

Global Prosodic Parameters and Anaphora Resolution¹

Ekaterina Jasinskaja, Ulrike Kölsch, & Jörg Mayer

{jasinsk ; ukoelsch}@rz.uni-potsdam.de / mayer@ling.uni-potsdam.de

University of Potsdam, Germany

Summary: Discourse structure is reflected by a number of global prosodic parameters. Pauses between discourse segments that pertain to different (sub)topics are longer than pauses between segments within one topic. Pitch range is expanded at the beginning of a discourse topic and compressed at the end. Discourse structure is also known to affect the accessibility/salience of antecedents of anaphoric expressions. If a new referent is introduced in a subordinated discourse segment (subtopic), it is no longer accessible for an anaphoric pronoun after shifting from the subtopic back to the main topic. Assuming these generalizations are correct, one can ask whether listeners use the information encoded in pauses and pitch range to resolve anaphoric references in ambiguous contexts. To examine this, we conducted a perception experiment with ambiguous discourses. The pitch range of the sentences and the pause duration between sentences was manipulated using PRAAT resynthesis techniques. The results of our experiment corroborate our main research hypothesis that global prosodic parameters influence the resolution of anaphoric pronouns. Moreover, the direction of the observed effect is clearly in accordance with the predictions of the existing theories of discourse anaphora and the current state of research on discourse prosody.

1. Introduction

This paper presents the results of an experiment on the influence of prosody on the interpretation of anaphoric pronouns. Until now, empirical studies in this area have predominantly concentrated on local prosodic features of pronouns, such as pitch accent (cf. e.g. Venditti et al. 2002, and references therein), whereas the impact of global prosodic parameters of an utterance, such as pitch range or pause duration, was almost entirely ignored. At the same time, the current state of research on discourse anaphora on the one hand and global prosody on the other strongly suggests that a link must exist between these two (seemingly unrelated) phenomena. On the one hand, it is an established fact that discourse structure affects accessibility/salience of possible antecedents to anaphoric expressions (Grosz and Sidner 1986, Polanyi 1988, Cristea et al. 1998, Asher and Lascarides 2003). On the other hand, pitch range and pauses have been shown to signal the structure of a spoken monologue. With the present study we want to fill the gap in empirical research and check whether the expectation that global prosodic parameters affect anaphora resolution is correct.

2. Background

2.1 Discourse structure and anaphoric accessibility

One of the factors that constrain the resolution of an anaphoric pronoun is the hierarchical discourse structure of the context in which the potential antecedent is mentioned. Roughly, if a new referent is introduced in a subordinated discourse unit (or subtopic) it is no longer *accessible* for anaphoric

¹ The research reported in this paper was sponsored by the German Research Community (DFG) as part of the Collaborative Research Center 632 (SFB) on "Information Structure" at the University of Potsdam. We would also like to thank Robin Hörnig, Stefanie Jannedy, Johannes Neubarth, Annika Neumann, Norman Schenk, Marcus Thienert and Thomas Weskott for valuable suggestions and practical help.

reference after a shift from the subtopic back to the main topic. The following example illustrates this idea:

- (1) (a) **Lena** war glücklich nach dem Tennisturnier.
Lena was happy after the tennis tournament
- (b) Die Silbermedaille war ein großer Erfolg.
the silver medal was a great achievement
- (c) **Die Trainer=**in gratulierte nach der Siegerehrung.
the coach=FEM congratulated after the award ceremony
- (d) Für das nächste Turnier wünscht **sie** sich
for the next tournament wishes she herself
 allerdings den ersten Platz.
however the first place
- (2) (a) **Lena** was happy after the tennis tournament.
 (b) The silver medal was a great achievement.
 (c) **The coach** congratulated [her] after the award ceremony.
 (d) For the next tournament, however, **she** hopes for the first place.

The German discourse in (1), as well as its English translation in (2) allow for at least two possible structures, shown in figures 1 and 2. In both cases the sentences (b) and (c) form a constituent that is connected to (a) by a subordinating discourse relation, e.g. *Explanation*: the segments (b) and (c) jointly present the cause of Lena's happiness. For sentence (d), however, two attachments are possible. If it is attached higher up in the tree at the level of sentence (a), as shown in figure 1, the predicted interpretation of the personal pronoun *sie/she* in the last sentence is *Lena*, since the referents introduced in the subordinated segment (b)-(c) are not accessible from (d)'s attachment site. Figure 2 illustrates the other possible structure, where (d) is connected directly to the last segment which makes *the coach* an accessible antecedent. Since it is also the most recent one, the pronoun will preferably resolve to *the coach*.

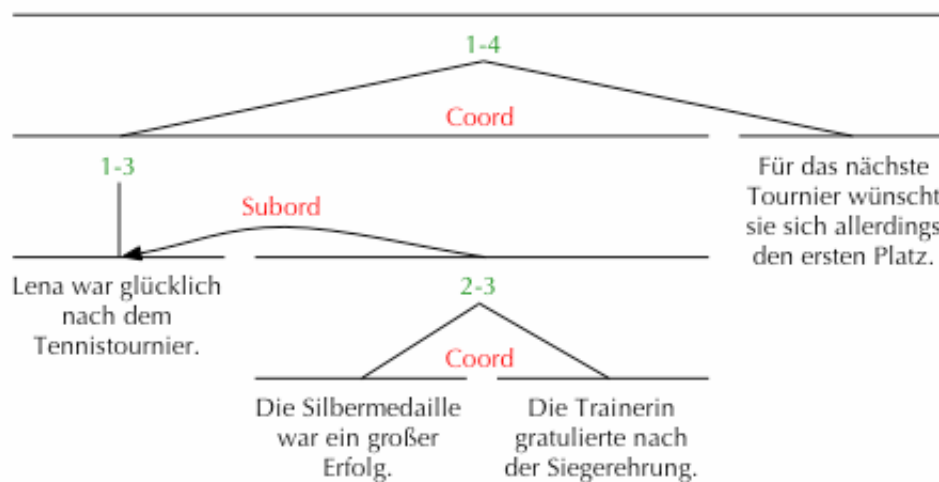


Figure 1: Example (1)/(2). High attachment of the last sentence.

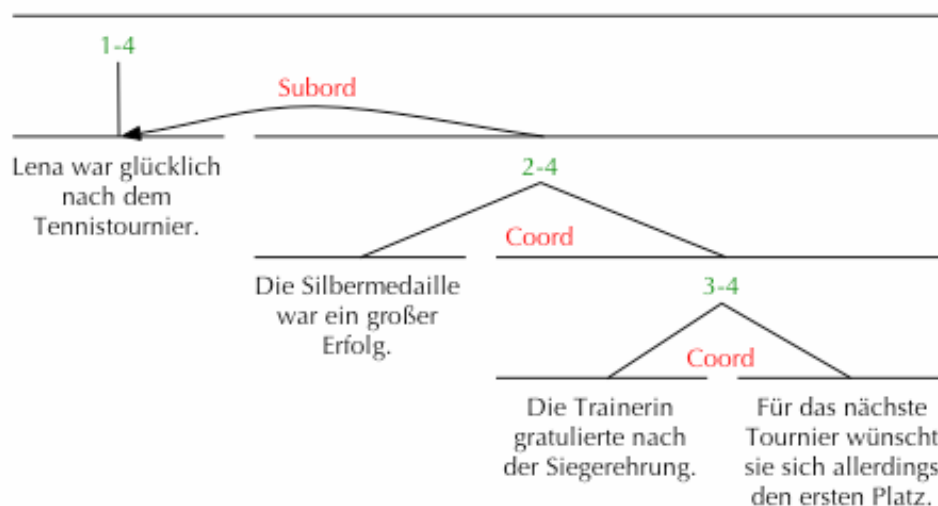


Figure 2: Example (1)/(2). Low attachment of the last sentence.

The generalization illustrated above is captured in one form or another by almost any discourse theory, although the precise formulation may differ depending on the underlying assumptions about the discourse structure and the nature of anaphoric accessibility. One of the most well-known formulations is the *Principle of the Right Frontier* (Polanyi 1988, Webber 1991, Asher 1993), which says that only the discourse units on the "right frontier" of the discourse graph are accessible for any further operations, including search for anaphora antecedents, whereas the right frontier includes the immediate left sister of the current discourse unit plus all the dominating nodes, but not the subordinated nodes. A very similar constraint is proposed by Grosz and Sidner (1986), although it is formulated in more procedural terms. In this approach, discourse referents are organized in a stack of focus spaces; a new focus space is pushed onto the stack when a subordinated discourse unit is opened, and popped off the stack once that unit is closed. Only the referents in the focus space on top of the stack are accessible for anaphoric reference. Further, both the Veins Theory (Cristea et al. 1998) and the Rhetorical Distance Theory (Kibrik 1999) define discourse-structural constraints on anaphora resolution on the basis of the RST tree architecture (cf. Mann and Thompson 1988), although the two approaches differ in their view of the nature of these constraints. Whereas the Veins Theory follows the other proposals mentioned above in assuming that discourse structure imposes *categorical accessibility* constraints on anaphora, i.e. a particular referent is either accessible at a certain point in the discourse or not; Kibrik takes discourse structure to contribute directly to the *gradual* notion of *salience*: under this view, roughly speaking, a referent can be less or more accessible. However, both approaches agree that referents introduced in a discourse segment subordinated to some previous sentence are inaccessible, or at least hard to access from outside that segment.²

It is not the purpose of the current study to compare different theories of discourse anaphora. Instead, we would like to concentrate on the phenomenon illustrated above, which seems to be most uncontroversial among the different approaches. Our interest focuses on the question whether prosody can eliminate the structural ambiguity of (1)/(2) and lead the hearer to the corresponding interpretation of the anaphoric pronoun.

² Subordinating relations between discourse segments roughly translate to RST nucleus-satellite relations, and coordination to multi-nuclear relations. See Chiarcos and Krasavina (2005) and references therein for a systematic comparison of different approaches to anaphoric accessibility.

2.2 Discourse prosody

Numerous studies have shown that the hierarchical structure of spoken discourse is reflected by prosody. The two most important and best researched prosodic means for structuring longer utterances are pitch range and pause duration. Pitch range is a global prosodic parameter of an intonational phrase and defines a subdivision of the total range of fundamental frequency variation of a given speaker. The pitch range can vary in width (e.g. expanded, normal, compressed) and in position relative to the total range (e.g. high, mid, low). It is the reference frame for local tonal events like pitch accents and boundary tones. For instance, a high tone is realized higher in a phrase with expanded pitch range compared to a high tone in a phrase with compressed pitch range. In general, most studies agree that expanded pitch range correlates with the introduction of new discourse topics and sub-topics or with the beginning of a paragraph; compressed pitch range, on the other hand, signals the end of a paragraph or the closing of a (sub-) topic. These results were already obtainable on the basis of a rather simplistic pre-theoretical notion of discourse structure, equating the latter either with the structure of a written text, i.e. the paragraph structure (Lehiste 1975, 1979; Sluijter & Terken 1993; van Donzel 1999) or with a discourse topic model adopted for the specific material of the study (Ayers 1994; Venditti & Swerts 1996; Swerts & Geluykens 1993; Nakajima & Allen 1993). Other studies that based their analyses on more elaborate, theoretically motivated hierarchical notions of discourse structure, such as RST or SDRT, have also shown that the width and position of the pitch range correlate significantly with the depth of embedding of discourse units (den Ouden et al. 2002, Mayer 1999, Möhler & Mayer 2001).

Similar results are reported for the duration of silent pauses. Pauses are longer before units introducing new discourse topics. The shortest pauses appear between intonational phrases dealing with the same topic (Grosz & Hirschberg 1992; Swerts & Geluykens 1993; Hirschberg & Nakatani 1996; Swerts 1997). As with the pitch range, den Ouden et al. (2002) again showed a strong correlation between pause duration and depth of embedding.

For example, these findings suggest that the discourse structures in figures 1 and 2 should have different prosodic realizations. If the last sentence is attached higher in the tree (figure 1) the greatest structural break occurs immediately before it. This break is likely to be associated with a longer pause and a pitch reset, i.e. pitch range of sentence (1c) is compressed, whereas pitch range of sentence (1d) is expanded, as shown in figure 3. By contrast, if sentence (1d) relates directly to the preceding utterance (figure 2), both nodes are embedded deep in the structure, so the pause between them will be shorter, and no pitch reset is expected, cf. figure 4. But given the considerations of anaphoric accessibility discussed in the previous section, these prosodic contrasts should ultimately correlate with the corresponding options of anaphoric pronoun interpretation: longer pause, pitch reset—high attachment in the discourse structure—pronoun *sie/she* is resolved to *Lena*; shorter pause, no pitch reset—low attachment—pronoun *sie/she* is resolved to *die Trainerin/the coach*.

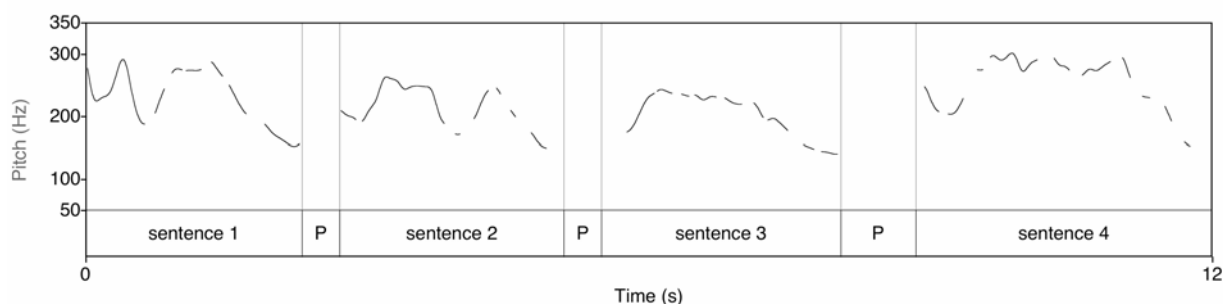


Figure 3: Predicted prosodic realization of (1). High attachment of the last sentence.

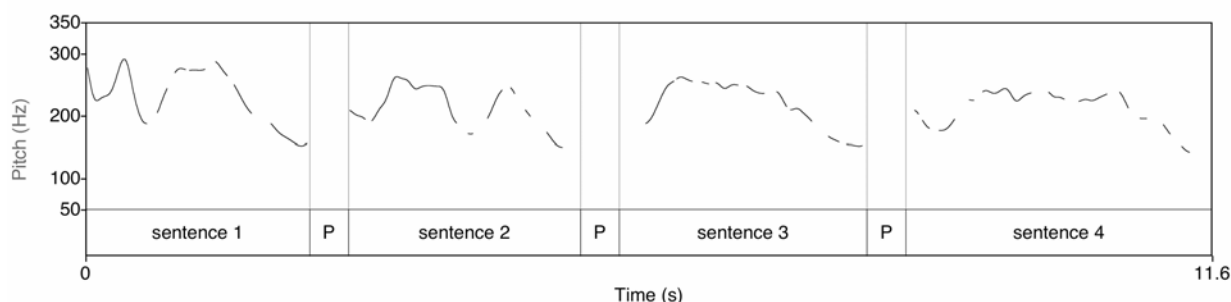


Figure 4: Predicted prosodic realization of (1). Low attachment of the last sentence.

Although the predictions are obvious, it is still largely an open question whether hearers actually use the information encoded by pauses and pitch range to disambiguate anaphoric references,³ which also bears on a more general theoretical issue, whether global prosodic parameters contribute to the *linguistic* interpretation of an utterance and should be represented in the grammar. In general, perception studies on global prosody are relatively few as compared to corresponding production studies (cf. above), and even those available concentrated mainly on more superficial aspects of perception. It has been found that synthesized speech with "paragraph intonation" sounds more natural than without it (Sluijter & Terken 1993), and that the hearers are able to identify discourse structural boundaries of different strength on the basis of prosodic cues (Swerts & Geluykens 1993, 1994; Swerts et al. 1994; van Donzel 1999). As for the impact of global prosodic parameters on the semantic/pragmatic interpretation of linguistic expressions, we are aware of only one study—Silverman (1987: chapter 6)—that addresses the issue. The results of two experiments suggest that pauses and pitch range can resolve ambiguities in quantification domain selection by a universal quantifier,⁴ as well as ambiguities in the scope of indirect speech. However the study was based on very limited material and the experiments were designed in such a way that (we suspect) the subjects could easily guess the hypothesis under investigation.⁵ Thus Silverman's results, though encouraging, call for replication in a methodologically more rigorous setting.

The purpose of the experiment presented in the current paper is thus to complement the previous perception studies of global prosody and test the hypothesis that global prosodic features contribute to the interpretation of linguistic expressions by disambiguating structurally ambiguous discourses. Among all the interpretation phenomena sensitive to discourse structure, we confine our attention to pronominal anaphora, which is a paradigmatic case among various types of anaphora as well as a central phenomenon of discourse interpretation.

3. Method

3.1 Materials

3.1.1 Discourses

For the purpose of this study we constructed 28 discourses like the one in (1). All the discourses comprised 4 sentences, where sentences 2 and 3 were connected to sentence 1 with a subordinating discourse relation—*Elaboration*, *Explanation*, or *Background* (Asher & Lascarides 2003: pp. 459-

³ The hypothesis that the global prosodic organization of discourse could help the hearer keep track of anaphoric references was expressed by van Donzel (1997), but as such, it was not tested in her study.

⁴ Note that it is generally accepted that quantification domain selection by quantifiers, like e.g. *all*, has anaphoric nature. It is natural to expect that discourse constraints on anaphora resolution extend to domain anaphora, as well.

⁵ The experimental materials included, for instance, only two instances of quantification domain ambiguity, and only two instances of indirect speech scope ambiguity. The instructions to the participants stated explicitly that the discourses are ambiguous and that "the way they are uttered" should disambiguate them. The author also does not report using distractors.

471). The intended discourse relation was achieved by a combination of relevant factors, primarily the lexical and compositional semantics of the sentences, as well as world knowledge. Sentence 4 was constructed in such a way as to allow for two more or less equally plausible interpretations: either as part of the elaboration/explanation/background initiated by sentences 2 and 3 (low attachment), or as related directly to sentence 1 (high attachment).

The critical discourse referents R1 and R2, e.g. *Lena* and *the coach* in (1), were introduced in sentence 1 and sentence 3, respectively, and were not mentioned elsewhere until the target sentence 4, which contained an ambiguous pronoun that could refer to R1 or R2.⁶ The referents R1 and R2 in sentences 1 and 3 were realized by proper names or definite descriptions, always constituted the grammatical subject of the sentences and occurred in the pre-verbal position (the German *Vorfeld*). The ambiguous pronoun was the grammatical subject of sentence 4, too,⁷ but it was realized immediately after the finite verb (the first position of the *Mittelfeld*), whereas the preverbal position was occupied by a different constituent, usually an adverbial or another argument of the verb. The post-verbal position for the target pronoun was chosen for two reasons. First, it is a prosodically weak position where the pronoun is most naturally realized without a pitch accent. Consistent deaccenting of the target pronoun in all the items was important, since accent placement on pronouns is known to affect anaphora resolution (see e.g. Venditti et al. 2002). The second reason for placing the pronoun after the verb was that in this way it occurred late enough in the sentence so that the first high pitch accent could be realized before it and the hearer could appreciate the pitch range of the utterance before (s)he interpreted the pronoun.

Additionally, 36 fillers were constructed. The filler discourses also comprised 4 sentences but showed no ambiguity in discourse structure or pronoun resolution. Each discourse (experimental and filler) was completed with a final *who?*-question of the form in (3). In the experimental items, the question was derived from sentence 4, as to reveal the hearer's interpretation of the target pronoun, cf. (3) and (1).

- (3) Wer wünscht sich den ersten Platz?
Who hopes for the first place?

In the filler items, the question could address any of the four sentences, the answer to the question was always some person mentioned in the discourse, which could be realized by a pronoun or a full DP in the sentence from which the question was derived.

3.1.2 Experimental items

All materials were recorded in an anechoic chamber using high quality equipment. The sentences were read by one female speaker in randomized order (i.e. not in the context of the respective discourses), aiming at producing constant pitch range and intensity settings. We then adjusted the pitch range parameters (see below) for each sentence and re-created the original discourses by concatenating the resynthesized sentences with specific pause durations (intervals of zero amplitude). Concerning the 28 experimental discourses, 2 versions of each discourse were created resulting in 56 experimental items. In the low attachment version, pauses were set to standard duration (400 ms) between all sentences. The pitch range of sentences 2 and 3 was set to normal, and pitch range of sentence 4 was compressed. The high attachment version had a lengthened pause (800 ms) between sentence 3 and sentence 4 and standard pause durations between the other sentences. Pitch range was set to normal in sentence 2, compressed in sentence 3, and expanded in sentence 4, i.e. there was a pitch reset between sentence 3 and 4. In both versions, sentence 1 was

⁶ R1 and R2 were singular human referents in all items, hence the target pronoun was always *er* 'he' or *sie* 'she'.

⁷ In this way we wanted to consistently maintain the parallelism in grammatical function between the target pronoun and its both potential antecedents, in order to switch off "parallel function" effects as a possible confounding factor (cf. e.g. Sidner 1983, Kameyama 1986).

always assigned expanded pitch range.

For the 36 fillers, pitch range and pause durations were set according to their discourse structure, which either matched one of the discourse structural patterns of the items, cf. figure (1) and (2), or exhibited a third pattern where the greatest structural break occurred between sentences 2 and 3, so that the discourse was divided into two equal "paragraphs". In such discourses, the 800 ms pause was introduced accordingly between sentences 2 and 3, the pitch range in sentences 1 and 3 was set to expanded, and in sentences 2 and 4 to normal.

The final question was spoken by a male speaker and was added to the sequence 1500 ms after the end of sentence 4 with the original unmanipulated question intonation.

3.1.3 Pitch range manipulation

Pitch range was defined as the range between the highest intonationally relevant high tone (HT) and the lowest relevant low tone (LT) within one phrase (sentence). Relevant tones were labeled manually in the original recordings and corresponded usually to high or low tonal targets of pitch accents. For pitch range manipulations, 3 different ranges were defined: normal, compressed and expanded. We determined the normal pitch range of the female speaker as ranging from 150 Hz (baseline) to 270 Hz (topline). Compression and expansion ratios were calculated from a fully annotated radio news corpus (Rapp 1998). Therefore, the differences to normal range were relatively moderate and less distinctive compared to spontaneous speech:

| | Baseline | Topline |
|------------------|-----------------|----------------|
| Normal range | 150 Hz | 270 Hz |
| Expanded range | 150 Hz | 310 Hz |
| Compressed range | 140 Hz | 250 Hz |

Based on the original HT and LT and the target range values, the pitch contour of each sentence was shifted so that the LT was set to the target baseline and multiplied so that the HT reached the target topline. All signal processing was done using PRAAT (Boersma & Weenink 2005).

3.1.4 Expectations

We expected that listening to the low attachment version with a short pause between sentence 3 and 4 and gradually decreasing pitch range over the whole sequence, hearers would preferably attach sentence 4 inside the subordinated constituent. This would be indicated by a preferred co-reference between the pronoun in sentence 4 and the second referent R2 introduced in sentence 3. Listening to the high attachment version with a long pause and pitch reset between sentence 3 and 4, the hearers would attach sentence 4 outside the subordinated constituent *more frequently than in the first condition*, which would be indicated by a more frequent resolution of the target pronoun to the referent R1 introduced in sentence 1.

3.2 Subjects and procedure

38 subjects participated in the experiment. Two of them were excluded from the analysis because they recognized the ambiguity and refused to give a definite answer to the final questions. Of the remaining 36 subjects, 25 were female, the mean age was 25.1 years, and all were native German speakers.

Two item sets were compiled from a total of 56 experimental items (28 x 2 versions) and 36 filler items. Each set contained only one version of the experimental items and all filler items, resulting in a total of 64 items per set in randomized order (14 experimental items/high attachment version + 14 experimental items/low attachment version + 36 fillers). Half of the subjects were given set A, the other half set B, so each subject heard each discourse exactly once.

The items were presented over loudspeakers. Subjects were asked to listen to the stories and answer the question at the end of each story orally. Neither written transcripts nor lists of possible answers were provided. The answers were immediately classified by the experimenter (R1/high attachment, R2/low attachment, or indefinite) using a computer interface. In addition, all answers were recorded to facilitate re-checking of the classifications.

4. Results

For the analysis the participants' responses were coded on a nominal scale with *1* for R1 (high attachment response) and *-1* for R2 (low attachment response). The data were aggregated within the experimental condition and two tailed paired t-tests were done for all participants (subject analysis: t_1) and items (item analysis: t_2).⁸ Figure 5 shows the results as an average value of the subject analysis (t_1 [35] = 2.992, $p < 0.01$) and the item analysis (t_2 [27] = 4.804, $p < 0.001$). The high attachment response (R1) was significantly more frequent in the high attachment prosody condition (39%) than in the low attachment prosody condition (28%). The predominance of the low attachment response (R2) in both conditions can be explained by the general preference for the nearest antecedent (referent R2), in our experiment mentioned in sentence 3.

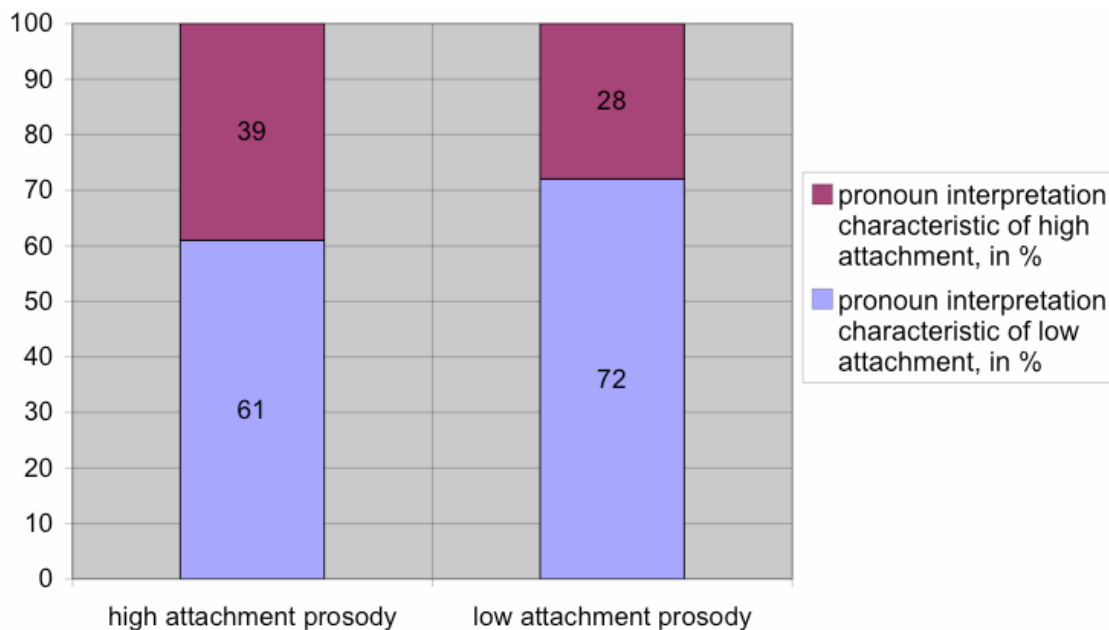


Figure 5: Average percentages for pronoun interpretation in high/low attachment prosody condition.

5. Discussion

The results of our experiment corroborate our main research hypothesis that global prosodic parameters such as pitch range and pause duration influence resolution of anaphoric pronouns. Moreover, the direction of the observed effect is clearly in accordance with the predictions of the existing theories of discourse anaphora and the current state of research on discourse prosody.

⁸ t_1 = aggregate data within experimental condition across items; t_2 = aggregate data within experimental condition across participants.

"High attachment prosody" favors "high attachment anaphora resolution" to a greater extent than low attachment prosody does. This result strongly suggests that indeed, the relationship between global prosody and the choice of the antecedent of an anaphoric pronoun is mediated by the choice of attachment site of an utterance in the discourse structure.

The current results are also consistent with those obtained by Silverman (1987), although the measured difference between the conditions turned out to be considerably more modest. In our experiment, the pronoun interpretation characteristic of high attachment was only 11% more frequent in the high attachment condition than in the low attachment condition, whereas Silverman measured differences of 40% to 68%. Of course, this discrepancy could be due to the design of our experiment, where the underlying hypothesis was arguably better concealed from the subjects than in Silverman's design. Nevertheless, apart from the strength of the observed effect, the presented experiment can be viewed as a replication of Silverman's result in a methodologically more rigorous setting. Furthermore, the current experiment complements the study conducted by Silverman by testing the hypothesis on a different discourse structure sensitive phenomenon (pronominal anaphora) and in a different language (German). The two studies taken together present substantial evidence for a *linguistically relevant* interaction between global prosodic parameters and discourse structure.

A few words should be said on the apparent overall preference for low attachment of the target sentence in both conditions (61% and 72%), cf. figure 5. This strong preference could be due to (a combination of) at least three reasons. First of all, note that the antecedent that we considered to be the indicator of low attachment (R2) always occurred in the sentence immediately preceding the sentence with the target pronoun, whereas the "high attachment antecedent" (R1) was mentioned 3 sentences earlier. Intuitively, it is not surprising that the pronominalization of a referent mentioned so far away is generally dispreferred. Thus listening to the high attachment versions of the experimental items, the subjects could have followed some kind of repair strategy in interpreting an overall suboptimal discourse. Although the importance of linear distance for anaphora resolution is uncontroversial, the issue of possible interaction between linear and structural constraints has received comparatively little attention from researchers so far. Most theoretical studies (cf. section 2.1) have been exclusively concerned with structural accessibility constraints (Grosz & Sidner 1986, Polanyi 1988, Asher 1993, Asher & Lascarides 2003). From the point of view of these theories, it would be completely unexpected if the interpreters attached the target sentence high in the discourse structure, but nevertheless resolved the anaphoric pronoun to the antecedent in the embedded segment, no matter if it is linearly the closest.⁹ The same holds for the Veins Theory (cf. esp. Cristea et al. 1999). Although the concept of *vein* in a way incorporates both structural and linear constraints on anaphora, one could say that it ranks the structural constraints higher than the linear ones, as the inaccessible segments are not part of the vein, and hence are not considered when calculating the linear distance. In this sense, the vein-based linear distance from the pronoun to R1 in the high attachment version of our items is 1 (since the intervening inaccessible sentences do not count), just as it is between the pronoun and R2 in the low attachment version. So the general preference for R2 with both low and high attachment prosody remains unexplained. A possible answer to this problem could be an approach along the lines of Kibrik (1999) where structural (rhetorical) and linear distance are considered as perhaps unequally strong, but independent factors.¹⁰ In this approach, it is a matter of weight assignment to the individual factors whether linear distance can sometimes "overrule" the structural constraints. Thus the observed overall preference for the closest antecedent in our experiment could be viewed as an indication that linear

⁹ However, sporadically one hears suggestions that the accessibility constraints should be relaxed to make the referents introduced in the last sentence accessible under any circumstances, even after a discourse pop (cf. e.g. Asher & Vieu 2005, fn. 9). The results of our experiment, in particular the preference for the most recent antecedent, support this idea.

¹⁰ It should be noted that Kibrik (1999) is primarily a theory of referential choice, i.e. choice of an anaphoric device by the speaker in the process of discourse production – rather than anaphora resolution, which is part of the discourse interpretation process.

distance should be assigned a rather high weight. Nevertheless, this is not intended to suggest that the results of our experiment present conclusive evidence for or against the above-mentioned approaches, since there are alternative explanations to the observed preference for R2.

The second factor that could have facilitated the choice of R2 in the high attachment condition is that prosody does not completely disambiguate between the discourse structures with high and low attachment. In principle, it is not excluded that a low attachment structure like that in figure 2 is realized with a strong prosodic break between sentences 3 and 4. This is possible especially if sentence 4 starts a relatively long discourse segment (a subtopic) which is too heavy to be realized as part of the same "prosodic paragraph", even though it might be attached low in the discourse structure (cf. Möhler & Mayer 2001 on complex subtopics). Although sentence 4 in our experimental discourses was always the last, it is possible that under the pressure of the linear recency of R2 (cf. above), the subjects preferred to accommodate an unspoken continuation with a subtopic and attach sentence 4 low despite the strong prosodic break before it.

Finally, the resolution of the target pronoun to the second antecedent in the high attachment condition could have been made possible by the fact that the antecedent was always phrased by a proper name or a definite DP. According to the standard assumption, the referents of proper names are always introduced *globally* in the discourse representation structure (DRS), and the existence presuppositions of definite descriptions prefer *global* accommodation (see e.g. Bos 2003, Beaver & Zeevat, to appear, and references therein). This implies that the referents introduced by such expressions should be accessible from anywhere in the discourse structure, even if the literal mention takes place in an inaccessible segment. Thus it is possible that both R1 and R2 were accessible both under low and under high attachment of the target sentence, whereas R2 was presumably preferred due to recency. However, if global accommodation were the whole story, we should not have found any difference at all between the high and the low attachment condition in our experiment. The fact that there was a significant difference could have two explanations. Either proper names and/or definite descriptions are in fact able to introduce their referents in the *local* DRS associated with the embedded discourse segment, at least as an option (contrary to the standard view, but see Geurts 1997 for related observations);¹¹ or structural constraints on anaphora resolution such as the Right Frontier Principle are after all not a matter of accessibility (but e.g. salience), or at least not accessibility in the standard sense that would result from global presupposition accommodation.

It goes beyond the scope of the current paper to resolve these issues. Surely, the results of the presented experiment provide material for further theoretical discussion. However, one issue seems clear: global prosodic features of short spoken monologues affect anaphoric links within those monologues in the way predicted by the Right Frontier Principle and the current models of global prosodic structure.

6. Conclusions

Prosodic features like pause duration and pitch range variation affect anaphora resolution by disambiguating structurally ambiguous discourses. We substantiated the hypothesis that not only local intonational events like pitch accents or boundary tones but also global prosodic characteristics of utterances are accessible for linguistic interpretation. Furthermore, we identified another surface property of an utterance that plays a role in anaphora resolution. The possible influence of global prosodic features should be taken into account by any experimental study of

¹¹ Although it is accepted that presuppositions of definite descriptions allow for local accommodation, this is usually thought of as a marked option that is only available if a number of constraints are met (see Beaver & Zeevat, to appear). As far as we can judge, these constraints were not met in our experimental discourses, i.e. nothing was there to prevent the preferred global accommodation. Thus, at least from a purely theoretical point of view, it would be equally unexpected if the observed effect were solely due to the items where R2 was expressed by a definite description.

anaphora resolution in spoken language. For instance, studies on the role of pitch accents in the interpretation of pronouns, such as Venditti et al. (2002), do not report controlling for global prosodic features, which could be a confounding factor in their setting. Finally, we want to point out that the presented experimental design could be used as a methodological framework for the study of a whole range of discourse-prosody interface phenomena.

References

- ASHER N. (1993), *Reference to Abstract Objects in Discourse*, Dordrecht, Kluwer.
- ASHER N. & LASCARIDES A. (2003), *Logics of Conversation*, Cambridge University Press.
- ASHER N. & VIEU L. (2005), Subordinating and coordinating discourse relations, *Lingua* 115-4, 591-610.
- AYERS G.M. (1994), Discourse functions of pitch range in spontaneous and read speech, *Ohio State University Working Papers in Linguistics* 44, 1-49.
- BEAVER D. & ZEEVAT H. (to appear), Accommodation, in Ramchand G. & Reiss C. (eds.), *Oxford Handbook of Linguistic Interfaces*, Oxford University Press.
- BOERSMA P. & WEENINK D. (2005), Praat: doing phonetics by computer. University of Amsterdam.
- BOS J. (2003), Implementing the binding and accommodation theory for anaphora resolution and presupposition satisfaction, *Computational Linguistics* 29-2, 179-210.
- CHIARCOS C. & KRASAVINA O. (2005), Rhetorical distance revisited—a parametrized approach, in Sassen et al. (2005), 63-70.
- CRISTEA D., IDE N. & ROMARY L. (1998), Veins theory: A model of global discourse cohesion and coherence, in *Proceedings of COLING-ACL 1998*, 281-285.
- CRISTEA D., IDE N., MARCU D. & TABLAN V. (1999). Discourse structure and co-reference: An empirical study, in *Proceedings of the ACL Workshop on The Relationship Between Discourse/Dialogue Structure and Reference*, USA, University of Maryland.
- DEN OUDEN H., NOORDMAN L. & TERKEN J. (2002), The prosodic realization of organizational features of texts, in *Proceedings of Speech Prosody 2002*, Aix-en-Provence, 543-546.
- GEURTS B. (1997), Good news about the description theory of names, *Journal of Semantics* 14, 319-348.
- GROSZ B. & HIRSCHBERG J. (1992), Some intonational characteristics of discourse structure, in *Proceedings of the 2nd International Conference on Spoken Language Processing*, Banff, Canada, 429-432.
- GROSZ B.J. & SIDNER C.L. (1986), Attention, intentions and the structure of discourse, *Computational Linguistics* 12-3, 175-204.
- HIRSCHBERG, J. & NAKATANI, C.H. (1996), A prosodic analysis of discourse segments in direction-giving monologues, in *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, 286-293.

- KAMEYAMA, M. (1986), A property-sharing constraint in Centering, in *Proceedings of the 24th Annual Meeting of the Association for Computational Linguistics*, New York, 200-206.
- KIBRIK, A.A. (1999), Reference and working memory: Cognitive inferences from discourse observations, in Van Hoek, K., Kibrik, A.A., & Noordman, L.G.M. (eds.), *Discourse Studies in Cognitive Linguistics*, Amsterdam, Benjamins, 29-52.
- LEHISTE, I. (1975), The phonetic structure of paragraphs, in Cohen, A. & Nooteboom, S.G. (eds.), *Structure and Process in Speech Perception*, Springer, 195-203.
- LEHISTE, I. (1979), Perception of sentence and paragraph boundaries, in Lindblom, B. & Öhman, S. (eds.), *Frontiers of Speech Research*, London, Academic Press, 191-201.
- MANN, W.C. & THOMPSON, S. (1988), Rhetorical Structure Theory: Toward a functional theory of text organization, *Text* 8-3, 243-281.
- MAYER, J. (1999), Prosodische Merkmale der Diskursrelationen, *Linguistische Berichte* 177, 65-86.
- MÖHLER, G. & MAYER, J. (2001), A discourse model for pitch-range control, in *Proceedings of the 4th ISCA Tutorial and Research Workshop on Speech Synthesis*, Perthshire, Scotland.
- NAKAJIMA, S. and ALLEN, J.F. (1993), A study on prosody and discourse structure in cooperative dialogues, *Phonetica* 50, 197-210.
- POLANYI, L. (1988), A formal model of the structure of discourse, *Journal of Pragmatics* 12, 601-638.
- RAPP, S. (1998), Automatisierte Erstellung von Korpora für die Prosodieforschung. Dissertation, University of Stuttgart.
- SASSEN, C., BENZ, A., & KÜHNLEIN, P., eds. (2005), *Proceedings of Constraints in Discourse*, June 2-5, 2005, Dortmund.
- SIDNER, C. (1983), Focusing in the comprehension of definite anaphora, in Brady, M. & Berwick, R. (eds.), *Computational Models of Discourse*, MIT Press, 267-330.
- SILVERMAN, K.E.A. (1987), *The Structure and Processing of Fundamental Frequency Contours*, PhD thesis, University of Cambridge.
- SLUIJTER, A.M.C. & TERKEN, J.M.B. (1993), Beyond sentence prosody: Paragraph intonation in Dutch, *Phonetica* 50, 180-188.
- SWERTS, M. (1997), Prosodic features at discourse boundaries of different strength, *Journal of the Acoustical Society of America* 101, 514-521.
- SWERTS, M., BOUWHUIS, D.G., & COLLIER, R. (1994), Melodic cues to the perceived "finality" of utterances, *Journal of the Acoustical Society of America* 96-4, 2064-2075.
- SWERTS, M. & GELUYKENS, R. (1993), The prosody of information units in spontaneous monologue, *Phonetica* 50, 189-196.
- SWERTS, M. & GELUYKENS, R. (1994), Prosody as a marker of information flow in spoken discourse, *Language and Speech* 37-1, 21-43.

VAN DONZEL, M. (1997), Perception of discourse boundaries and prominence in spontaneous Dutch speech, *Lund University, Dept. of Linguistics Working Papers* 46, 5-23.

VAN DONZEL, M. (1999), *Prosodic Aspects of Information Structure in Discourse*, PhD thesis, The Hague University.

VENDITTI, J. and SWERTS, M. (1996), Prosodic cues to discourse structure in Japanese, in *Proceedings of the International Conference on Spoken Language Processing*, Philadelphia, 725-728.

VENDITTI, J.J., STONE, M., NANDA, P. & TEPPER, P. (2002), Discourse constraints on the interpretation of nuclear-accented pronouns, in *Proceedings of the 2002 International Conference on Speech Prosody*, Aix-en-Provence, France.

WEBBER, B.L. (1991), Structure and ostension in the interpretation of discourse deixis, *Natural Language and Cognitive Processes* 2-6, 107-135.