

Discourse Structure and Information Structure: Interfaces and Prosodic Realization

Ekaterina Jasinskaja, Jörg Mayer and David Schlangen

Universität Potsdam

In this paper we review the current state of research on the issue of discourse structure (DS)/information structure (IS) interface. This field has received a lot of attention from discourse semanticists and pragmatists, and has made substantial progress in recent years. In this paper we summarize the relevant studies. In addition, we look at the issue of DS/IS-interaction at a different level—that of phonetics. It is known that both information structure and discourse structure can be realized prosodically, but the issue of phonetic interaction between the prosodic devices they employ has hardly ever been discussed in this context. We think that a proper consideration of this aspect of DS/IS-interaction would enrich our understanding of the phenomenon, and hence we formulate some related research-programmatic positions.

Keywords: discourse structure, information structure, prosody

1 Introduction

In this paper we review the current state of research on the issue of discourse structure (DS)/information structure (IS) interface. Although a recent special issue of the *Journal of Logic, Language, and Information* (Kruijff-Korbayová and Steedman, 2003) has addressed the same topic, the rapid development in the field calls, in our opinion, for another update. Progress has been made both in the study of specific DS/IS interface phenomena (e.g. Buring, 2003; Umbach, 2004; Hendriks, 2004; Zeevat, 2004) and in the development and formalization of the underlying theoretical concepts (e.g. van Rooy, 2003; Ginzburg and Cooper, 2004). These as well as some previous studies will be summarized and discussed in a systematic way.

Interdisciplinary Studies on Information Structure 1 (2004): 151–206

Ishihara, S., M. Schmitz and A. Schwarz (eds.):

©2004 Ekaterina Jasinskaja, Jörg Mayer and David Schlangen

In addition, we look at the issue of DS/IS-interaction at a different level—that of phonetics. It is known that both information structure and discourse structure can be realized prosodically, but the issue of phonetic interaction between the prosodic devices they employ has hardly ever been discussed in this context. We think that a proper consideration of this aspect of DS/IS-interaction would enrich our understanding of the phenomenon.

Naturally, new findings raise new questions. The ultimate purpose of this paper is to articulate some research areas and formulate hypotheses that should be investigated in order to supply the missing parts of the overall picture.

The paper has three major parts. In Section 2, we present what we take the notions of discourse structure and information structure to mean in isolation, and how each of these structures impacts prosody. In Section 3, we give an overview of the literature that investigates what is traditionally included in the notion of DS/IS-interface. Finally, we present some ideas on the interaction between DS and IS at the level of phonetics in Section 4.

2 General Remarks

Before we take up the issue how discourse structure and information structure interact, it is necessary to say a few words about what we take these two structures to be in isolation, and how they manifest themselves in speech. We will concentrate on their prosodic manifestation.

2.1 What is discourse structure?

Morphology and syntax seek to characterize the well-formedness of words and sentences; similarly, work on discourse structure attempts to describe the coherence of units larger than the sentence, up to the level of entire texts. This enterprise aims to discover and investigate of elementary discourse units, groups

of units that form larger units, as well as the relations between them, which constitute the hierarchical discourse structure.

Hierarchical discourse structure is motivated largely by three kinds of linguistic phenomena that bear on discourse coherence. First, just like syntactic structure constrains the anaphoric relations within a sentence (cf. binding theory), discourse structure affects the accessibility of antecedents for discourse anaphors—anaphoric expressions not captured by binding theory. Second, it is generally accepted that the meaning of a discourse is more than a sum of the meanings of its sentences. In addition, there are various semantic relations (e.g. causal or temporal relations) that hold between the meanings of individual sentences and groups sentences, and which speakers of a language appear to be able to successfully infer from a text, even when they are not overtly signaled e.g. by discourse connectives. Discourse structure provides the blocks to fill the arguments of such semantic relations, whereas in some theories the relations themselves are considered an essential part of the discourse structure, as well. The third kind of linguistic motivation for discourse structure comes from prosody—and we will discuss this in some more detail in Section 2.2. In brief, the crucial observation is that speakers control a number of global prosodic parameters of speech, such as pitch range, speech rate and pause duration, in a systematic way, and that the way they do it intuitively seems to serve the purpose of grouping single utterances into larger chunks.

The current approaches to discourse structure are often classified into two major trends—the *informational* and the *intentional* approaches.¹ Informational approaches attempt to characterize discourse coherence in terms of semantic relations between the information conveyed by successive units. The approaches differ in what role they grant to discourse structure in the overall architecture of language. According to one view, discourse structure is part of the *conceptual structure* (i.e. semantics) of a text. This view is represented by such theories as

¹ See Zaenen et al. (2001) and Asher and Lascarides (2003) for recent surveys.

the *Rhetorical Structure Theory* (RST, Mann and Thompson, 1988), *Discourse Representation Theory* (DRT, Kamp and Reyle, 1993) as well as *Segmented Discourse Representation Theory* (SDRT, Asher, 1993; Asher and Lascarides, 2003), which combines elements of RST and DRT. In particular, RST and SDRT assume large ontologies of semantic discourse relations and the process of computing discourse coherence boils down largely to linking all utterances of a text to one another with these relations in a sensible way. The other group of informational theories treat discourse structure and discourse semantics along the same lines as sentence structure and semantics. That is, discourse structure is allocated at the level of “discourse *syntax*”, coherence is treated on a par with sentence-level well-formedness and discourse processing is parsing. The most prominent theories that implement this approach are the *Linguistic Discourse Model* (LDM, Polanyi, 1988) and the *Discourse Tree Adjoining Grammar* (D-TAG, Gardent, 1997; Webber et al., 2003). What unites all the informational approaches, is that they reject using theoretical constructs, such as the speakers’ intentions, that reach beyond the text’s “syntax” and semantics.

By contrast, the *intentional* approach emphasizes the pervasive role of the speakers’ *plans* and *intentions* for discourse coherence. Within this approach, the hierarchy of discourse segments results from hierarchically organized intentions, or *discourse goals* (e.g. Grosz and Sidner, 1986). The idea is that each discourse segment fulfills a goal (conveying certain information, urging the hearer to perform a certain action), and the structural relations between segments reflect the relations between goals. Here, coherence correlates with the efficiency with which a discourse serves the goals of the communication participants. Within the intentional trend, one group of studies will be discussed in particular detail, cf. Section 3.2. These studies represent discourse goals as *Questions under Discussion* (QUD), which the speakers try to answer cooperatively (e.g. Klein and von Stutterheim, 1987; van Kuppevelt, 1995a; Ginzburg, 1996a; Roberts, 1996; Büring, 2003). This framework is particularly interesting

in the current context, since it pays the most attention to information-structural phenomena.

In addition to the major philosophical separation into information and intention orientations, approaches to discourse differ in the richness of the postulated discourse structures, which correlates only partly with the attitude towards intentions. First, the minimum of what a discourse structure is assumed to incorporate is *discourse constituency*, i.e. the information that some discourse segments belong closer together than others, forming larger segments. All the theories mentioned above assume some kind of discourse constituency.

Second, most theories make a further distinction between *subordination* and *coordination* of discourse segments.² This distinction affects the accessibility of referents for anaphora in a crucial way (Grosz and Sidner, 1986; Asher and Vieu, 2003). In place of subordination vs. coordination, RST makes a distinction between *mononuclear* and *multinuclear* relations. Although the original motivation for this opposition was rather different (nuclearity is supposed to reflect the communicative weight of an utterance: in a mononuclear relation one sentence is more central and the other one is more peripheral with regard to the message of the text; in a multinuclear relation each part has equal communicative weight), a number of recent studies have shown that the distinction has impact on anaphora resolution (Cristea et al., 1998) and the generation of anaphoric pronouns (Grüning and Kibrik, 2002). In terms of their influence on anaphora, mononuclear relations are roughly comparable to subordination, and multinuclear relations to coordination. The only theories that do not or not systematically make the distinction between subordination and coordination are the ones in the QUD trend, which could be simply due to the fact that the issues of anaphora resolution were not in the center of attention within this frame-

² In classical DRT, there is strictly speaking no counterpart to coordination: subordination corresponds roughly to the embedding of DRSs, whereas what is viewed as coordinated discourse segments in other theories, would map onto expressions that belong to the same DRS in DRT.

work. However, the few speculations by the proponents of this approach that are found in the literature suggest the general intention to reconstruct the subordination/coordination opposition from the relations between the underlying discourse goals (questions), rather than postulating two kinds of relations between utterances, cf. pp. 184–185.

Finally, some theories assume much richer ontologies of discourse relations than just coordination vs. subordination. These ontologies make further distinctions between relations according to their semantics or pragmatics. E.g. subordinating relations include: *Elaboration*, *Explanation*, *Restatement*, *Comment* etc.; the coordinating relations typically include *Contrast*, *Narration* (or *Sequence*), *Parallel*, *List* etc. For instance if an *Explanation* relation holds between utterances u_1 and u_2 (*Explanation*(u_1, u_2)), the utterance u_2 explains why the event introduced in u_1 happened, e.g. by mentioning the causes of that event. Whereas if u_1 and u_2 are connected by *Contrast*, the speaker intends to compare u_1 and u_2 with respect to their similarities and differences. The approaches that adhere to this rich notion of discourse structure are, for instance, RST, SDRT, LDM and D-TAG. Other proposals in discourse theory reject the idea of encoding such semantic relations as ontological primitives of discourse structure, and either view them as epiphenomenal wrt. speakers' intentions, or acknowledge their existence only to the extent to which they are signaled by overt cues, or are inferable from lexical and world knowledge associated with individual sentences.

We are not going to take sides in this discussion. In this survey we would like to include studies that deal with the interaction of information structure with any of the aspects of discourse structure mentioned above: discourse constituency, coordination vs. subordination, specific discourse relations, as well as hierarchically organized discourse goals. However, it should be noted that we are primarily concerned with one-speaker discourse, i.e. structural relations between utterances in a monologue or within a single dialogue turn. Therefore we have not mentioned turn taking as a further parameter that imposes struc-

ture on discourse. Of course, we will not completely avoid discussing dialogue issues, especially in connection with the question-answer relationship and various question-question relations, which play an important role both in discourse structuring (cf. Section 3.2) and information structural issues. However, we do not pretend to cover the full range of discourse structural, information structural or prosodic phenomena in dialogue. We will touch on these issues only to the extent that they help us understand how DS and IS interact in monologue.

2.2 How is discourse structure realized prosodically?

Recent studies regarding the correlation between discourse structure and prosody differ with respect to the applied discourse model. Three types of approaches can be distinguished: text-oriented approaches, intuitive discourse analysis, and theoretically motivated discourse structures. Studies of the first type (Lehiste, 1975, 1979; Sluijter and Terken, 1993; van Donzel, 1999) equate the structure of written text with discourse structure and concentrate the prosodic analysis on paragraph boundaries in read speech ('paragraph intonation'). Studies of the second type (Ayers, 1994; Venditti and Swerts, 1996; Swerts and Geluykens, 1993; Nakajima and Allen, 1993) are based on discourse or topic models which are intuitively adopted to the specific material of the study. These models consist of very simplified sets of discourse units and relations which are not formally defined or derived from established discourse theories. The third approach, which we consider as the most promising, is to apply one of the independently developed discourse theories and to examine the correlation between the various theoretically identified concepts and the prosodic features of spoken discourse. Most studies of the third type use the intentional approach by Grosz and Sidner (1986), cf. Grosz and Hirschberg (1992); Passonneau and Litman (1993); Hirschberg and Nakatani (1996). But other theories are applied as well in recent studies, namely RST (den Ouden et al., 2002) and SDRT (Mayer, 1999;

Möhler and Mayer, 2001). This third approach allows a far more fine-grained analysis of discourse prosody, increases result comparability, and possibly leads to a generalized model of the interface between discourse structure and prosody.

After these basic remarks regarding the discourse theoretical background of the relevant literature, we will now summarize the results of the above mentioned studies. The two most important prosodic means for structuring longer utterances, which are reported in numerous studies, are pitch range and pause duration. The pitch range is a property 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, respectively. Compressed pitch range, on the other hand, signals the end of a paragraph or the closing of a (sub-) topic. Furthermore, some studies assuming hierarchical discourse structures showed that width and position of the pitch range correlate significantly with the depth of embedding of discourse units (Ayers, 1994; den Ouden et al., 2002; Mayer, 1999; Möhler and 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 and Hirschberg, 1992; Swerts and Geluykens, 1993; Hirschberg and 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. However, pitch range (Ayers, 1994) and, even more, pause duration as prosodic correlates of discourse structure depend undoubtedly on speaking style. The clearest results

are obtained from read speech using professional speakers. In a study by Gustafson-Capková and Megyesi (2002), this group yielded results where every silent pause of considerable length was correlated with a topic change, while in the spontaneously speaking group only 34% of the pauses correlated with a topic change.

Apart from pitch range and pause duration, two additional prosodic parameters are considered as relevant for the organization of spoken discourse—speaking rate and intensity variation—but so far, with less convincing results. Concerning speaking rate, Hirschberg and Nakatani (1996) showed that topic-final phrases are produced faster as compared with phrases within the same topic. In contrast, a decrease of speaking rate in phrases preceding a topic change was reported in the study by Smith et al. (2002). A possible explanation for the diverging results could be speaking style again: Hirschberg & Nakatani used spontaneous speech while Smith et al. used read speech. Den Ouden et al. (2002) didn't find any connection between speaking rate and topic structure or depth of embedding, but they found a strong correlation between speaking rate and the nuclearity of discourse segments. Nuclear segments, i.e. segments which are more important concerning the overall coherence of the discourse than others, were realized more slowly. Hirschberg and Nakatani (1996) also reported decreased intensity for topic-final phrases as compared to non-final phrases. These results were confirmed in a study by Herman (2000).

At the end of this section on the prosodic realization of discourse structure we would like to point out that surprisingly little is known about the perceptual relevance of discourse prosodic features. Are these features evaluated by the hearer and if they are evaluated, how are they evaluated? Do we for example perceive pitch range variations gradually or categorically? Is it possible to resolve ambiguous discourse structures with the aid of prosody alone? Are certain prosodic features more important than others? What we know is that synthesized speech with paragraph intonation sounds more natural than without

it (Sluijter and Terken, 1993) and that hearers make use of melody and pauses to identify major discourse units (Swerts and Geluykens, 1994). We assume that not only the research on production aspects of discourse prosody but also the research on perceptual aspects will profit from the integration of formal discourse semantic models and experimental phonetic research as proposed in this paper and in the studies by den Ouden et al. (2002) and Möhler and Mayer (2001).

2.3 What is information structure?

In this paper, we adopt a common view that information structure is constituted by a set of features such as $[\pm F]$, focus; $[\pm T]$, topic; $[\pm CF]$, contrastive focus; $[\pm CT]$, contrastive topic, etc., defined for all the syntactic constituents of a sentence. The distribution of these features in a syntactic tree affects, on the one hand, the prosodic realization of the sentence (e.g. Selkirk, 1995).³ On the other hand, these features have their specific semantic/pragmatic interpretations, which affect the presuppositions and sometimes the truth conditions of the sentence, and most crucially, constrain the set of contexts in which the utterance is felicitous (Rooth, 1985, 1992; Schwarzschild, 1999; Büring, 1997). Below, we briefly introduce three most well-known contextual effects of focus ($[\pm F]$): felicity of question-answer pairs (Section 2.3.1); felicity and truth conditions of utterances that involve alternatives (Section 2.3.2); and contextual newness vs. givenness (Section 2.3.3). These IS concepts and phenomena will be relevant for our discussion of IS/DS-interface in Section 3.

But before concentrating on these issues, some reservations have to be made. First of all, of course, the approach to information structure we adopt here is only one of a great many proposed in the literature. However, a comprehensive review goes beyond the scope of this paper. The reader is referred to van Kup-

³ How exactly information structure is realized prosodically, will be recapitulated in more detail in the next section (2.4).

pevelt (1999) and Kruijff-Korbayová and Steedman (2003) for recent surveys and further references.

Second, we restrict our discussion to *sentential* information structure, i.e. focus and topic features defined and interpreted not higher than at the sentence level. Further, we will remain open as to whether there is a single information-structural partition of a sentence, or whether foci and topics can be embedded in other foci and topics. It should be noted though that the issue is not at all trivial and has been discussed in the literature, cf. Partee (1996); Komogata (2003). In the current context it is particularly interesting to note that, at least in some languages, we find cases that suggest that foci can be interpreted above the sentence level, i.e. one full sentence (a discourse) may serve as focus with respect to another sentence, which is then completely deaccented, cf. (Kodzasov, 1996). This opens a completely different perspective on possible DS/IS interactions—topic and focus features taking scope over discourse constituents. However, we will not go further into this issue in this paper.

Finally, we are primarily interested in information-structural categories motivated by prosodic phenomena. It is well known that also syntax, morphology and certain lexical items (particles) are sensitive to information structure, and that frequently, the IS categories motivated by morphology and syntax and IS categories motivated by prosody are not the same (Vallduví and Vilkuna, 1998). Therefore we restrict our discussion to the latter.

2.3.1 Focus in question-answer pairs

One of the contexts most universally acknowledged as a diagnostic for focus are question-answer pairs. The constituent of the answer that corresponds to the *wh*-word of the question should be focused (F-marked), and hence bear a nuclear accent (indicated by small caps in the examples below). This makes (1-a)–(2-a) and (1-b)–(2-b) felicitous question-answer pairs, whereas (1-a)–(2-b) and (1-b)–

(2-a) are infelicitous.

- (1) a. Who did Mary vote for?
 b. Who voted for John?
- (2) a. Mary voted for [JOHN]_F.
 b. [MARY]_F voted for John.

This simple example illustrates the contextual function of focus in answers to questions. Some linguists propose to reduce the pragmatics of focus to this function alone. They postulate that every sentence answers a question, but the question may be implicit, whereas further constraints at the level of discourse structure regulate which implicit questions are admissible in which contexts. Basically, this approach postpones the analysis of most focus-related contextual phenomena to the level of discourse structure. This will be discussed in some detail in Section 3.2. By contrast, other approaches attempt to capture a broader range of the contextual effects of focus at the level of IS by giving more general definitions of focus pragmatics, cf. below.

2.3.2 Focus and alternatives

the well-known approach (cf. Rooth, 1985, 1992) sees the pragmatic function of focus in highlighting information that is contrary to some *alternative(s)*, anaphorically recoverable from the context. That's why B is a felicitous denial of A in (3), whereas B' is not: *voted for Bill* is a legitimate focus alternative for the VP *voted for* [JOHN]_F, but not for [VOTED]_F for *John*.

- (3) A: Mary voted for Bill.
 B: No, she voted for [JOHN]_F.
 B': # No, she [VOTED]_F for John.

The interpretation of focus in terms of alternatives proved particularly fruitful as a way to account for the semantics of sentences with so-called focus-sensitive particles, such as *only*. Roughly speaking, *only* conveys the idea that the focus-alternatives of the expression in its scope do not hold. For instance, (4-a) implies that Mary did not vote for Bill, or George; whereas (4-b) says that Mary did not do anything else for John, e.g. she did not campaign for him, but she could have voted for other people, as well.

- (4) a. Mary only voted for [JOHN]_F.
b. Mary only [VOTED]_F for John.

The alternative-based approach also gives an adequate account of the effect of focus in question-answer pairs, as well as a whole range of other contexts (cf. Rooth, 1992).

2.3.3 Givenness

Finally, we would like to mention the proposal by Roger Schwarzschild (1999), according to which, a specific pragmatic function is carried by the *unfocused* material, i.e. the [−F]-constituents, rather than [+F], as in the approaches discussed above. This is motivated by the observation that [−F]-constituents can be uniformly interpreted as *given*, or ‘anaphorically recoverable’, whereas [+F] needs, according to Schwarzschild, at least three distinct pragmatic definitions. Two of them were presented in the previous sections: it is (a) focused material as ‘replacing the *wh*-element in a presupposed question’ (cf. 2.3.1); and (b) focused material as ‘being contrary to some predicted or stated alternative’ (cf. 2.3.2). The third definition of [+F] relates it to ‘textually and situationally non-derivable information’, such as *make* in B’s answer in (5).

- (5) From (Schwarzschild, 1999, p. 142):

A: Why don't you have some French TOAST?

B: I've forgotten to MAKE French toast.

According to Schwarzschild (1999) this multitude of (unrelated) definitions is due to a redundancy in the conceptualization: $[-F]$ (*given*) and $[+F]$ (*new*) really are complements, and so only one notion is needed in the theory. His solution is thus to make the notion of *givenness* elementary, and to link only this to intonation.

We will gloss over the technical details here,⁴ and focus on the main points: (6) shows the (simplified) definition of *given* in this approach; (7) links it to F-marking; and (8) gives the additional constraint that is needed to restrict F-marking (since (6) on its own does not say anything about the status of focused information).

(6) *Given* (simplified)

An utterance U counts as *given* iff it has a salient antecedent A which entails the non-F-marked parts of U.

(7) *Givenness*

If a constituent is not F-marked, it must be *given*.

(8) *AvoidF*

F-mark as little as possible, without violating *Givenness*.

Augmented with constraints that relate F-marking to accentuation, the theory makes plausible predictions wrt. accent placement, i.e. which constituents of a sentence must, may, or must not be accented in which contexts.

⁴ Just briefly: to be able to define *given* as 'being entailed by salient parts of the previous discourse', Schwarzschild (1999) defines a semantic operation called 'existential type shifting' that takes arbitrary parts of (the meaning of) antecedent utterances to the type of formulae which can then be tested for whether they entail non focused parts (similarly type-shifted to type *t*) of (the meaning of) the new utterance.

Thus we have briefly sketched three closely related approaches to the pragmatic interpretation of focus features [$\pm F$]. In spite of some differences in the details, all three belong to the same tradition and reflect the notion of information structure presupposed in this paper. The pragmatic definitions of other IS-features, such as topic, discussed in sections below, build up on the presented view of focus.

2.4 How is information structure realized prosodically?

In the previous section we discussed the pragmatic interpretation of focus features, now we turn to their prosodic realization. The best-studied prosodic means of marking focus is the placement of the (nuclear) accent. Some word of the focused constituent receives an accent (cf. the words in small caps in the examples (2)-(5) in the previous sections). This word is then called the *focus exponent*. Roughly, if the focused constituent coincides with the focus exponent, one talks about *narrow focus*; a focused constituent that is larger than its focus exponent is called *broad focus*. In the latter case, special syntactic and pragmatic rules regulate which word of the focused constituent is accented (Selkirk, 1995; Schwarzschild, 1999).

In many languages, e.g. English and German, focus exponents receive pitch accents. However, it is also very common that besides accent position, focus constrains accent type or tune: focus-related pitch accents on new material are usually falling accents (HL-sequences in analyses based on the tone sequence model by Pierrehumbert, 1980), whereas topic accents are often—for example in German—rising accents (LH-sequences, cf. Büring, 1997). Furthermore, languages can prosodically differentiate between broad and narrow focus by means of categorically distinct pitch accent realizations (e.g. Frota (2000) for Portuguese). While it is a standard view that prosodic prominence on phrase level is expressed essentially by local pitch movements, we know little about

the supporting function of other prosodic parameters like e.g. the highlighting of focused material by filled or silent pauses (Arnold et al., 2003; Horne et al., 2004) or the variation of speaking rate to differentiate between given and new material.

Another interesting area in the field of prosody and information structure is the mechanisms of prominence reduction. In Germanic intonation languages like Dutch contextually given material is deaccentuated (complete deletion of pitch accents), while in Romance languages like Italian pitch accents on given material are realized, but with significantly reduced accents range compared to new material (Swerts et al., 1999). In Swedish, a language with lexical accent, pitch accents on contextually given material can also be realized if they occur early in the sentence, but they differ from accents on new material in peak-alignment (Horne et al., 1999). Focus exponents in the scope of a focus-sensitive particle are deaccentuated (deletion of pitch accent), if they occurred already in the identical construction in the immediately preceding context ('second occurrence focus'), but are still marked prosodically prominent by means of durational and intensity features (Beaver et al., 2002).

The last area we would like to sketch out briefly is prosodic phrasing. While in neutral constructions phonological phrasing is a mere reflection of syntactic structure, both levels can vary independently under the influence of information structure. For example, word order variations due to information structural constraints can result in identical phonological phrasing patterns. On the other hand, identical syntactic structures can be phonologically implemented with different numbers of phrases to meet information structural requirements, such as, for example, narrow emphatic focus.

2.5 Interim conclusions and further questions

In the previous sections we formulated our assumptions about the essence of discourse structure and information structure and reviewed recent findings about their prosodic realization. In our opinion, the concepts and facts presented above suggest (at least) two possible ways in which DS and IS could interact.

First, DS and IS could interact at the level of pragmatics. On the one hand, the pragmatics of topic and focus contributes to the appropriateness of an utterance in a given context. On the other hand, discourse structure provides a highly structured representation of the context. Depending on the discourse structure, not all parts of the context might be equally relevant for the distribution of topic and focus features in a particular utterance. In turn, information structure could impose constraints on how the utterance is connected to the context. What these constraints are like, whether these constraints are “direct” or mediated by other structures, what aspects of discourse structure interact with which information-structural features—all these questions and the relevant discussion in the literature will be recapitulated in Section 3.

Second, it is obvious that information structure and discourse structure (see Section 2.2) use the same prosodic devices, namely pitch, durational and intensity parameters, to express different kinds of structures, relations, and prominence. This proximity suggests phenomena of interaction and conflicting requirements. We formulate some related hypotheses in Section 4.

3 DS and IS Interact at the Level of Pragmatics

In this section, we review the current state of research on the issue how discourse structure and information structure communicate at the level of (semantic/pragmatic) interpretation. The field has made substantial progress in recent years. Two major theoretical trends seem to be emerging. In one of them, the

interaction between discourse and information structure is assumed to be mediated by the speakers' joint intention to discuss a certain issue, or a *Question under Discussion* (QUD). The other trend does not make use of this construct and tries to relate DS and IS more or less directly. Since this approach is in some sense simpler, we will start by presenting it in Section 3.1. The approach based on QUD will be discussed in Section 3.2. The QUD framework is conceived in such a way that practically any phenomenon related to information structure (e.g. those discussed in Section 2.3) has to be analyzed as a DS/IS interface phenomenon. Therefore we will dwell somewhat longer on the general architecture of that approach.

3.1 Relational approaches

In this subsection we discuss the (at first glance seemingly unconnected) approaches by Schwarzschild (1999), Asher and colleagues (in the framework of SDRT, see in particular Asher and Txurruka, 1995) and Nakatani (1997). What unites these approaches for our purposes is the common assumption that there are coherence relations that directly link utterances (without recourse to structures like QUDs) and that have an influence in licensing information structure—be they the relations of centering theory that hold between forward- and backward-looking centers (as in Nakatani, 1997), rhetorical relations (as in SDRT), or the relations between anaphor and antecedent, mediated by rhetorical relations (as in Schwarzschild, 1999). We begin with a discussion of the approaches and then look at some phenomena that have been dealt with in these approaches.

3.1.1 The approaches

Givenness as anaphora: Schwarzschild (1999)

In Section 2.3.3 we already introduced the main positions of Schwarzschild's (1999) work, in particular, the notion of *givenness* and its relation to focus marking and accent placement. Here, we address the predictions this approach makes wrt. interactions between accent placement and discourse structure. The part of the proposal that is of the most interest here is the claim that givenness is a form of anaphora, which has to search for a *salient* antecedent, cf. definition (6) in Section 2.3.3. Schwarzschild (1999, p. 165) conjectures that salience is mediated by rhetorical relations (and hence by discourse structure). The author does not mention any particular theory of discourse structure or accessibility, but it would be interesting to test whether this *given*-anaphora behaves in the same way as pronominal anaphora. To give a brief illustration of this question, consider (9) below. The SDRT-conditions on accessibility of antecedents (Asher and Lascarides, 2003) explain why the continuation (e) in this mini-discourse is odd (namely because utterance (e) is in a narrative sequence with (b), which renders the evaluation (c)–(d) of (b) 'inaccessible').

- (9) (a) Sandy had a great evening.
(b) First she called Peter.
(c) They talked about her mother.
(d) And about some common friends.
(e) ??Then she phoned her.

The better acceptability of the intonational contours in (10) below suggests that *givenness*-anaphora underlies weaker constraints than SDRT-accessibility (since (e'') is only licensed by attaching to (c), which in SDRT is not available); this however remains to be tested in a more rigorous empirical manner.

- (10) (e') Then she phoned her MOTHER.
 (e'') Then she PHONED her mother.

Information packaging in SDRT

Asher and Txurruka (1995) integrates a theory of *information packaging* into the discourse theory SDRT (Asher, 1993; Asher and Lascarides, 2003)—‘*information packaging*’ (rather than ‘*information structure*’) being the term that Chafe (1976) introduced (and Vallduví (1992) took up and further developed) to emphasize the function (possibly one among many) that intonation has of indicating to the hearer how the parts of an utterance fit into the context. Asher and Txurruka (1995) make the following assumptions: (a) both the topic/comment-dichotomy and the focus/background-dichotomy have a role in a theory of information packaging; (b) one of the dichotomies becomes *dominant* in a given utterance (in a given context), resulting in utterances being either *focus-dominant* or *topic-dominant*; (c) information packaging is a pragmatic notion, which is only partially determined by intonation or syntax; (d) *rhetorical relations* connecting utterances can constrain the informational structure of their relata (or conversely, a given informational structure can be used as evidence for inferring that a certain rhetorical relation holds).

This already represents the essential ideas of the approach; in the following, we will try to illustrate them a bit more. The distinction in (b) above is motivated by the difference between questions as in (11-a), which set up a *topic* for discussion, and those in (11-b), which set up a *situation* for discussion—and in this approach, this is what the discourse context has to provide, namely either an entity (a topic) or a situation (a background).

- (11) a. A: What about John? What did he do?
 B: JOHN (↗) talked to MARY (↘)

- b. A: Who did John talk to?
B: John talked to MARY.

According to the authors, certain rhetorical relations are sensitive to kinds of information packaging; e.g. relations connecting the utterances to the first type of question in (11) above require topic-dominant structures, as probably do relations like *Narration* and *Elaboration* (although this is not developed in the paper), whereas, for example, *Corrections* require a focus-dominant structure, just like the second kind of question-answer-pair above.

Now, what are those information packaging-sensitive constraints? While the technical details of the proposal are involved (making use of partial isomorphisms between semantic representations, i.e. the SDRSs), the main idea is simple: the focused element in the new information must be mapped to an element in the antecedent utterance, and the remaining parts of both utterances (the results of abstracting over these elements) must be logically equivalent (possibly in a non-monotonic logic that allows for ‘normality’-modalities).^{5,6} An example will hopefully make this clearer:

- (12) A: Who came to the party?
B: JOHN came to the vernissage.

If we map “John” to “who”, we get “came to the party” as the remaining elements of the first sentence, which is equivalent to “came to the festivity” (and

⁵ This amounts to essentially the same idea as in Schwarzschild’s definition of *given*, only in a more direct way: where Schwarzschild ensures through use of type-shifting and existential closure that the antecedent and the given-elements in the utterance are of the right type for the first to entail the second, in this approach, the antecedent utterance itself (or rather, its logical form, LF) and a constructed utterance (LF) as described above are related by the entailment relation.

⁶ In Asher and Txurruka (1995) only focus-dominant-sensitive rhetorical relations are worked out in detail; in later work in SDRT (e.g. Asher and Lascarides (2003)), the distinction between focus-dominance and topic-dominance is not even mentioned.

this shows why logical equivalence is needed, namely because the situation can be referred to differently in either sentence). For (this type of) question, the authors can then additionally require that the focus be mapped onto a *wh*-element; in corrections such as (13) below, the mapping is between elements of the same type (the condition on the background is the same).

- (13) A: John came to the party.
B: No, it was PETER who came to the fete.

Note that contrary to Schwarzschild's motivating assumptions, this approach does not see it as problematic that there is no uniform notion of *newness* (as the relation that connects focused material to an antecedent): there are different rhetorical relations that hold between the utterances in any case, and it is only natural for them to impose different constraints on their relata. Information-packaging constraints on the relation *Narration* aren't worked out in this approach as far as we can see, so there is no developed treatment of our accessibility example above (9), but it should be clear that the mechanism is there in this approach to explain the observed phenomena.

Intonation and Centering Theory

The last approach we will discuss here, that of Nakatani (1997), sticks out a bit in that it is not an attempt to formalize constraints on accent placement, but rather is an analysis of empirical data, in the framework of a specific theory of discourse structure. More precisely, she looked at referring expressions, and correlated features of their form (pronouns vs. full forms), the grammatical function they play in the utterance (subject vs. object), and intonation (prominent vs. non-prominent) to their function according to *Centering Theory* (Grosz et al., 1995). As we will see, this analysis can be seen as keeping the link between intonational (non-)prominence and newness (givenness), while giving a

more nuanced analysis of the second set of terms.

The purpose of Nakatani's study was to test "the general claim about the accentuation of given/new information" which "predicts that (1) pronouns are unaccented and full noun phrases and proper names are accented, and (2) subjects are unaccented and direct objects are accented" (Nakatani, 1997, p. 140). While Nakatani found a supporting tendency, there were also many exceptions to this rule, which she claims can be explained by looking at the status of these expressions with respect to centering theory. Before we can discuss this, we will briefly review the fundamentals of this theory.

Centering looks at patterns of referential connections between utterances, i.e. at how subsequent sentences in a (coherent) discourse keep referring to the same entities or introduce new ones, and formulates rules about which kinds of referential configurations produce "better", i.e. more coherent, discourse. To give an example, centering predicts, corresponding to intuitions, that the discourse in (14-a) is relatively more coherent than that in (14-b), which both mention exactly the same entities, only in different order.

- (14) a. Peter went to a music store.
He really liked that store.
He bought a piano.
- b. Peter went to a music store.
It was a store that he really liked.
A piano was what he bought there.

To formulate the transition rules, centering looks at two aspects of utterances, namely how they set up entities for further reference in subsequent discourse (the *forward-looking* potential), and how they take up entities from previous utterances (the *backward-looking* aspect). The former aspect is captured by assigning each utterance a list in which the entities that are mentioned in the utter-

ance are ranked according to criteria like the grammatical function they play in the sentence (with subjects being higher ranked than objects, for example), or linear order; this is the list of *forward-looking centers* or C_f . Broadly speaking, this ranking is meant to reflect the likelihood that an entity will be mentioned again in the next utterance, or the preference that it be rementioned; hence the highest ranked member is also called the *preferred center* or C_p . The other aspect is represented in the *backward-looking center* C_b , which is the highest ranked member of the previous utterance's C_f that is realized in the current utterance. Differences in perceived coherence (as in example (14) above) are then explained by ranking movements of configurations of forward and backward-looking centers: e.g. *retaining* a center (as in all utterances in (14-a), where “Peter” remains the highest ranked element of C_f) is preferred over *rough shifting* centers from one utterance to the next (as in (14-b), where the highest ranked element is “Peter”, “the store”, and “a piano”, respectively).

This theory of centering is embedded in a more general theory of discourse structure (Grosz and Sidner, 1986); it suffices here to say that in this theory new discourse segments introduce new *centering spaces* which are put on a stack (i.e. which are removed once a segment is closed, to return to the previous centering configuration). These two elements form the theoretical framework in which Nakatani formulates her analysis,⁷ which is presented in Table 1.

Perhaps the only surprising result is the occurrence of prominent pronouns in subject position; according to the author, these were cases like the following (our example):

(15) Peter likes Sandy.

⁷ A word on the scope of the analysis. Nakatani (1997) looked at spoken narrative (i.e. not dialogue), and centering has indeed been developed for monologues (but see Taboada (2002) for a review of attempts to extend centering to dialogue), hence there are no immediate predictions of this approach with respect to the examples we have seen so far in the previous sections.

p	gf	f	$discourse\ function$
+	S	p	shift local attention to new C_b
+	S	ef	Introduce new global referent as C_p
+	DO	ef	Introduce new global referent
–	S	p	Maintain C_b referent in primary local focus
–	DO	p	Maintain non- C_b referent in primary local focus
–	DO	ef	Maintain referent in global focus

Table 1: pr stands for “prominence”; gf for “grammatical function”— S being subject and DO being direct object; and f for “form”, with p for “pronoun” and ef for “explicit form”.

SHE hates him, though.

However, an important point is made explicit in this analysis (whereas it is implicit in both approaches that have been discussed above), namely that *givenness* or *newness* of a given entity is relative to previous discourse structure, and is not absolute over the discourse as a whole, or in other words, that *salience* has a role to play.

As a point of criticism, it has to be pointed out that it’s not entirely clear what exactly Nakatani is talking about: is it nuclear accent, or sentential focus? As far as we can see, she doesn’t say whether one element being prominent implies that all other elements are non-prominent, or whether there were configurations where both subject and object were intonationally prominent.

While Nakatani phrases her analysis rather procedurally in terms of instructions to the listener (e.g. “shift attention to”), it should be possible to reformulate her approach so that it is more in line with the ones discussed above (e.g. “intonational prominence on a subject pronoun is licensed if it is a new C_b ”), and hence, this difference does not favor or disfavor any approach.

3.1.2 The phenomena

In this section we very briefly mention some discourse configurations where IS has an effect on felicity conditions, and say whether, and if so, how the discussed approaches handle them.⁸

Questions and answers

The well-known IS constraints on questions and answers (already mentioned above), illustrated by (16), are handled by Schwarzschild (1999) through the *givenness* mechanism which handles all discourse effects: in B' below “Sandy” is not given, and hence its non-prominence violates that constraint.

- (16) A: Who came to the party?
B: SANDY came to the party.
B': *Sandy came to the PARTY.

Asher and Txurruka (1995), on the other hand, have specific constraints for question-answer-pairs, as discussed above.

Contrast

Example (17) illustrates the IS-constraints on contrastive sequences. Again the fact that “speaks” is non-prominent can be explained by Schwarzschild (1999) with his *givenness* constraint. This approach, however, has nothing to say about why the other elements have to be accented with this particular contour. Moreover, to explain the realization of the first sentence, the approach would have

⁸ Note that all those dialogue acts (answering, correction, clarification) can also be performed with fragmental utterances (e.g. “A: Peter came to the party.—B: No, SANDY.”) where the fragment consists only of the focus. This has been noted before (Ginzburg (1999b); Schlangen (2003)), but has not yet been systematically studied, to our knowledge.

to allow *given* to be *cataphoric*, i.e. to take its “ante”cedent from subsequent discourse.

- (17) JOHN (↗) speaks FRENCH (↘).
BILL (↗) speaks GERMAN (↘).

IS-constraints on *Contrast* are mentioned in passing in Asher and Lascarides (2003), and again the general idea is that this relation requires this particular contour, and conversely that this contour can be used to infer this relation.

Correction and focus

Corrections have already been discussed above; the treatment in the approaches follows the same general lines as for other phenomena.

Clarification and focus

The influence of focus-marking on the interpretation of clarification requests (i.e. questions addressing the question of understanding a previous utterance, including acoustic understanding; cf. Ginzburg (1999a); Schlangen (2004)) like those in (18) has so far not been studied systematically, but it seems that there are constraints on the information structure which work very similarly to those on corrections, and which help determine which elements are being clarified.

- (18) A: Peter hit Sandy.
B: Peter HIT Sandy? (= “Did you say that what Peter did to Sandy was to hit her?”)
B': Peter hit SANDY? (= “Did you say that the person Peter hit was Sandy?”)

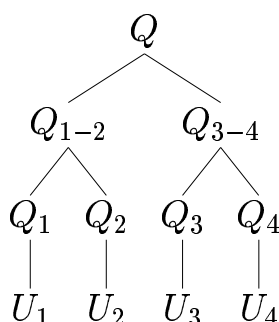


Figure 1: Abstract QUD-structure of a sequence of utterances $\langle U_1, U_2, U_3, U_4 \rangle$. The utterances are immediately dominated by (i.e. answer) the questions Q_1, \dots, Q_4 , respectively. The dominance relations between questions correspond to subquestion relations. The whole discourse is dominated by the root node Q , i.e. the discourse answers the question Q .

3.2 Questions under discussion

3.2.1 Theoretical background

The so-called *Question under Discussion* (QUD) approach is based on the assumption that *each* utterance in discourse answers an explicit or implicit question (under discussion). If the question is actually uttered by one of the conversation participants, it is *explicit*, and *implicit* otherwise. The relations between utterances are construed as relations between the underlying questions.⁹ This contrasts with the assumptions made in other frameworks, such as RST, SDRT, D-TAG, in which discourse relations connect the utterances directly (cf. Section 2.1). We are not going to discuss the specific gain of such a complication in the theory. For the time being it is enough to say that QUD is at present the most prominent framework addressing the relation between discourse structure and information structure as one of its central issues.

It is necessary to note that the QUD approach does not form a coherent framework comparable to SDRT or D-TAG. Rather, it is a collection of partly

⁹ A schematic discourse structure (Büring's style, cf. Büring, 2003) is given in Figure 1.

isolated independent proposals, which include Klein and von Stutterheim (1987, 1992, 2001); von Stutterheim and Klein (1989); van Kuppevelt (1995a,b); Roberts (1996, 1998); Ginzburg (1996a,b); Larsson (2002); Cooper (2003); Ginzburg and Cooper (2004); Büring (2003). In this section, we try to formulate something that could be viewed as the “core” of the QUD framework, paying tribute to the individual proposals where possible.

The cue notions of the QUD framework are those of *question*, *answer*, *inquiry strategy* and various relations between questions within a strategy.

Questions and answers:

According to the standard view, a question is identified with the set of its possible answers. This idea was first introduced by Hamblin (1973) in connection with the semantic analysis of interrogative sentences. It was further developed by Karttunen (1977), Higginbotham and May (1981), Groenendijk and Stokhof (1984), among others, covering more and more aspects of linguistic behavior of interrogatives. However, all these proposals were aimed at providing a purely context-independent compositional semantics for interrogative sentences. It was acknowledged that our understanding of a question, and hence our ability to give an appropriate answer is strongly dependent on what is *relevant* in the current situation, but only with the work of Ginzburg (1995) was the idea explicitly integrated into the formal analyses of interrogative semantics. Most of the work in the QUD-framework adopted a notion of question based on Ginzburg’s context-sensitive approach to the semantics of interrogatives.¹⁰ The idea, or at least one of its aspects, is illustrated in examples (19) and (20), which show two possible answer-set interpretations of the question *Who attended the meeting?*

¹⁰ In the work by Klein and von Stutterheim (1987, 1992, 2001) and van Kuppevelt (1995a,b) sets of propositions associated with the nodes of a discourse tree are called *topics*, rather than *questions*. However, all QUD-based proposals seem to agree on the type of object associated with discourse nodes. Therefore, we ignore these terminological differences and uniformly call these objects *questions*.

- (19) a. John attended the meeting.
 b. Mary attended the meeting.
 c. John and Mary attended the meeting., etc.
- (20) a. Eight linguists attended the meeting.
 b. Fifteen psychologists attended the meeting., etc.

Which interpretation is actually chosen, i.e. what is currently relevant, seems to be regulated by two major classes of constraints. First of all, as was argued by Ginzburg, questions (as discourse goals) depend on the domain-level goals of the interlocutors. See van Rooy (2003) for a formal account of this relationship in terms of statistical decision theory. Second, questions addressed in a discourse depend on each other. For instance, the set of propositions associated with question Q_2 in Figure 1 depends on the sets Q_1 and Q_{1-2} —the nodes Q_2 is structurally related to. This class of constraints is discussed in more detail in the next section.

Relations between questions:

As mentioned above, in the QUD framework relations between utterances in discourse are cast in terms of relations between the (implicit) questions they answer. There are two major ways in which such relations are characterized, which we could dub as *static*, or declarative, and *dynamic*, or procedural. However, it appears that both approaches can be successfully used to model interdependencies between questions in discourse.

The static approach: Within the static approach, certain relations between questions play the role of constraints on allowed discourse structures, the most important being the relation of *subquestion* and the underlying notion of *inquiry strategy*. Informally, the set S of questions $\{Q_1, \dots, Q_n\}$ is a *strategy* for answering question Q_0 iff answering all of the questions in S gives an answer

to Q_0 . A sequence of utterances U_1, \dots, U_n forms a discourse constituent dominated by question Q_0 only if U_1, \dots, U_n are dominated by (i.e. answer) the questions Q_1, \dots, Q_n respectively which in turn form a strategy for answering Q_0 . Various versions of the notion of *strategy* are present in both Klein and von Stutterheim's and van Kuppevelt's work, although they do not use the term.¹¹ Roberts (1996) seems to have introduced the term in the discourse-structural context. She relates the notion to Groenendijk and Stokhof's question entailment, whereas Büring (2003) has explicitly used it to formulate constraints on discourse structures in a consistently declarative way.

Given the notion of strategy, a *subquestion* is defined as follows: Question Q is a *subquestion* of question Q' iff it belongs to some strategy of answering Q' (cf. Roberts' (1996) and Büring's (2003) notion of relevance of a question). It should be noted though that frequently subquestionhood is given an "absolute", strategy-independent definition: Q is a *subquestion* of Q' iff in order to answer Q' , Q has to be answered first (cf. van Kuppevelt (1995a) and Ginzburg's (1996a) notion of *dependent* questions). However, such a definition appears too strong, because often there are multiple alternative strategies of answering a question that are equally good. For instance, a question like *Who ate what?* can be addressed via at least two strategies: going "by people", cf. Figure 2; or going "by food", cf. Figure 3. Of course, we would like to treat a question like *What did John eat?* as a legitimate subquestion of *Who ate what?*, but strictly speaking, it need not be addressed in order to find out who ate what. Namely, one could do without asking *What did John eat?* if one chose the by-food strategy.

To summarize, within the static approach questions under discussion are used to define the hierarchical discourse structure based on the relation of *sub-*

¹¹ Jan van Kuppevelt's analogue to strategies is the relation of *subquestion* (cf. below) and, as he calls it, the "conjunctive property of subquestions" (van Kuppevelt, 1995a, p. 125, 24ff). Klein and von Stutterheim use the term *referentielle Bewegung* 'referential movement'. They do not give a general definition, but formulate a number of text type specific constraints on referential movement, which can be seen as special cases of strategies in the above sense.

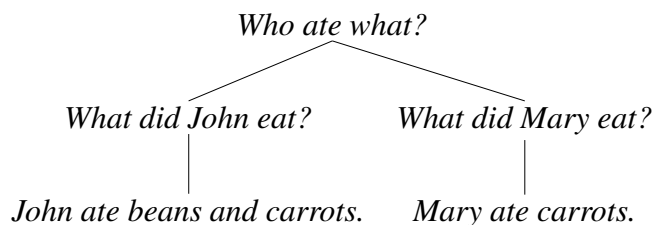


Figure 2: Addressing the question *Who ate what?*: The “by-person” strategy.

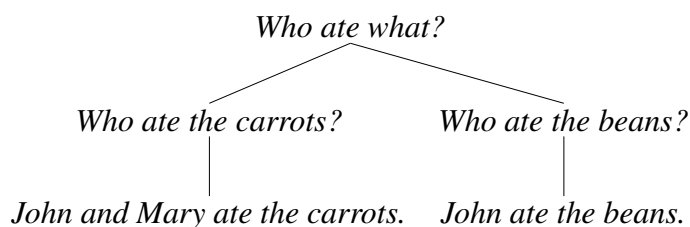


Figure 3: Addressing the question *Who ate what?*: The “by-food” strategy.

question and *strategy*. The possible combinations of question nodes in the resulting tree structure is constrained by the definition of *subquestion* relation, based on Groenendijk and Stokhof’s notion of question entailment or Ginzburg’s dependent questions. However, as we will see in Section 3.2.2, these constraints are not enough, for instance, to predict the correct distribution of topic accents.¹²

The dynamic approach: A different style of formulating the same constraints is represented by a considerable bulk of work on dialogue (Ginzburg, 1996a,b; Larsson, 2002). Rather than formulating constraints on possible discourse trees, the context of an utterance is modeled by a highly structured *information state* and a set of rules for dealing with it that models the behavior of discourse par-

¹² See our discussion of Büring (2003) on pp. 187–189.

ticipants during a conversation. The information state includes, among other things, a stack of questions under discussion (the QUD stack).¹³ The question at the top of that stack corresponds to the current question under discussion, whereas the other questions below it correspond to its superquestions.¹⁴ The QUD stack is *updated* and *downdated* by the interlocutors in the course of conversation. If a question is explicitly asked by one of them, it is pushed on the QUD stack, i.e. the QUD stack is updated. A question can also be pushed on the stack, if it is “presupposed” by some utterance, in which case the question has to be *accommodated*.¹⁵ Accommodated questions roughly correspond to implicit questions in the static view. Once a question is on the QUD stack, the conversation participants are committed to address it until it is either answered or determined to be practically unanswerable (cf. Roberts, 1996). However, once a sufficient answer is provided, the question must be popped, i.e. the QUD stack is downdated (cf. van Kuppevelt’s (1995a) Dynamic Principle of Topic Termination). When a question is popped off the QUD stack, its immediate superquestion becomes topmost and the procedure is repeated until there are no more questions left (cf. van Kuppevelt’s (1995a) Principle of Recency). The QUD stack management in the dynamic approach can be viewed as a way of processing hierarchical discourse trees defined in the static approach.¹⁶

It should also be noted that in addition to hierarchical relationships between questions and their processing, the QUD-based dynamic analyses of dialogue have concentrated in particular on modeling the distinction between the public and the interlocutors’ private part of the information state, and the process of *grounding*. The idea is that an utterance is not *automatically* accepted by the

¹³ The notion that gave its name to the whole trend in discourse theory.

¹⁴ A superquestion is the reverse of subquestion: if Q_1 is a subquestion of Q_2 , then Q_2 is a superquestion of Q_1 .

¹⁵ Roberts (1996) argues that it is a general function of focus to introduce a question presupposition.

¹⁶ Note that processing a context-free grammar requires a stack.

interlocutors, i.e. a question is not immediately added to the QUD stack and an assertion does not immediately update the common ground (the set of shared facts). First it has to be understood by all the interlocutors and acknowledged to be relevant to the current question under discussion. Most importantly, the question whether an utterance should be accepted is by itself a discussable issue, and should be modeled in such a way that it can be integrated in the overall hierarchy (stack) of questions. See in particular Ginzburg (1997), Ginzburg and Cooper (2004) and Ginzburg (forthcoming). These issues have not received much attention in the analysis of monologue, whereas, as we already noted, the notion of grounding in a monologue is probably key for a QUD-based definition of subordination.¹⁷

Relations between declarative utterances

Relations between declarative utterances in a monologue have received much less attention in the QUD framework than in “relational” theories of discourse, such as RST or SDRT. Indeed, it seems that what the hierarchy of questions gives us is discourse constituency, i.e. which utterances belong closer together than others (one strategy vs. different strategies), but it does not seem to provide any distinction between coordination and subordination, or various semantic relations such as *Elaboration*, *Explanation*, and *Contrast*.

At present, there is no systematic account of these aspects of discourse structure in the QUD framework, but there are a few ideas circulating on how these issues could be approached. A promising starting point is van Kuppevelt’s (1995a; 1995b; 1996) classification of unsatisfactory answers and the notion of *grounding* in the dynamic trend (Ginzburg and Cooper, 2004). A speaker, even a speaker of a monologue that gets no on-line feedback from the audience, sometimes has reasons to assume that the utterance he has produced gives an

¹⁷ See Zaenen et al. (2001) for a more detailed comparison of the QUD-based approaches to dialogue and monologue.

unsatisfactory answer to a question he was addressing, or in other words, that the utterance could not be grounded yet. Such a situation licenses “reraising” the same question, or alternatively, raising a question whether the utterance is true. Cooper (2003, p. 372) suggests that such question reraising corresponds to subordination structures. E.g. in (21) both sentences appear to address the question *How does John look?*, (21-b) *elaborating* on (21-a).¹⁸ Apparently, the speaker assumes that (21-a) does not give a satisfactory account of John’s appearance, and decides to dwell on the same issue in (21-b). By contrast, in (22), which exhibits a coordination structure, the first sentence is accepted as a satisfactory answer, and the speaker moves on immediately to the next question *What does John do?*

- (21) a. John is a nice looking guy.
b. He has blond hair.
- (22) a. John is a nice looking guy.
b. He works for a bank.

Within the class of subordinating relations, the opposition between *Elaboration* and *Explanation* could be reconstructed using van Kuppevelt’s distinction (van Kuppevelt, 1995b, 1996) between *quantitatively* and *qualitatively* unsatisfactory answers (cf. Zaenen et al. (2001) for related discussion). A further development of this idea can be found in Jasinskaja (2004). The paper proposes a method to compute the semantic effects associated with the discourse relation of *Elaboration* using a QUD-based approach.

¹⁸ Note that *Elaboration* is a subordinating relation.

3.2.2 QUD-based explanations

Focus accent and focus-sensitive particles

In one of its most radical versions (e.g. Roberts, 1996), the QUD-based theory reduces the semantics of focus to its function in question-answer pairs, already discussed in Section 2.3.1. Focus introduces a *question presupposition*: the presupposed question can be simply read off the focus structure of this sentence (by replacing foci by *wh*-phrases). For instance, the sentence *Mary voted for [John]_F* presupposes the question *Who did Mary vote for?*; *[Mary]_F voted for John* presupposes *Who voted for John?* (cf. examples (1) and (2) in Section 2.3.1); whereas a sentence with double focus *[Mary]_F voted for [John]_F* presupposes a double question *Who voted for whom?* In other words, focus accent placement is completely determined by the underlying question in this rather straightforward way.

According to this view, the analysis of all contextual functions of focus (cf. also Sections 2.3.2 and 2.3.3) becomes a matter of DS/IS-interface. For instance, Roberts' analysis of the focus-sensitive character of *only* crucially relies on the notion of strategy and the subquestion relation between the items on the QUD stack. Roughly, a sentence with *only*, e.g. (23-b), presupposes a question "with *only*" such as (23-a). This question can in turn be successfully accommodated in the context of a corresponding question "without *only*" (24). The set of propositions associated with this question determines the alternatives that *only* quantifies over in (23-b).

- (23) a. Mary only invited "who" for dinner?
 b. Mary only invited [LYN]_F for dinner
- (24) Who did Mary invite for dinner?
 a. Mary invited John for dinner.
 b. Mary invited Bill for dinner., etc.

The process of accommodation is governed by constraints on admissible strategies at the level of discourse structure. On the one hand, this is supposed to make sure that (23-a) can be accommodated when (24) is under discussion which renders the sequence (24)-(23-b) felicitous. On the other hand, this mechanism should provide that, for instance, the question *Who did Mary vote for?* (1-a) *cannot* be accommodated under *Who voted for John?* (1-b) to account for the fact that the sequence (1-b)-(2-a) (*Who voted for John?—Mary voted for [John]_F*) is infelicitous. Roberts' definition of strategy based of Groenendijk and Stokhof's notion of question entailment captures this fact.

Topic accent

Building up on Roberts' (1996) proposal, Büring (2003) uses the notions of question under discussion and strategy in order to predict the occurrence of falling A-accents (\searrow) and (falling-)rising B-accents (\nearrow) in English. He relates the former to focus [$\pm F$], and the latter to *contrastive topic* [$\pm CT$]. The basic idea is that contrastive topics, unlike foci, do not just presuppose a single question, but a whole strategy around the current utterance. Büring formalizes this idea in the definitions (25) and (26) below.¹⁹

(25) CT-Congruence:

An utterance U containing a contrastive topic can map onto a move M_U within a d-tree D only if U indicates a strategy around M_U in D .

(26) U indicates a strategy around M_U in D iff there is a non-singleton set Q' of questions such that for each $Q \in Q'$, (i) Q is identical to or a sister of the question that immediately dominates M_U , and (ii) $\|Q\|^o \in \|U\|^{ct}$

Ignoring technical details, these definitions predict, for example, that the utterance [*Fred*]_{CT} ate the [*beans*]_F (where *Fred* is the contrastive topic and bears

¹⁹ A *d-tree* is a discourse tree, and the term *move* is used to refer to nodes of a d-tree.

a B-accent, and *beans* is a focus and bears an A-accent) indicates a by-people strategy (cf. Section 3.2.1, pp. 180–182): *What did Fred eat?*, *What did Mary eat?*, etc. Whereas if we switch the focus and the topic marking, as in $[Fred]_F$ *ate the* $[beans]_{CT}$, the strategy indicated by this utterance is the by-food strategy: *Who ate the beans?*, *Who ate the carrots?*, etc. This analysis predicts that the CT+F and the F+CT accentuation pattern cannot be freely exchanged. For instance, (27) should be ill-formed according to this approach.

- (27) Who ate what?
- a. $[FRED \nearrow]_{CT}$ ate the $[BEANS \searrow]_F$.
 - b. #And $[MARY \searrow]_F$ ate the $[CARROTS \nearrow]_{CT}$.

Further, Büring (2003) uses Schwarzschild's notion of *givenness*, which we already discussed in Sections 2.3.3 and 3.1.1 (pp. 169–170), to account for the choice between a topic accent and no accent at all. Whereas a topic accent can indicate a strategy that may include *implicit* questions (cf. Section 3.2.1, p. 178), givenness only takes into account overt moves, i.e. declarative utterances in the previous context and explicit questions. Therefore, if the subquestion in (28-b) remains implicit, the accentuation pattern in (28-d) is infelicitous, the lack of accent on *female* violates givenness. However, if (28-b), which contains the word *female*, is uttered explicitly, both (28-c) and (28-d) are possible answers.

- (28)
- a. What did the pop stars wear?
 - b. (What did the female pop stars wear?)
 - c. The $[FEMALE \nearrow]_{CT}$ pop stars wore $[CAFTANS \searrow]_F$.
 - d. #The female pop stars wore $[CAFTANS \searrow]_F$.

Thus Büring's theory describes mutual constraints imposed by, on the one hand, the configuration of discourse nodes in the vicinity of an utterance and, on the other hand, topic and focus accentuation in that utterance. These constraints are

cast in terms of the notions of well-formed strategy and subquestion. However, as Buring (2003, p. 530) notes himself, in some cases a more refined rating of strategies would be necessary in order to account for the accentuation facts. He suggests that the *efficiency* of strategies could be one such measure. For instance, there is nothing wrong with the implicit question (29-b) as a subquestion of (29-a). However, going “by clothes” appears a much less efficient strategy in this case than going “by groups of pop stars”, since there are normally too many different kinds of clothes, which would give rise to very long lists. Therefore, the accentuation pattern in (29-c) appears less appropriate than the pattern in (28-c) in the same (explicit) context.

- (29) a. What did the pop stars wear?
 b. #(Who wore caftans?)
 c. #The [FEMALE ↘]_F pop stars wore [CAFTANS ↗]_{CT}.

Buring does not elaborate this part of his proposal. To our knowledge, the impact of relative efficiency of strategies on discourse coherence has not yet been sufficiently investigated in connection with the contextual effects of information structure. We think, however, that this would be an important concept which would make the QUD-based theories somewhat less permissive in general, and make it possible to formulate interesting hypotheses in the domain of DS/IS-interface.

Information structure and discourse relations

Contrast: Contrast is an important notion of both discourse structure and information structure, although different things are usually meant. A recent special issue of the *Journal of Semantics* (de Hoop and de Swart, 2004) has concentrated on clearing the terminological confusion and studying the interaction of the different aspects of this notion. On the one hand, focus (contrastive fo-

cus, contrastive topic) in general involves a kind of “contrast” between the focus alternatives it induces. On the other hand, the discourse relation of contrast is said to hold between utterances that “compare” two situations with respect to their similarities and differences, or between utterances where the second contradicts a default expectation associated with the first one. In the latter case, one also often talks about *concession*, a discourse relation closely related to contrast proper and typically signaled by the same set of discourse connectives, e.g. *but* or *although*.

Umbach (2004) proposes a QUD-based analysis that integrates both the discourse-structural and the information-structural notion of contrast. First, Umbach investigates various restrictions on the sets of alternatives—the quantification domains of exhaustive operators such as *only* and bare exhaustification. Then she proposes an analysis in which the contrastive and the denial-of-expectation uses of *but* are correlated with different properties of underlying alternative sets. See also Jasinskaja (2002) and Zeevat (2004) for a more elaborate discussion and formalization of the quantification domain restrictions, as well as Kruijff-Korbayová and Webber (2000, 2001) for an account of focus sensitivity and the contrastive vs. denial-of-expectation uses of *although*.

Elaboration: Like Umbach wrt. *Contrast*, Jasinskaja (2004) uses constraints on alternative sets and the notion of exhaustification to infer the semantic effects associated with the discourse relation of *Elaboration*. By definition, *Elaboration* holds between two utterances where the first one introduces an event, and the second “elaborates” that event, i.e. adds more detail to the description of that event or some part of it. Jasinskaja does not discuss the influence of accent placement, but the general architecture of the approach is the same: the exhaustification operator quantifies over a set of alternatives constrained, on the one hand, by the current question under discussion, and on the other hand, additional constraints, such as *distinctness* (cf. Zeevat, 2004). The *Elaboration*

relation is inferred whenever two utterances address the same QUD.

Clarification: Finally, we should mention Ginzburg and colleagues' QUD-based approach to ellipsis, and various discourse relations that frequently involve ellipsis, e.g. *Correction*, *Acknowledgement* and *Clarification* (Ginzburg and Sag, 2000; Ginzburg and Cooper, 2004, Ginzburg, forthcoming). All these relations are specific to dialogue and connect utterances by distinct interlocutors. For instance, *Clarification* is a relation between an utterance by one interlocutor, and a question asked by another interlocutor, in order to clarify the content of that utterance. Two notions of the QUD framework play a central role in this analysis: the current question under discussion and the notion of grounding, cf. pp. 182–184. First, the content of an elliptical utterance is almost entirely constructed from the contextually salient question (Ginzburg and Sag, 2000). Second, Ginzburg and Cooper (2004) argue that before an utterance is sufficiently understood by all the participants of the conversation and can be *grounded*, not only the semantic objects it introduces, but also elements of its syntactic and phonological representation must be available for reference in subsequent utterances, since this kind of reference is necessary for a proper analysis of *Clarification* questions. In accordance with these insights, Ginzburg and Cooper account for the ambiguous character of interrogatives like *Finagled?* (in the context of *Did Bo finagle a raise?*) which can be paraphrased either as *Are you asking if Bo (of all actions) FINAGLED a raise?*, or as *What does it mean to finagle?*

3.3 Discussion

We do not intend to engage in a full scale comparison of the predictions of the theoretical approaches presented above. Instead, we would like to summarize the findings of previous studies and formulate our tentative answer to the

question of which aspects of discourse structure interact with which aspects of information structure, as far as their pragmatic interaction is concerned.

In Section 2.1, we articulated three types of information that can be encoded in discourse structure: (1) discourse constituency, (2) discourse coordination vs. subordination, and (3) specific semantically/pragmatically motivated discourse relations, such as *Elaboration*, *Contrast*, *Narration*, etc.

On the basis of the findings in the literature, one could claim that information structure, as manifested by accent choice and placement, stands in especially close interaction with the third of these elements—the choice of a discourse relation. That is, once we know (or hypothesize) that the utterances u_1 and u_2 form a discourse constituent, the information structure of these utterances helps us constrain the choice among possible discourse relations. The most uncontroversial is the connection between the discourse relation of *Contrast* and contrastive topic/focus marking. But *Concession*, as well as various dialogue-specific relations, also appears to be affected by information structure, mainly by the location of the focus accent. At the same time, the impact of other discourse relations (e.g. *Narration*) on information structure is not so well-studied and awaits further research.

On the other hand, discourse relations are also the most controversial part of discourse structure, and some theories e.g. Grosz and Sidner (1986) or the QUD-approaches do not acknowledge them. In this group of theories the relation between information structure and the choice of a semantic/pragmatic relation between utterances is not direct, but is mediated by speakers' intentions (e.g. as QUDs). That is, information structure helps the hearer recover the set of issues the speaker intends to discuss, which in turn affects the semantic relations between the individual utterances. However, up to now there has been no extensive theoretical proposal that accounts for the connection between questions under discussion and the semantic relations between declarative utterances. This is again a desideratum for the future.

But as was mentioned, the choice of a discourse relation between two utterances presupposes the existence of a structural connection between these utterances (they have to form a discourse constituent). So what about discourse constituency, and other “properly structural” aspects of discourse structure? Do they have anything to do with information structure and accenting or not? Sporadic mentions in the literature (e.g. Schwarzschild, 1999, pp. 165–166) suggest that they do, and Nakatani (1997) confirmed it at least for pronouns. In brief, the presence vs. absence of a structural relation, as well as coordinating vs. subordinating character of the relation determines which individuals, properties, situations etc. are *salient* at the current point in discourse, and that in turn affects which items can be considered as given or new, and hence whether they should be accented or uttered without an accent. This is another field in the domain of pragmatic IS/DS-interface which calls for further investigation.

4 DS and IS Interact at the Level of Phonetics

As previously mentioned, prosodic correlates of information structure and discourse structure exploit the same phonetic parameters. IS-driven prosodic phenomena like pitch accent position and type, deaccentuation, short pauses in the vicinity of focus exponents, or decreased speaking rate on focus exponents are mostly local phenomena operating on words or syllables. DS-driven phenomena, on the other hand, tend to be global, i.e. affecting whole phrases (pitch range, general speaking rate and intensity) or the strength of separation of adjacent phrases (pause duration and type). But this distinction along the local/global dimension is not complete. There is a gray area where IS- and DS-driven prosodic phenomena come very close and may impose conflicting requirements. According to this situation, we distinguish two types of interaction on the phonetic/phonological level: regular interactions and instances of conflicting requirements.

Conflicting requirements can be expected particularly with pauses and speaking rate. Both parameters are used to signal IS- as well as DS-driven prosodic phenomena. Information structure is one source of prosodic phrasing which is implemented—among other features—by pauses between phrases and by phrase-final decrease of speaking rate (‘final lengthening’). Discourse prosody, on the other hand, usually falls back on the existing phrasing and controls pause duration and more general variations of speaking rate, i.e. rate differences between phrases, not within phrases. But besides the dividing function, pauses and speaking rate variation may also be employed by IS to isolate and highlight focused items, as mentioned in Section 2.4. This leads to a revision of the neutral, syntax-based phrasing and potentially entails conflicts with DS-derived phrasing requirements, since discourse structure is crucially linked to syntactic structure.

Regular interactions are due to the similarity of the expressive phonetic parameters. In these cases, both IS- and DS-requirements are fulfilled, but with noticeable modifications in one area depending on specific parameter settings in the other area. An interesting example of this type of interaction is the influence of compressed pitch range on the realization of pre-nuclear pitch accents. In phrases with an overall falling contour, the first accent is usually realized with the highest pitch peak, while following accents are downstepped, resulting in the lowest pitch-peak assigned to the nuclear (i.e. last) accent (see Figure 4, top). In phrases with compressed pitch range this mechanism would lead to very low pitch peaks on nuclear accents (see Figure 4, bottom left), further disguised by creaky voice which often co-occurs with compressed pitch range. It was never systematically investigated whether this really happens or whether other strategies are available to the speaker to preserve an appropriate prominence level of the nuclear accent. This question is of particular interest to the discussion on the theoretical status of discourse prosody. Assuming the standard view that discourse related variation of global prosodic parameters is not part

of the phonological representation of an utterance but directly implemented at the phonetic level, the phonetic process of gradual lowering of pitch accents is the only possibility of the production system to adapt local pitch events to the compressed range. If we assume, however, that discourse prosody is phonologically represented, then local and global prosodic aspects of an utterance can interact at the phonological level, which allows at least two additional strategies to handle the problem at hand. If the phonological system is aware of the compressed pitch range one possibility would be to inhibit the phonological process of downstep, resulting in a nuclear accent with equal height to the prenuclear accents. This, however, is not likely because it would possibly change the metrical pattern and subsequently the interpretation of the utterance (Ladd, 1996). We assume that a deaccentuation strategy is more likely (see Figure 4, bottom/right). Deaccentuation of prenuclear pitch accents leaves the full range to the realization of the nuclear accent, while the metrical strength of the prenuclear items can be maintained using intensity or durational prominence features (cf. Beaver et al. (2002) for the realization of second occurrence focus). An investigation of this problem has to be based on production data as well as perception experiments, since particularly the perceptual function of non-pitch prosodic features regarding sentence-level prominence is still unclear.

5 Conclusions

In this paper we presented the current state of research on the issue of how discourse structure and information structure interact. We propose that this interaction should be investigated at (at least) two levels. First, these two structures impose mutual constraints at the pragmatic level. This aspect of DS/IS interaction has been in the focus of linguists' interest for a some time now, and we have tried to give a comprehensive survey of relevant studies. However, we have found a lot of theoretical and empirical issues that still have to be clarified in

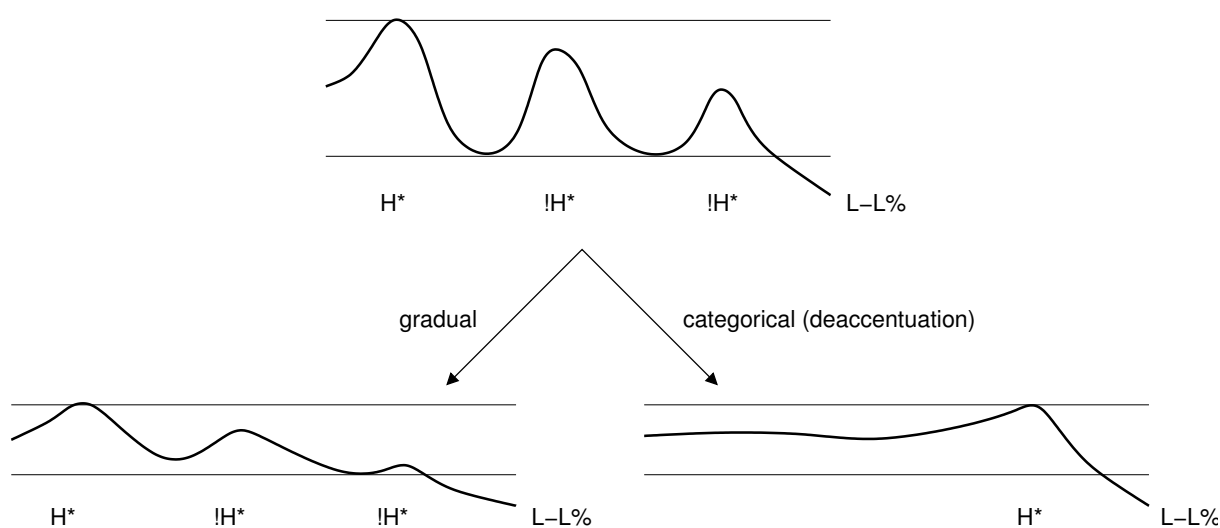


Figure 4: Phonetic (bottom/left) and phonological (bottom/right) adaption of multi-accent utterances to compressed pitch range.

order to complete the picture.

Empirical testing is particularly called for by theories that treat the contextual effects of IS as a species of anaphora, such as Schwarzschild's (1999) theory of givenness, cf. Section 3.1.1, pp. 169–170. Combined with a theory of anaphoric accessibility (based on one of the available discourse models, e.g. DRT, SDRT, or LDM), an anaphoric approach to IS provides clear testable hypotheses about the influence of discourse constituency and subordination relations on accent placement. Verifying these hypotheses appears to be a highly relevant and realistic research task.

On the theoretical side, the QUD-based approach provides a vast field for further work. As we have seen in Section 3.2, the whole architecture of this approach is such that it “postpones” the treatment of most IS phenomena to the level of discourse structure. Contextual effects that are traditionally treated at the level of focus semantics are now cast in terms of constraints on accommodation of (implicit) questions, or more generally, constraints on possible inquiry strategies. But the solutions proposed so far are not sufficient. One possible di-

rection for strengthening these theories is to rate strategies according to their efficiency, as suggested by Büring (2003, p. 530). In general, it should be said that if the QUD framework's ambition is to provide a general theory of discourse structure and semantics, a number of rather central components still have to be developed. For instance, comprehensive accounts of discourse anaphora, as well as semantic relations between utterances, are still missing. Once these missing parts are supplied, the framework is likely to provide a rather broad notion of DS/IS interface and at the same time a uniform treatment of various interface phenomena.

Finally, the interaction of prosodic topic/focus marking with the choice of semantic discourse relation leaves space for both theoretical and empirical investigations. So far, the only discourse relations whose information-structural effects have been studied reasonably well are *Contrast*, *Correction* and *Question-Answer Pair*, cf. Section 3.1.1, pp. 170–172 and Section 3.2.2, pp. 189–190. It is still unclear whether other relations, such as *Narration*, *Elaboration*, *Explanation*, have their characteristic information structures and accentuation patterns. Some interesting ideas on this issue were sketched out by Asher and Txurruka (1995) and need to be both worked out theoretically and tested empirically.

These research areas are suggested by the current literature and they all pertain to the study of DS/IS interface at the level of pragmatics. However, as we emphasized in this paper, discourse structure and information structure address a number of prosodic devices which in turn interact at the level of phonetics (or phonology). These phenomena have received much less attention in general and have never been brought up in connection with the issue of DS/IS interface. We have outlined some directions in which this line of research could go.

Bibliography

Arnold, J., Fagnano, M., and Tanenhaus, M. (2003). Disfluencies signal thee, um, new information. *Journal of Psycholinguistic Research*, 32:25–36.

- Asher, N. (1993). *Reference to Abstract Objects in Discourse*. Studies in Linguistics and Philosophy. Kluwer Academic Publisher, Dordrecht.
- Asher, N. and Lascarides, A. (2003). *Logics of Conversation*. Studies in Natural Language Processing. Cambridge University Press.
- Asher, N. and Txurruka, I. G. (1995). Extending SDRT to integrate a theory of the informational partition. Technical Report ILCLI-95-LIC-3, Institute for Logic, Cognition, Language and Information, Euskal Herriko Universidad.
- Asher, N. and Vieu, L. (2003). Subordinating and coordinating discourse relations. *Lingua, to appear*.
- 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., Clark, B., Flemming, E., and Wolters, M. (2002). Second occurrence focus is prosodically marked. Ms., Stanford University.
- Büring, D. (1997). The great scope inversion conspiracy. *Linguistics and Philosophy*, 20:175–194.
- Büring, D. (2003). On D-trees, beans and B-accent. *Linguistics and Philosophy*, 26:511–545.
- Chafe, W. (1976). Givenness, contrastiveness, definiteness, subject, topics, and point of view. In Li, C., editor, *Subject and Topic*, pages 27–55. Academic Press, New York.
- Cooper, R. (2003). Afterword (to the special issue on discourse and information structure). *Journal of Logic, Language and Information*, 12:369–374.
- Cristea, D., Ide, N., and Romary, L. (1998). Veins theory: A model of global discourse cohesion and coherence. In *Proceedings of COLING-ACL 1998*, pages 281–285.
- Culicover, P. W. and McNally, L., editors (1998). *The Limits of Syntax*, volume 29 of *Syntax and Semantics*. Academic Press.

- de Hoop, H. and de Swart, P. (2004). Contrast in discourse: Guest editors' introduction. *Journal of Semantics*, 21:87–93.
- den Ouden, H., Noordman, L., and Terken, J. (2002). The prosodic realization of organizational features of texts. In *Proceedings of Speech Prosody 2002*, pages 543–546, Aix-en-Provence.
- Frota, S. (2000). *Prosody and Focus in European Portuguese. Phonological Phrasing and Intonation*. Garland, New York.
- Gardent, C. (1997). Discourse tree adjoining grammar. Claus report 89, University of the Saarland, Saarbrücken.
- Ginzburg, J. (1995). Resolving questions 1 & 2. *Linguistics and Philosophy*, 18:459–527, 567–609.
- Ginzburg, J. (1996a). Dynamics and the semantics of dialogue. In Seligman, J. and Westerståhl, D., editors, *Logic, Language and Computation*, volume 1. CSLI Publications, Stanford, CA.
- Ginzburg, J. (1996b). Interrogatives: Questions, facts and dialogue. In Lappin, S., editor, *The Handbook of Contemporary Semantic Theory*. Blackwell, Oxford.
- Ginzburg, J. (1997). On some semantic consequences of turn taking. In Dekker, P., Stokhof, M., and Venema, Y., editors, *Proceedings of the 11th Amsterdam Colloquium*, ILLC-Publications, Amsterdam.
- Ginzburg, J. (1999a). Fragmenting meaning: Clarification ellipsis and nominal anaphora. In Bunt, H., editor, *Proceedings of the 3rd Workshop on computational semantics*, Tilburg, The Netherlands.
- Ginzburg, J. (1999b). Semantically-based ellipsis resolution with syntactic pre-suppositions. In Bunt, H. and Muskens, R., editors, *Computing Meaning: Current issues in computational semantics*. Kluwer, Dordrecht.
- Ginzburg, J. (forthcoming). *Semantics and Interaction in Dialogue*. CSLI Publications and University of Chicago Press, Stanford CA. Draft chapters available from <http://www.dcs.kcl.ac.uk/staff/ginzburg>.

- Ginzburg, J. and Cooper, R. (2004). Clarification, ellipsis, and the nature of contextual updates in dialogue. *Linguistics and Philosophy*, 27:297–365.
- Ginzburg, J. and Sag, I. A. (2000). *Interrogative Investigations: The Form, Meaning and Use of English Interrogatives*. CSLI Publications, Stanford CA.
- Groenendijk, J. and Stokhof, M. (1984). *Studies in the Semantics of Questions and the Pragmatics of Answers*. PhD thesis, University of Amsterdam.
- Grosz, B. and Hirschberg, J. (1992). Some intonational characteristics of discourse structure. In *Proceedings of the 2nd International Conference on Spoken Language Processing*, pages 429–432, Banff, Canada.
- Grosz, B., Joshi, A. K., and Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225.
- Grosz, B. J. and Sidner, C. L. (1986). Attention, intentions and the structure of discourse. *Computational Linguistics*, 12(3):175–204.
- Grüning, A. and Kibrik, A. A. (2002). Referential choice and activation factors: A neural network approach. 4th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC'2002). Lisbon, Portugal.
- Gustafson-Capková, S. and Megyesi, B. (2002). Silence and discourse context in read speech and dialogues in Swedish. In *Proceedings of Speech Prosody 2002*, pages 363–366, Aix-en-Provence.
- Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language*, 10:41–53.
- Hendriks, P. (2004). Coherence relations, ellipsis and contrastive topics. *Journal of Semantics*, 21:133–153.
- Herman, R. (2000). Phonetic markers of global discourse structures in English. *Journal of Phonetics*, 28:466–493.
- Higginbotham, J. and May, R. (1981). Questions, quantifiers and crossing. *Linguistic Review*, 1:41–79.

- Hirschberg, J. and 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*, pages 286–293, Santa Cruz.
- Horne, M., Frid, J., and Roll, M. (2004). Hesitation disfluencies after the clause marker *att* 'that' in Swedish. *Lund University Working Papers in Linguistics*, 51.
- Horne, M., Hansson, P., Bruce, G., and Frid, J. (1999). Prosodic correlates of information structure in Swedish human-human dialogues. In *Proceedings of EURO-SPEECH99*, pages 29–32, Budapest.
- Jasinskaja, E. (2004). Exhaustification and semantic relations in discourse. In Geurts, B. and van der Sandt, R., editors, *Proceedings of the ESSLLI'2004 Workshop on Implicature and Conversational Meaning*, pages 14–19, Nancy, France.
- Jasinskaja, K. (2002). Relevance and other constraints on the quantification domain of *only*. In *Proceedings of the Workshop on Information Structure in Context*. IMS Stuttgart.
- Kamp, H. and Reyle, U. (1993). *From Discourse to Logic: Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*. Kluwer Academic Publishers.
- Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and Philosophy*, 1:3–44.
- Klein, W. and von Steutterheim, C. (1987). Quaestio und die referentielle Bewegung in Erzählungen. *Linguistische Berichte*, 109:163–185.
- Klein, W. and von Steutterheim, C. (1992). Textstruktur und referentielle Bewegung. *Zeitschrift für Literaturwissenschaft und Linguistik*, 86:67–92.
- Klein, W. and von Steutterheim, C. (to appear, 2001). How to solve a complex verbal task: Text structure, referential movement and the quaestio. In Espinas, G. S. and Brum, M., editors, *Recherche sur l'acquisition des langues*.

- Kodzasov, S. V. (1996). Zakony frazovoj akcentuacii. In Nikolaeva, T. M., editor, *Prosodičeskij stroj rusškoj reči*, pages 181–204. Institut rusškojazyka Rossijskoj akademii nauk, Moskva.
- Komogata, N. (2003). Information structure in subordinate and subordinate-like clauses. *Journal of Logic, Language and Information*, 12:301–318.
- Kruijff-Korbayová, I. and Steedman, M. (2003). Discourse and information structure. *Journal of Logic, Language and Information*, 12:249–259.
- Kruijff-Korbayová, I. and Webber, B. L. (2000). Discourse connectives, inference and information structure. In Bos, J. and Kohlhase, M., editors, *Workshop on Inference in Computational Semantics (ICoS-2)*, pages 105–120, Schloß Dagstuhl, Germany. International Conference and Research Center for Computer Science.
- Kruijff-Korbayová, I. and Webber, B. L. (2001). Concession, implicature, and alternative sets. In Bunt, H., editor, *4th International Workshop on Computational Semantics (IWCS-4)*, pages 227–248, Tilburg, The Netherlands.
- Ladd, D. R. (1996). *Intonational Phonology*. Cambridge University Press.
- Larsson, S. (2002). *Issue-based Dialogue Management*. PhD thesis, Göteborg University.
- Lehiste, I. (1975). The phonetic structure of paragraphs. In Cohen, A. and Nooteboom, S. G., editors, *Structure and Process in Speech Perception*, pages 195–203. Springer.
- Lehiste, I. (1979). Perception of sentence and paragraph boundaries. In Lindblom, B. and Öhman, S., editors, *Frontiers of Speech Research*, pages 191–201. Academic Press, London.
- Mann, W. C. and 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. and 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.
- Nakatani, C. H. (1997). Discourse structural constraints on accent in narrative. In van Santen, J., Sproat, R. W., Olive, J., and Hirschberg, J., editors, *Progress in Speech Synthesis*, pages 139–156. Springer, New York.
- Partee, B. H. (1996). Allegation and local accomodation. In Partee, B. H. and Sgall, P., editors, *Discourse and Meaning: Papers in Honor of Eva Hajičová*, pages 65–86. John Benjamins, Amsterdam.
- Passonneau, R. J. and Litman, D. J. (1993). Feasibility of automated discourse segmentation. In *Proceedings of the 31th Annual Meeting of the Association for Computational Linguistics*, pages 148–155, Ohio.
- Pierrehumbert, J. (1980). *The Phonology and Phonetics of English Intonation*. PhD thesis, MIT.
- Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics*, 12:601–638.
- Roberts, C. (1996). Information structure in discourse: Towards an integrated formal theory of pragmatics. *OSU Working Papers in Linguistics*, 49:91–136.
- Roberts, C. (1998). Focus, the flow of information, and universal grammar. In Culicover and McNally (1998), pages 109–160.
- Rooth, M. E. (1985). *Association with Focus*. PhD thesis, University of Massachusetts, Amherst.
- Rooth, M. E. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1:75–116.
- Schlangen, D. (2003). *A Coherence-Based Approach to the Interpretation of Non-Sentential Utterances in Dialogue*. PhD thesis, School of Informatics, University of Edinburgh, Edinburgh, UK.

- Schlangen, D. (2004). Causes and strategies for requesting clarification in dialogue. In *Proceedings of the 5th Workshop of the ACL SIG on Discourse and Dialogue*, Boston, USA.
- Schwarzschild, R. (1999). GIVENNESS, AVOIDF and other constraints on the placement of accent. *Natural Language Semantics*, pages 141–177.
- Selkirk, E. (1995). Sentence prosody: Intonation, stress and phrasing. In Goldsmith, J., editor, *Handbook of Phonological Theory*, pages 550–569. Blackwell, Oxford, UK.
- Sluijter, A. M. C. and Terken, J. M. B. (1993). Beyond sentence prosody: Paragraph intonation in Dutch. *Phonetica*, 50:180–188.
- Smith, C. L., McCraw, M. O., and Hogan, L. A. (2002). The role of topic structure in speech timing: A comparison of English and Japanese. Talk at LabPhon VIII.
- Swerts, M. (1997). Prosodic features at discourse boundaries of different strength. *Journal of the Acoustical Society of America*, 101:514–521.
- Swerts, M., Avesani, C., and Krahmer, E. (1999). Reaccentuation or deaccentuation: A comparative study of Italian and Dutch. In *Proceedings of the XIVth International Congress of Phonetic Science*, pages 1541–1544, San Francisco.
- Swerts, M. and Geluykens, R. (1993). The prosody of information units in spontaneous monologue. *Phonetica*, 50:189–196.
- Swerts, M. and Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, 37(1):21–43.
- Taboada, M. (2002). Centering and pronominal reference: In dialogue, in Spanish. In *Proceedings of the 6th Workshop on the Semantics and Pragmatics of Dialogue (EDILOG)*, Edinburgh, UK.
- Umbach, C. (2004). On the notion of contrast in information structure and discourse structure. *Journal of Semantics*, 21:155–175.
- Vallduví, E. (1992). *The Informational Component*. Garland, New York.

- Vallduví, E. and Vilkuna, M. (1998). On rheme and kontrast. In Culicover and McNally (1998), pages 79–108.
- van Donzel, M. (1999). *Prosodic Aspects of Information Structure in Discourse*. PhD thesis, The Hague University.
- van Kuppevelt, J. (1995a). Discourse structure, topicality and questioning. *Journal of Linguistics*, 31:109–147.
- van Kuppevelt, J. (1995b). Main structure and side structure in discourse. *Linguistics*, 33:809–833.
- van Kuppevelt, J. (1996). Directionality in discourse: Prominence differences in subordination relations. *Journal of Semantics*, 13:363–395.
- van Kuppevelt, J. (1999). Topic and comment. In Brown, K. and Miller, J., editors, *Concise Encyclopedia of Grammatical Categories*, pages 369–374. Elsevier, Amsterdam.
- van Rooy, R. (2003). Questioning to resolve decision problems. *Linguistics and Philosophy*, 26:727–763.
- Venditti, J. and Swerts, M. (1996). Prosodic cues to discourse structure in Japanese. In *Proceedings of the International Conference on Spoken Language Processing*, pages 725–728, Philadelphia.
- von Stutterheim, C. and Klein, W. (1989). Referential movement in descriptive and narrative discourse. In Dietrich, R. and Graumann, C. F., editors, *Language Processing in Social Context*, pages 39–76. North Holland, Amsterdam.
- Webber, B., Joshi, A., Stone, M., and Knott, A. (2003). Anaphora and discourse structure. *Computational Linguistics*.
- Zaenen, A., Larsson, S., Sandor, A., and Traum, D. (2001). Text and dialogue. In *The Trindi Book*, chapter 9, pages 261–296.
- Zeevat, H. (2004). Contrastors. *Journal of Semantics*, 21:95–112.

Ekaterina Jasinskaja, Jörg Mayer and David Schlangen
Universität Potsdam
SFB 632, Institut für Linguistik
Postfach 601553
14415 Potsdam
Germany
jasinsk@rz.uni-potsdam.de
{mayer,das}@ling.uni-potsdam.de