or ratings of the intonation of read-out texts. In contrast, speakers in this study constructed discourse on-line using only pictures as guidance for what to say. As described in the Introduction (Section 1.3), read speech and spontaneous speech have been observed to differ from each other in various aspects. It is not inconceivable that speaking mode affects intonational information status marking as well.

To test if speaking mode can account for the differences between my findings and those of earlier studies, I conducted a second study, using the same stimuli, but this time letting the speakers read prepared texts that described what was happening in the picture stories.
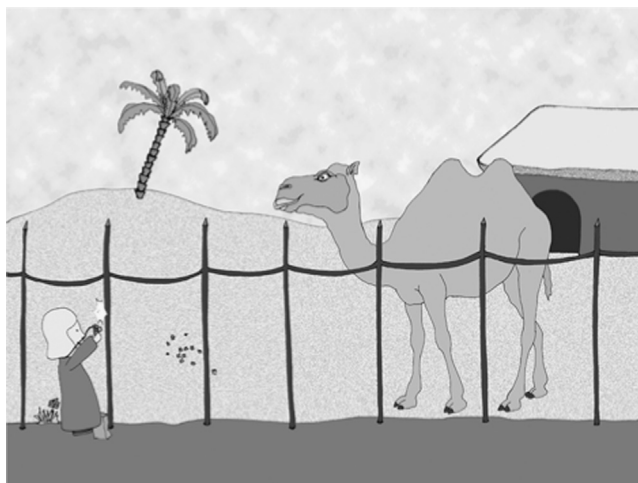

Fig. 5. Picture from the *Kamel* given story. The accompanying text was: "Das Mädchen nimmt seine Kamera und fotografiert das Kamel" 'the girl takes her camera and takes a picture of the camel'.

4. Study 2: Marking information status in read speech

4.1. Methods

4.1.1. Materials

The picture stories were the same as in Study 1. In contrast to the previous experiment, the pictures were this time accompanied by text. Under each picture, a sentence was provided that described the picture. An example is given in Fig. 5. In many cases, constructions were used that had been produced by the speakers in Study 1. In keeping with the selection criteria from Study 1, the target referent was always the last referring expression in the sentence, and in 10 out of 12 cases also the last word. In two sentences, the target referent was followed by a verb particle (*aus* 'out' and *mit* 'with'). Thus the sentences were very similar in structure to the sentences that speakers had produced spontaneously and which had been selected for analysis in Study 1. Again four different lists were created which had a different trial order.

4.1.2. Speakers

Eight adult native speakers of German participated in the experiment (all female, mean age 22.5 years). They were all students of Linguistics or Psychology at the University of Potsdam and had no reported history of any speech or hearing deficits. None of the participants had taken part in the previous experiment. The participation was rewarded with a financial reimbursement or course credit.

4.1.3. Procedure

The technical set-up was the same as in Study 1. Participants were randomly assigned to one of the four lists. As in Study 1, participants were told that someone listening to their story should be able to retell it. They were instructed to first look at the picture and read the printed sentence silently before reading it out aloud.

4.1.4. Data annotation, data selection and accent label consistency check

Of the 96 items (8 participants × 4 referents × 3 conditions), 11 had to be excluded because of disfluencies or creaky voice. The remaining 85 items were annotated as described for Study 1. The various accentuation types were again compared along a number of acoustic parameters, using the same procedure as in Study 1. The results showed that in this data set, too, potentially confusable accent pairs (e.g., L+H* and L*+H) differed from each other significantly in the way predicted from the GToBI literature. (The details of the analysis are provided in the Supplementary materials).

4.2. Results

Speakers produced the referents with six different pitch accent types and with deaccentuation. Almost all utterances (83) were produced with a low boundary tone. New referents were always accented. The most frequent realization for new referents was H+!H* (34.5%), followed by H+L* (31.0%), the two accents that have recently been taken to be of one category, early peak accents. Accessible referents were most often produced with H+L* accents (32.1%), followed by H+!H* accents (28.6%). Given referents were deaccented almost all of the time (92.9%). Table 3 shows the distribution of the different pitch accent types across the three experimental conditions (new, accessible, given).

An example of the *F*0 contour of a read utterance with a referent in the new condition is given in Fig. 6.

The results from the reading experiment thus confirm Hypotheses 2 and 3, but not Hypothesis 1, since new referents were not predominantly marked with H* accents.

Table 3
Distribution (absolute and relative) of accentuation types by condition in the reading experiment. The grey-shaded areas mark the most frequent realization in each condition.

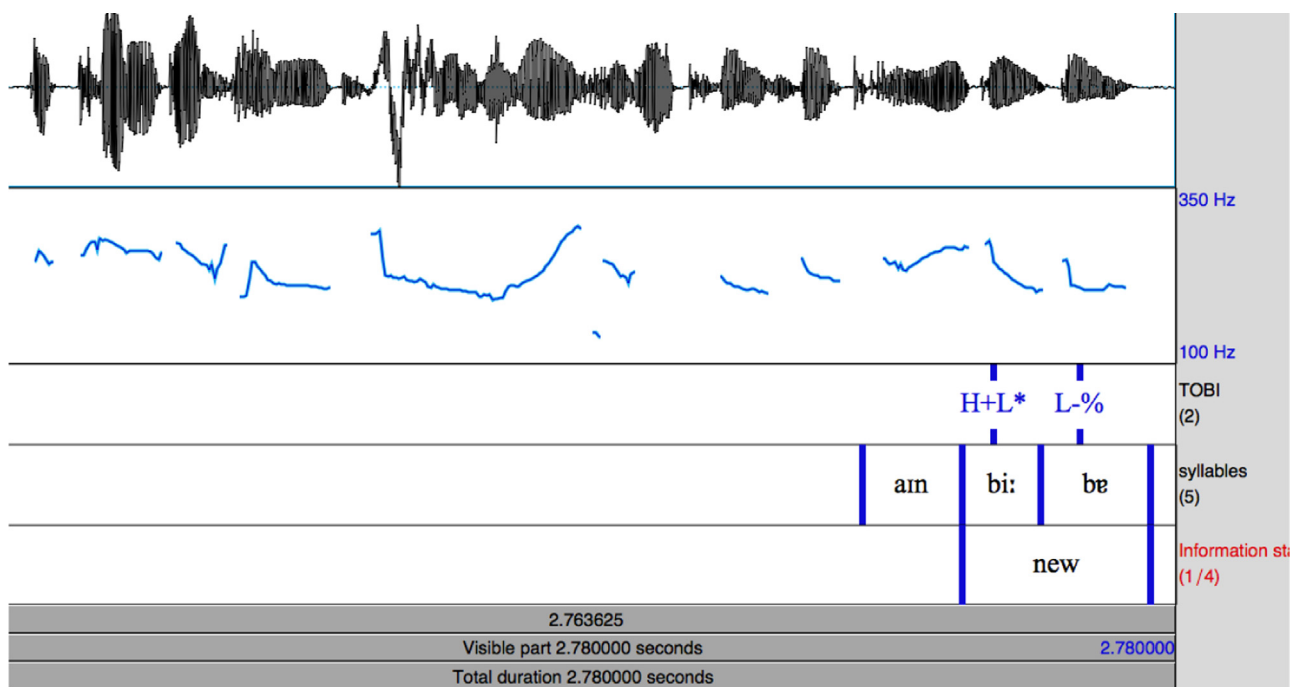|		Information status					
		New		Accessible		Given	
		N	%	*N*	%	*N*	%
Accent	deacc.	0	0	5	17.9	26	92.9
	H*	7	24.1	2	7.1	0	0
	L+H*	3	10.3	1	3.6	0	0
	L*+H	0	0	1	3.6	0	0
	H+L*	9	31.0	9	32.1	0	0
	H+!H*	10	34.5	8	28.6	2	7.1
	!H*	0	0	0	0	0	0
	L*	0	0	2	7.1	0	0
	Total	**29**	**100.0**	**28**	**100.0**	**28**	**100.0**

Fig. 6. *F*0 contour of the read out utterance *und streichelt einen Biber* ('and strokes a beaver'), in the new condition, realized with a H+L* accent and a low boundary tone.
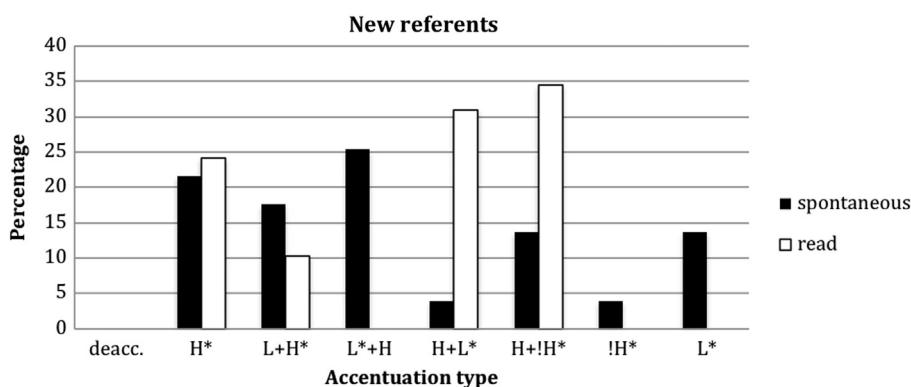

Fig. 7. Distribution of accentuation types in spontaneous speech (*N*=51) and read speech (*N*=29) for the new condition.

5. Comparing spontaneous and read speech results

The accent distributions from Study 1 and Study 2 seem to differ considerably. For ease of presentation, the distributions of accentuation types for the three conditions in both speaking modes are shown in Figs. 7–9.

I used Fisher's exact tests (Fisher, 1922) to test this impression statistically. (Both data sets are unbalanced and contain cells with only few cases or even zero, which precludes the use of linear mixed-effect models, generalized linear models as well as Pearson's chi-squared tests.) In a first step, I calculated the expected outcomes for the read data set (Study 2) from the distribution in the spontaneous data set (Study 1). I then tested for each information status (new, accessible, given) if the observed outcomes (i.e., the distribution of accentuation types in the read data) differ significantly from the expected outcomes (i.e., the distribution derived from the spontaneous data set). It turned out that for all three information statuses, the observed outcomes differ significantly from the expected ones (New condition: $p<.001$; Accessible condition: $p<.001$; Given condition: $p<.0001$). Four differences between spontaneous and read speech are particularly striking[7]:

- In contrast to spontaneous speech, new referents in read speech are never realized with L*+H accents.
- Compared to spontaneous speech, the percentage of H+L* accents for accessible referents is higher in read speech.
- Compared to spontaneous speech, accessible referents in read speech are much more often realized with H+!H* accents.
- Compared to spontaneous speech, the percentage of deaccented given referents is much higher in read speech.

---

[7] Note again that the analysis is restricted to referents in nuclear position.
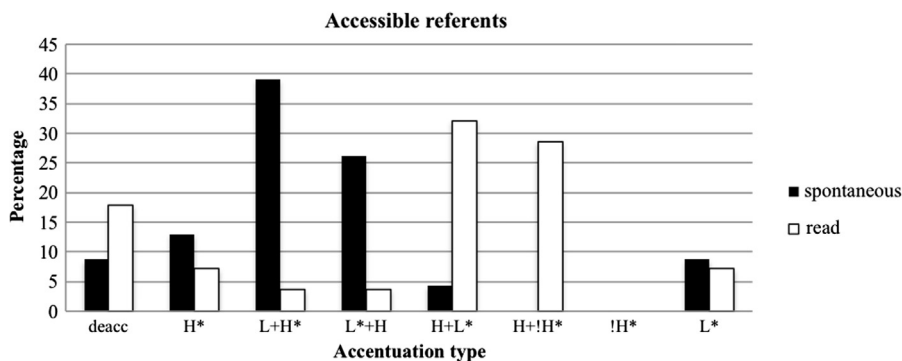
Fig. 8. Distribution of accentuation types in spontaneous (*N*=23) and read speech (*N*=28) for the accessible condition.
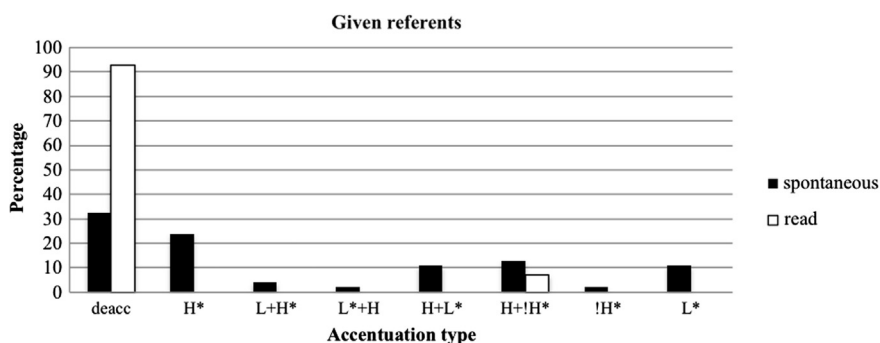

Fig. 9. Distribution of accentuation types in spontaneous (*N*=46) and read speech (*N*=28) for the given condition.

I tested whether these percentages (e.g., the percentage of deaccented given referents in read speech and the percentage of deaccented given referents in spontaneous speech) differed from each other significantly using a *z*-test for two proportions. For this test, proportions need to be transformed into *z*-values (Fisher *z*-transformation) and then compared to each other (see Hinkle, Wiersma, & Jurs, 2003).[8]

All four differences were statistically significant: The percentage of new referents realized with L*+H accents in read speech is significantly lower than in spontaneous speech ($z=2.92$, $p<.01$). In read speech, accessible referents were significantly more often realized with both H+L* ($z=2.48$, $p<.01$) and with H+!H* accents ($z=2.73$, $p<.01$). Finally, given referents were significantly more often deaccented in read speech than in spontaneous speech ($z=5.05$, $p<.0001$). In addition, there were significantly more low-ending utterances in read speech than in spontaneously produced speech ($z=6.84$, $p<.0001$).

6. General discussion

Unlike the results from Study 1 (spontaneous speech), the results in Study 2 (read speech) mostly confirm the hypotheses formulated in Section 1.4. Different results were thus obtained depending on speaking mode, suggesting that the way speakers mark information status in read speech generally differs from what they do in spontaneous speech. In the following, I will discuss these results in connection with a recent study by Baumann and Riester (2013), which also looked at effects of givenness on prosody in read and spontaneous speech.[9] I will then go on to some general deliberations on the multifunctionality of intonation and the special status of reading intonation. Finally, I will discuss the implications that the choice of methodology in intonation research can have for the study of language development in children.

Before entering the discussion of the results, however, I would first like to address some potential criticisms that might be raised against the methodological approach used here. First, there is the relatively small data set, which can be a concern for reliability. Due to the strict selection criteria, which were necessary to control for several potentially confounding factors, only 205 items were included (120 in Study 1, 85 in Study 2). But even in this dataset, significant differences between the speaking styles emerged, showing that this factor does have a clear impact on intonation. Second, one might be inclined to argue that although the narrations that speakers produced in the two studies were very similar (recall that the pre-set stories were formulated in such a way as to match closely the structures the speakers in Study 1 had produced), they were not strictly identical, which could potentially have reduced the validity of the comparison. Specifically, in Study 1 speakers sometimes used different syntactic constructions different from those

---

[8] In the case of zero occurrences (e.g., for L*+H accents in the new condition in read speech) a value of .1 was entered to enable the calculation of a proportion.

[9] Note that this study had not yet appeared at the time the study reported upon in this article was conducted.

given to speakers in the reading experiment. For instance, one speaker in Study 1 uttered the sentence "And now he is coming back to his camel", where the target word *camel* is just an adjunct, instead of the pre-set "He is returning home and stroking his camel" in the reading experiment, where the target word is a direct object of the verb. The spontaneously produced sentences containing the target words also varied in the number of words. Moreover, when talking "on the fly", speakers might create contrastive constructions, which typically induce contrastive intonation that overrides deaccentuation. However, the presence of these differences does not prevent us from extracting useful information with regard to the primary research question at hand. The scenes in the picture stories did not, for instance, invite constructing any contrasts between the target referent and some other object. Due to the simplicity of the depicted stories, speakers produced rather similar descriptions across the board, which makes it likely that they also "perceived" the referents they were talking about to have the same information statuses. Of course, as an observer, one can never be sure what goes on in a speaker's mind. But this is a general methodological constraint that applies to, for instance, written texts as well. On a more general level, one might expect linguistic hypotheses (like "given referents are deaccented") to be robust against variations in the number of words in a sentence. Thus, although there are some inevitable uncertainties attached to the method, it is sound enough to warrant a valid comparison of the two speaking modes.

In this comparison, I will first consider the use of boundary tones, which was strikingly different in both experiments. While in the spontaneous narrations nearly half of all utterances ended in a high boundary tone, in the read-out narrations almost all utterances were low ending. Differing distributions of boundary tones across speaking modes have been reported before. For Dutch, Blaauw (1995) observed that when speakers produced instructions spontaneously, they used predominantly rising tunes (i.e., high boundary tones). In contrast, when they read out their orthographically transcribed instructions later, speakers showed a clear preference for low-ending phrases. Baumann and Riester (2013) report similar observations. They investigated the role of different types and degrees of givenness of referential expressions on prosodic markings both in read speech and in uncontrolled monologues. Accordingly, they did not – unlike the present study – use the same stimuli for both conditions (i.e., read out speech and spontaneously produced speech). They found that in uncontrolled monologues, more than 40% of all phrases ended high, often in combination with L* accents. What may be the reason for this preference for high boundary tones in spontaneous speech? It has been suggested that in spontaneous speech non-low boundary tones are used to indicate 'topic continuity' (Brown, Currie, & Kenworthy, 1980; Swerts & Geluykens, 1994), while the end of a topic (topic finality) is signalled with a low boundary tone. Blaauw (1995) assumes that this is different in read speech, where finality is signalled at the level of the sentence. As a consequence, there are more low boundary tones in read speech. However, pitch direction in read speech is also clearly dictated by punctuation: A full stop canonically suggests falling intonation, so speakers who read sentences that end in a full stop will go down with their $F0$.

Turning to the choice of pitch accent type, clear differences between spontaneous speech and read speech emerged as well. Firstly, while L*+H accents figured prominently in the speakers' spontaneous productions, they were virtually absent in read speech. Secondly, the accents H+!H* and H+L* were produced much more often in read speech, with H+L* accents notably being the most frequent realization for accessible referents. Thirdly and perhaps most strikingly, only about 33% of given referents were deaccented in the spontaneous data, whereas it was more than 90% in the read data. The comparison of the two sets of results clearly shows that different intonation patterns are obtained depending on the task. Although the participants saw the same picture stories with the target referents having the same information status as when the stories were told freely, they used accent types and deaccentuation in very different distributions.

The L*+H accents in the spontaneous narrations – which were used for new and accessible referents – were always combined with high boundary tones. As mentioned before, Baumann and Riester (2013) made a similar observation in their analysis of spontaneous, unconstrained monologues. In their data, the accent L* was the most frequently used nuclear accent, and was typically followed by a high boundary tone. It has been noted before that there is generally a high correlation between the type of boundary tone and the type of pitch accent (Dainora, 2002), in the sense that high boundary tones typically occur with low pitch accents and vice versa. If speakers have a preference for combining contrasting pitch accents and boundary tones, it is conceivable that the use of low pitch accents in the narrations was a consequence of the speakers' decision to use high boundary tones to indicate continuity.[10]

Unlike spontaneous speech, read speech was furthermore characterized by a preponderance of H+L* and H+!H* accents. Recall from the Introduction (Section 1.1) that it is unclear if they should be taken to be variants of a single accent. Baumann and Riester (2013), following Kohler (1991), collapsed them into one accent category, early peak. In the present data, H+L* accents and !H+H* accents were distributed very similarly in both the new and the accessible condition (each accent occurred around 30% of the time). This suggests that speakers use them interchangeably, lending support to the assumption that these two accents are simply variants of a single early peak accent. The finding that accessible referents were predominantly marked by early peak accents is in line with Hypothesis 2. However, the fact that new accents were equally often realized with early peak accents casts doubt on the idea that early peak accents are markers of accessibility. If this were the case, we would not expect them to be used to such a great extent for new referents, too. In Baumann and Riester's (2013) read data, early peak accents were the most frequently used accent for new referents[11] and referents accessible via inference as well. Both results indicate that – even in read speech – speakers appear not to

---

[10] Note that this assumption would imply that pitch accents and boundary tones are not functionally independent from each other. For more on this topic, see Dainora (2006).

[11] The authors classified information statuses using the RefLex scheme (Baumann & Riester, 2012), which differentiates between referential and lexical givenness. An entity is referentially given if it has occurred in the discourse before, but not necessarily using the same lexical form (e.g., *die Bundeskanzlerin* – *Frau Merkel*, 'the Chancellor' – 'Mrs. Merkel'). An item is lexically given, on the other hand, if the same word has been used before, although it need not necessarily refer to the same referent. Different levels of referential and lexical givenness can thus be combined. The present article deals only with the "extreme cases" of referents that are either both lexically *and* referentially new, lexically given *and* referentially given, or accessible. For this reason, the discussion of Baumann and Riester's (2013) results is restricted to those cases as well.

distinguish intonationally between newness and at least certain types of accessibility. In this context it is also interesting to note that previous studies on German have found similar patterns: Braun (2006) reports that speakers produced mainly (!)H+L* nuclear accents in her reading experiment. Braun's study was not concerned with information status marking, but the elicitation of contrastive and non-contrastive topics and foci. However, the participants read short paragraphs that were in their format comparable to the short stories used in Study 2. Both the present findings and the reports in the literature (Baumann, 2006; Braun, 2006) can be taken as evidence that H+L*/H+!H* accents have to some extent become conventionalized elements of read-out narration and reporting – a point which will be taken up again below.

The third remarkable outcome was that speakers deaccented given referents in read speech quite consistently (in accordance with Hypothesis 3), but did so in only about a third of the cases when they narrated spontaneously. Baumann and Riester (2013) also found differences in the proportional use of deaccentuation in their two corpora, with deaccentuation of given referents occurring more often in read (66%) than in spontaneous speech (17%). The rate of deaccentuation in read speech was lower in Baumann and Riester's data than in the present study. This is most likely due to the fact that they also included referents in prenuclear position. Prenuclear accents are often produced purely for rhythmic reasons (e.g., Féry & Kügler, 2008), so it is not surprising to find more accents on given referents when prenuclear positions are included in the analysis. But why is deaccentuation generally less common in spontaneous speech? Baumann and Riester (2013) offer as an explanation that their spontaneous data contained more instances of predications (e.g., "Pete is the BOSS.") and more contrastive expressions that may have overridden information status marking prosodically (Baumann & Riester, 2013: 34). This appears to be a plausible explanation, especially since – unlike the speakers in the present study – the monologue speakers were absolutely unconstrained with respect to the contents and structure of their contributions.[12] In addition, we need to take into account that in spontaneous speech, functions other than information status or focus marking may be more important: signalling discourse structure, keeping the floor (in a dialogue situation), flagging uncertainty or indicating attitudes towards the contents of the message (e.g., ironic stance) etc. Depending on the situation, these different functions of intonation will be more or less important.[13]

What is more, in reading aloud speakers do not have to allocate as many resources to speech planning as a speaker in spontaneous discourse (see also Brown et al., 1980, p. 47). With "burdens" like these taken away from the speakers, they can concentrate on using intonation to a greater extent to reflect the information structure of the text – by deaccenting given referents, or by marking contrasting topics, for instance. Intonational information structure marking may therefore be especially pronounced in reading aloud. Written texts are often syntactically more complex and have a higher informational density than unscripted spoken language (Biber, 1988; Chafe, 1982). When reading such texts aloud, marking relations between parts of the sentence (e.g., topic-comment) or disambiguating syntax intonationally makes the text easier to process for the listener. As a consequence, using intonation in this way has probably become a codified element of "good reading" and should be viewed as a cultural skill that is learned in the course of literacy training. In fact, researchers working on the development of oral reading fluency have even remarked that, as children become skilled readers, their reading prosody takes on "a culturally normative character" (Schwanenflugel, Hamilton, Kuhn, Wisenbaker, & Stahl, 2004: 127). In several studies children were categorized as good readers if (among others) their intonation contour matched that of adults closely and if their rendition of the text exhibited steep $F0$ declination at the end of declarative sentences (e.g., Miller & Schwanenflugel, 2006, 2008). This shows that people have a clear idea of what constitutes good reading intonation, and also that this intonation does not vary strongly, since the values from adult readers are typically averaged to yield norm values. This somewhat prescriptive view of intonation in the context of reading aloud may also explain the findings from the perception studies referred to in the Introduction (Baumann, 2006; Baumann & Grice, 2006). Here, participants listened to read-out stimuli. It has been found that people can identify speech as read-out rather than spontaneous, even when ungrammatical structures and hesitations have been removed (Blaauw, 1994; Laan, 1997). If clear information status marking (e.g., deaccentuation of given referents) is an important feature of reading intonation, listeners – fully aware that they were listening to read speech – are likely to rate intonation deviating from this as less appropriate. It is unlikely that reading intonation would be completely different from what people do in natural speech, but some aspects may have become exaggerated, modified and conventionalized. The questions of why reading intonation may have developed in a certain way, and how exactly it deviates from spontaneous intonation are not the primary concerns of this study (but see Monschau, 2004 for a review of research on reading intonation in English). What the two studies presented here have found, however, is that there *were* clear differences between spontaneous and read speech in the way speakers marked information status. This has implications for researchers who study this topic in children: Since children learn language mostly from spontaneous speech, intonation obtained in reading studies cannot straightforwardly be assumed to be the target for children. In fact, in doing so, one would arrive at an incorrect assessment of children's development. To take the example of information status marking: Read-out data (in Study 2, but also in Baumann & Grice, 2006) suggested that accessible referents are typically realized with early peak accents. The children in De Ruiter (2014), who narrated the same picture stories as the adults in Study 1, however, did not produce these early peak accents. The logical conclusion from this observation in combination with the read-out results would be that children have not yet learned how to mark textual accessibility. But this is not true, because – as has

---

[12] In this context it is interesting to note that a new approach to the study of accent choice (for spontaneous American English) is based on a probabilistic model of accent assignment (Calhoun, 2010), taking into account prosodic structure, different information structural dimensions and lexical 'accentability'. Although highly interesting, evaluating the present results within this framework would be outside of the scope of this paper.

[13] One anonymous reviewer remarked that this most probably holds for read speech as well. For instance, a speaker may want to mark his utterance in a particular way (as 'read') in order to control the extent to which he is held responsible for what he is saying. This is an important point, a discussion of which is unfortunately beyond the scope of this article. I do believe, however, that for instance distancing oneself from the content of the written texts that speakers are typically used to read in the studies mentioned so far probably plays only a minor role.

been demonstrated in this article – in spontaneous speech, accessible referents are not marked like that. Similarly, De Ruiter (2014) observed that 7-year-old children used continuation rises (i.e., L* H-% contours), while 5-year-olds did not. When this child data is compared to read-out adult data (which does not contain continuation rises), this would suggest that 5-year-olds are more adult-like, but 7-year-olds are not. Having investigated spontaneous speech, however, we now know that quite the opposite is the case: 7-year-olds are more similar to adults, exactly *because* they have already learned that continuity rises are used to signal discourse structure. This shows that child language researchers interested in phonetic and phonological acquisition need to be very careful when collecting adult speech data to determine the target model. Reading aloud can only be used if it has been established that it yields the same results as spontaneous speech. This has been shown to be the case in voice onset time production, for instance (Brown & Docherty, 1995). But it is evidently not true for (at least certain aspects of) intonation.

7. Summary and conclusion

The present study investigated if speaking mode – that is, whether something is produced from a prepared script or spontaneously – influences the way speakers mark the information status of discourse referents. The findings showed that there is a clear difference in intonation between read and spontaneous speech. In spontaneous speech, speakers always accented new referents, but they did not use designated pitch accent types to differentiate between new and accessible referents, nor did they deaccent given referents throughout, contrasting with claims made in the literature. In addition, these speakers often made use of low pitch accents in combination with high boundary tones to indicate continuity. In contrast (but in accordance with reported findings in the literature), read speech was characterized by low boundary tones, consistent deaccentuation of given referents and the use of early peak (H+L*/H+!H*) accents, for both new and accessible referents.

With respect to linguistic theory, these findings contribute to our growing understanding that there is no one-to-one mapping between information statuses and pitch accents types (cf. Baumann & Riester, 2013), however attractive this idea may be. With respect to the methodology of intonation research, the findings raise an important issue regarding the ecological validity of reading studies. Reading tasks of course have a place in phonological and phonetic research. But if certain intonational patterns can only be obtained under very controlled circumstances, such as having speakers read out especially designed materials, we have to be very cautious in interpreting these results when our aim is to study intonation patterns in real-time verbal communication.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.wocn.2014.10.008.

References

Ariel, M. (1988). Referring and accessibility. *Journal of Linguistics*, *24*, 65–87.
Ariel, M. (1990). *Accessing noun-phrase antecedents*. London: Routledge.
Baayen, H. R. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.
Barik, H. C. (1977). Cross-linguistic study of temporal characteristics of different types of speech materials. *Language and Speech*, *20*, 116–126.
Baumann, S. (2006). *The intonation of givenness – Evidence from German*. Vol. 508. Tübingen: Niemeyer.
Baumann, S., Brinckmann, C., Hansen-Schirra, S., Kruijff, G. -J., Kruijff-Korbayova, I., Neumann, S., & et al. (2004). Multi-dimensional annotation of linguistic corpora for investigating information structure. In *Proceedings of the NAACL/HLT frontiers in corpus annotation*.
Baumann, S., Brinckmann, C., Hansen-Schirra, S., Kruijff, G.-J. M., Kruijff-Korbayova, I., Neumann, S., &et al. (2004). The MULI Project: Annotation and analysis of information structure in German and English. In *LREC*.
Baumann, S., & Grice, M. (2006). The intonation of accessibility. *Journal of Pragmatics*, *38*, 1636–1657.
Baumann, S., & Riester, A. (2012). Referential and lexical givenness: Semantic, prosodic and cognitive aspects. In G. Elordieta, & P. Prieto (Eds.), *Prosody and meaning*, Vol. 25 (pp. 119–162). Berlin, New York: Mouton De Gruyter.
Baumann, S., & Riester, A. (2013). Coreference, lexical givenness and prosody in German. *Lingua*, *136*, 16–37.
Beckman, M., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309.
Benzmüller, R., & Grice, M. (1997). Trainingsmaterialien zur Etikettierung deutscher Intonation mit GToBI. *Phonus*, *3*, 9–34.
Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
Biber, D. (1993). The multi-dimensional approach to linguistic analyses of genre variation: An overview of methodology and findings. *Computers and Humanities*, *26*, 331–345.
Blaauw, E. (1994). The contribution of prosodic boundary markers to the perceptual difference between read and spontaneous speech. *Speech Communication*, *14*, 359–375.
Blaauw, E. (1995). *On the perceptual classification of read and spontaneous speech* (Doctoral dissertation). Utrecht University.
Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.55, retrieved September 2013 from http://www.praat.org/.
Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech*, *49*, 451–493.
Braun, B., & Chen, A. (2010). Intonation of 'now' in resolving scope ambiguity in English and Dutch. *Journal of Phonetics*, *38*, 431–444.
Brown, A., & Docherty, G. J. (1995). Phonetic variation in dysarthric speech as a function of sampling task. *European Journal of Disorders of Communication*, *30*, 17–35.
Brown, G. (1983). Prosodic structure and the given/new distinction. In A. Cutler, & D. R. Ladd (Eds.), *Prosody: Models and measurements* (pp. 67–77). Berlin: Springer.
Brown, G., Currie, K. L., & Kenworthy, J. (1980). *Questions of intonation*. London: Routledge.
Calhoun, S. (2010). How does informativeness affect prosodic prominence?. *Language and Cognitive Processes*, *25*, 1099–1140.
Chafe, W. L. (1974). Language and consciousness. *Language*, *50*, 111–133.
Chafe, W. L. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 35–53). Norwood: Ablex.
Chafe, W. L. (1987). Cognitive constraints on information flow,. *Coherence and Grounding in Discourse*, *11*, 21–51.
Chafe, W. L. (1994). *Discourse, consciousness and time*. Chicago/London: The University of Chicago Press.
Dainora, A. (2002). Modeling intonation in English: A probabilistic approach to phonological competence. In L. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory phonology* (p. *8*). Berlin: Mouton de Gruyter.
Dainora, A. (2006). Modeling intonation in English: A probabilistic approach to phonological competence. In L. Goldstein, D. H. Whalen, & C. Best (Eds.), *Laboratory phonology*, Vol. 8 (pp. 107–132). Berlin, New York: Mouton De Gruyter.

De Ruiter, L. E. (2009). The prosodic marking of topical referents in the German "Vorfeld" by children and adults. *The Linguistic Review*, *26*, 329–354.

De Ruiter, L. E. (2014). How German children signal information status in narrative discourse. *Journal of Child Language*, *41*, 1015–1061.

Eskénazi, M. (1992). Changing speech styles: Strategies in read speech and casual and careful spontaneous speech. In *Second international conference on spoken language processing*. Banff, Alberta, Canada.

Féry, C., & Kügler, F. (2008). Pitch accent scaling on given, new and focused constituents in German. *Journal of Phonetics*, *36*, 680–703.

Fisher, R. A. (1922). On the interpretation of $\chi 2$ from contingency tables, and the calculation of *P. Journal of the Royal Statistical Society*, 87–94.

Grice, M., & Baumann, S. (2002). Deutsche Intonation und GToBI. *Linguistische Berichte*, *191*, 267–298.

Grice, M., Baumann, S., & Benzmüller, R. (2005). German intonation in autosegmental-metrical phonology. In J. Sun-Ah (Ed.), *Prosodic typology. The phonology of intonation and phrasing* (pp. 55–83). Oxford: Oxford University Press.

Grice, M., Baumann, S., & Jagdfeld, N. (2009). Tonal association and derived nuclear accents – The case of downstepping contours in German. *Lingua*, *119*, 881–905.

Grice, M., Ladd, D. R., & Arvaniti, A. (2000). On the place of phrase accents in intonational phonology. *Phonology*, *17*, 143–185.

Grice, M., Reyelt, M., Benzmüller, R., Mayer, J., & Batliner, A. (1996). Consistency in transcription and labelling of German intonation with GToBI. In: *Proceedings of the 4th international conference on spoken language processing (icslp)* (pp. 1716–1719). Philadelphia.

Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, *69*, 274–307.

Gussenhoven, C. (2005). Transcription of Dutch intonation. In S.-A. Jun (Ed.), *Prosodic typology and transcription: A unified approach* (pp. 118–145). Oxford: Oxford University Press.

Hinkle, D., Wiersma, W., & Jurs, S. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Boston: Houghton Mifflin Company.

Hirschberg, J. (1993). Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence*, *63*, 305–340.

Hirschberg, J., Litman, D., Pierrehumbert, J., & Ward, G. (1987). Intonation and the intentional structure of discourse. In: *Proceedings of the 10th international joint conference on artificial intelligence* (pp. 636–639): Morgan Kaufmann Publishers.

Howell, P., & Kadi-Hanifi, K. (1991). Comparison of prosodic properties between read and spontaneous speech material. *Speech Communication*, *10*, 163–169.

Kohler, K. (1991). Terminal intonation patterns in single-accent utterances of German: Phonetics, phonology and semantics. *Arbeitsberichte des Instituts für Phonetik und digitale Sprachverarbeitung der Universität Kiel* (*AIPUK*) (Vol. 25, pp. 115–185).

Laan, G. P. M. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication*, *22*, 43–65.

Ladd, D. R. (1980). *The structure of intonational meaning: Evidence from English*. Bloomington: Indiana University Press.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge, England; New York, NY, USA: Cambridge University Press.

Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge; New York, NY, USA: Cambridge University Press.

Lehiste, I. (1979). Sentence boundaries and paragraph boundaries – Perceptual evidence. In: *Clyne, P.R., Hanks, W.F. and Hofbauer, C.L. (Eds.), The element: A parasession on linguistic units and levels. Chicago Linguistic Society* (Vol. 15, pp. 99–109).

Levin, H., Schaffer, C. A., & Snow, C. (1982). The prosodic and paralinguistic features of reading and telling stories. *Language and Speech*, *25*, 43–54.

Miller, J., & Schwanenflugel, P. J. (2006). Prosody of syntactically complex sentences in the oral reading of young children. *Journal of Education Psychology*, *98*, 839–853.

Miller, J., & Schwanenflugel, P. J. (2008). A longitudinal study of the development of reading prosody as a dimension of oral reading fluency in early elementary school children. *Reading Research Quarterly*, *43*, 336–354.

Monschau, J. (2004). *Input and output in oral reading in English: The interaction of syntax, semantic-pragmatics and intonation* (Unpublished Ph.D. thesis). Bonn: Rheinische Friedrich-Wilhelms-Universität.

Pierrehumbert, J. B. (1980). *The phonetics and phonology of English intonation* (Unpublished doctoral dissertation). Bloomington: MIT.

Pierrehumbert, J. B., & Hirschberg, J. (1990). The meaning of intonational contours in the interpretation of discourse. In P. R. Cohen, J. Morgan, & M. E. Pollack (Eds.), *Intentions in communication* (pp. 271–311). Cambridge: MIT Press.

Pitrelli, J. F., Beckman, M., & Hirschberg, J. (1994). Evaluation of prosodic transcription labelling reliability in the ToBI framework. In *the 3rd international conference on spoken language processing (ICSLP)* (Vol. 2, pp. 123–126). Yokohama.

R Development Core TeamM. E. (2008). *R: A language and environment for statistical computing* (2.6.2 ed.). Vienna: R Foundation for Statistical Computing.

Rathcke, T., & Harrington, J. (2010). The variability of early accent peaks in Standard German. *Laboratory Phonology*, *10*, 533–555.

Röhr, C., & Baumann, S. (2010). Prosodic marking of information status in German. In *Proceedings of speech prosody*. Chicago.

Röhr, C., & Baumann, S. (2011). Decoding information status by type and position of accent in German. In *Proceedings of the 17th ICPhS* (pp. 1706–1709).

Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M., & Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology*, *96*.

Silverman, K. E., Blaauw, E., Spitz, J., & Pitrelli, J. F. (1992). Towards using prosody in speech recognition/understanding systems: Differences between read and spontaneous speech. In *Proceedings of the workshop on speech and natural language* (pp. 435–440). Association for Computational Linguistics.

Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information-flow in spoken discourse. *Language and Speech*, *37*, 21–43.

Swerts, M., & Heldner, M. (1996). *F*0 declination in spontaneous and read-aloud speech. In *Proceedings of the 4th international conference on spoken language processing (ICSLP)* (Vol. 3, pp. 1501–1504). Philadelphia, PA.

Syrdal, A. K., & McGory, J. (2000). Inter-transcriber reliability of ToBI prosodic labeling. In *Proceedings of the 6th international conference on spoken language processing (ICSLP)* (Vol. 3, pp. 235–328). Beijing, China.

Terken, J. M. B. (1984). The distribution of pitch accents in instructions as a function of discourse structure. *Language and Speech*, *27*, 269–289.

Terken, J. M. B., & Hirschberg, J. (1994). Deaccentuation of words representing 'given' information: Effects of persistence of grammatical role and surface position. *Language and Speech*, *37*, 125–145.

Terken, J. M. B., & Nooteboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes*, *2*, 145–163.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). Elan: A professional framework for multimodality research. In *Proceedings of LREC* (Vol. 2006).