Introduction

MARK STEEDMAN
Informatics, University of Edinburgh, 2 Buccleuch Place, Edinburgh EH8 9LW, U.K.
steedman@cogsci.ed.ac.uk

IVANA KRUIJFF-KORBAYOVÁ
Computerlinguistik, Universität des Saarlandes, Saarbrücken, Germany
korbay@coli.uni-sb.de

Two Dimensions of Information Structure in Relation to Discourse Semantics and Discourse Structure

The following introduction is intended to do no more than to establish some basic distinctions and definitions within which we believe all the theoretical positions represented at the workshop can be compared and/or reconciled and their often close relationships to each other be understood.

1 Some Tentative Definitions

The title of the workshop implies a three-way distinction between Information Structure, Discourse Semantics, and Discourse Structure. While on occasion these terms have been used interchangeably or in overlapping senses, we believe that it is useful to distinguish them as follows:

- Information Structure is a Sentence Internal partition of the information in an utterance according to its relation to the discourse context under dichotomies such as topic/comment, theme/rheme, given/new, focus/background, etc. Such categories are essentially Referential in nature.

- Discourse Semantics centrally concerns the nature of the contextual model, and the entities in it to which Information Structural categories relate, in the form of alternative sets, filecards, presupposed propositions, etc.
• Discourse Structure concerns the **inter-clausal** relations of explanation, elaboration, exemplification, and illocutionary force that hold between successive utterances of a discourse or dialog, supporting inference about the domain and purposes of the discourse.

### 2 A Computational Model

We have tried in setting up the workshop to keep computational models of discourse and dialogue in focus. In particular we believe it may be useful to think of these distinctions in terms of some extension to Grosz and Sidner’s (Grosz and Sidner(1986))) computational account. Specifically:

- Information structure seems to belong in what they call “linguistic structure”—essentially, the grammar.

- Its Discourse Semantic extensions seem to concern the “Attentional State” component. (Although the structural representation of this state as a stack of focus spaces seems questionable, and the ontology of types that may inhabit focus spaces must be greatly extended—see Poesio and Di Eugenio, and Forbes et al. below.)

- Other rhetorical and illocutionary aspects of Discourse Structure appear to relate to their “Intentional Structure” component.

### 3 The Two Dimensions of Information Structure

The terminology that is used to describe Information Structure and its semantics is simultaneously various, and under-formalized. Yet it seems that all definitions have some elements in common. They all draw at least one of the following distinctions: (i) a “topic/comment” or “theme/rheme” distinction between the part of the utterance that relates it to the purpose of the discourse, and the part that advances the discourse; (ii) a “given/new” distinction, between parts of the utterance—actually, words—which contribute to distinguishing the content from other alternatives that the context makes available and those parts that are common to all of them.

There are differences among the theories of course. Some, like Halliday’s (Halliday(1967)), view these two distinctions as orthogonal, applying at independent levels of structure. Others in the Bolingerian tradition, like Vallduví, McCoy, Molnár, and ourselves in the present volume view them as different aspects of a single level of structure. An important issue that further differentiates the Bolingerian theories is that of whether pitch accent corresponds to a single contrastive notion of focus applying to both theme and rheme, or whether “contrastive focus” is a distinct notion, applying to explicitly mentioned entities and associated with topic or theme alone (see Molnár, and Umbach, below). Many of these questions
await fuller empirical and corpus based studies: we welcome the studies of this kind by Spenader, McCoy, Prévot, and Poesio and Di Eugenio.

There are further similarities: while some of the theories leave the associated discourse semantics at an intuitive level, the theories which do address formal semantic issues all tend to use some version of “update” semantics, such as the filecard semantics and segmented DRSs of the Kampo-Heimian synthesis (see Prévot, von Heusinger, and Gundel et al., below), or the Alternative semantics of Rooth and Büring (see Nouwen, Komagata, Umbach, and Kruijff-Korbayová and Webber, in the present collection).

The diagram on the next page displaying our view of the influences and terminological dependencies in theories of Information Structure and the associated Discourse Semantics may help to make these commonalities more obvious.

4 An Example that Exemplifies Both Dimensions

The terminology that we use ourselves may serve to illustrate these two dimensions at work. (We take the terms theme and rheme from Firbas and Bolinger, rather than Halliday, for our topic/comment distinction, and use Dahl’s background and focus for the Hallidean given/new distinction, and apply these elements in a single level of information structure.)

(1) Q: I know that Mary likes the man who wrote the musical. But who does she admire?

A: (Mary admires) (the woman who directed the musical)

In English we have claimed that Information Structure is homomorphic to Intonation Structure, and that the contour described in Pierrehumbert’s (Pierrehumbert(1980)) notation as L+H* LH% is one of the “theme tunes” that identify the intonational phrase Mary admires as the theme of this utterance, while H* LL% is a “rheme tune”. Within both the theme and the rheme, the presence of one or more pitch accents identifies words which contribute to distinguishing that theme and/or that rheme from other themes and rhemes that the context affords. Thus we follow (Bolinger(1965)) and (Lambrecht(1994)) in viewing the role of accent in English as a single undifferentiated contrastive meaning applying to both informational components.

We think of that context as an updatable database including two Roothian alternatives sets, respectively called the “Theme Alternative Set” and the “Rheme Alternative Set”—cf. “Alternative Semantics” (Rooth(1985); Rooth(1992)), (Büring(1995)).

In Combinatory Categorial Grammar (Steedman(2000)), Information Structure and the associated structured meanings are associated directly with surface syntactic derivational structure, uniting Information Structure, Intonation Structure, and Surface Syntax in a single module of Grammar.
5 Some Questions for the Workshop

In terms of this we hope uncontentious systematization of the terms of reference of the workshop, the following questions can be offered for consideration among the many others raised by the presenters at the workshop:

- What is the Discourse Semantics of Information Structure?
- What part does Information Structure play in Discourse Structure?
- How does Information Structure relate to Intonation Structure in English and related languages?
- How does Information Structure relate to Syntax and Semantics in the languages of the world?
- *Your question here*

6 Some Answers that Emerged

- **What is the Discourse Semantics of Information Structure?**
  There seemed to be two flavors of semantics: Forbes et al., Korbayová & Webber, and McCoy used forms of Alternative Semantics. Prévot; Vallduví, and Poesio were in the Kampo-Heimian Structured Meaning tradition. This appears to be a mainly notational difference and the systems seem in some sense equivalent.

- **What part does Information Structure play in Discourse Structure?**
  Many Discourse structural and Discourse Semantic phenomena discussed at the workshop appeared to depend at least in part on on information-structural representations in the context, including anaphora, presupposition, and the meaning of connectives like *but, otherwise and although*, and discourse particles like the Russian *zhe, ved’*. (Polanyi; Spenader; Poesio, Nouwen; Korbayová & Webber; Umbach)

- **How does Information Structure relate to Intonation Structure in English and related languages?**
  Several papers claimed or were compatible with the view that Intonation Structure in English is identical to an underspecified form of Information Structure, and in particular that intonational phrases correspond to information units. (Vallduví; Umbach; Prévot; McCoy; Komagata; Steedman.) There were interesting questions raised about dialect variation, the consistent specification of information structural role by prosodic tones, and the status of All-Theme utterances (Kamp; Gundel.)
How does Information Structure relate to Syntax and Semantics in the languages of the world?

There was general agreement that the same pair of orthogonal distinctions—in Vallduvand Villkuna’s terms, theme/rheme and Kontrast/ground—were at work in other languages, marked by other devices such as morphology, verbal particles, word-order etc. (Prévot; Komagata; McCoy; Molnár; von Heusinger). There was an open question how does Umbach’s *questio* fit in—is it a theme?

7 Further Questions that Emerged

- *How are Contextual Representations to be Managed?*
  Contexts, whether viewed as collections of alternative sets or of DRSs or filecards have a complex structure. Alternative representations of Attentional State were discussed, including those related to computational dialog management, such as that proposed in the TRINDI framework, which maintains lists of QUDs. Themes and QUDs seem closely related. The question of representing changing salience of discourse objects was also discussed.

- *The Role of Accomodation*
  Accomodation was central to a number of proposals. What types of discourse entity can accomodate, and when is accomodation blocked?

- *The Role of Information Structure Partitioning*
  What is the linguistic unit of discourses structure? It seemed clear the themes and rhemes could be discontinuous within a single utterance. The question of whether any such units were obligatory remains unresolved. The question of how to reliably identify informational structural units in real texts remained problematic, but a number of tests were proposed including looking at effects of systematically alternating intonation (and equivalent parameters in other languages) and the effect on the interpretation of IS-dependent discourse connectives like *although*, focus particles, etc.

- *What is the best empirical agreement for investigating these questions?* There was general agreement that dialog, rather than extended exposition, was the domain that really stretched the theories.

**Bibliography**


6


