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.

Paper presented at the ESSLLI 2001 Workshop on
Information Structure, Discourse Structure and Discourse Semantics
Edited by Ivana Kruijff-Korbayová and Mark Steedman.

- 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 file-card 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?

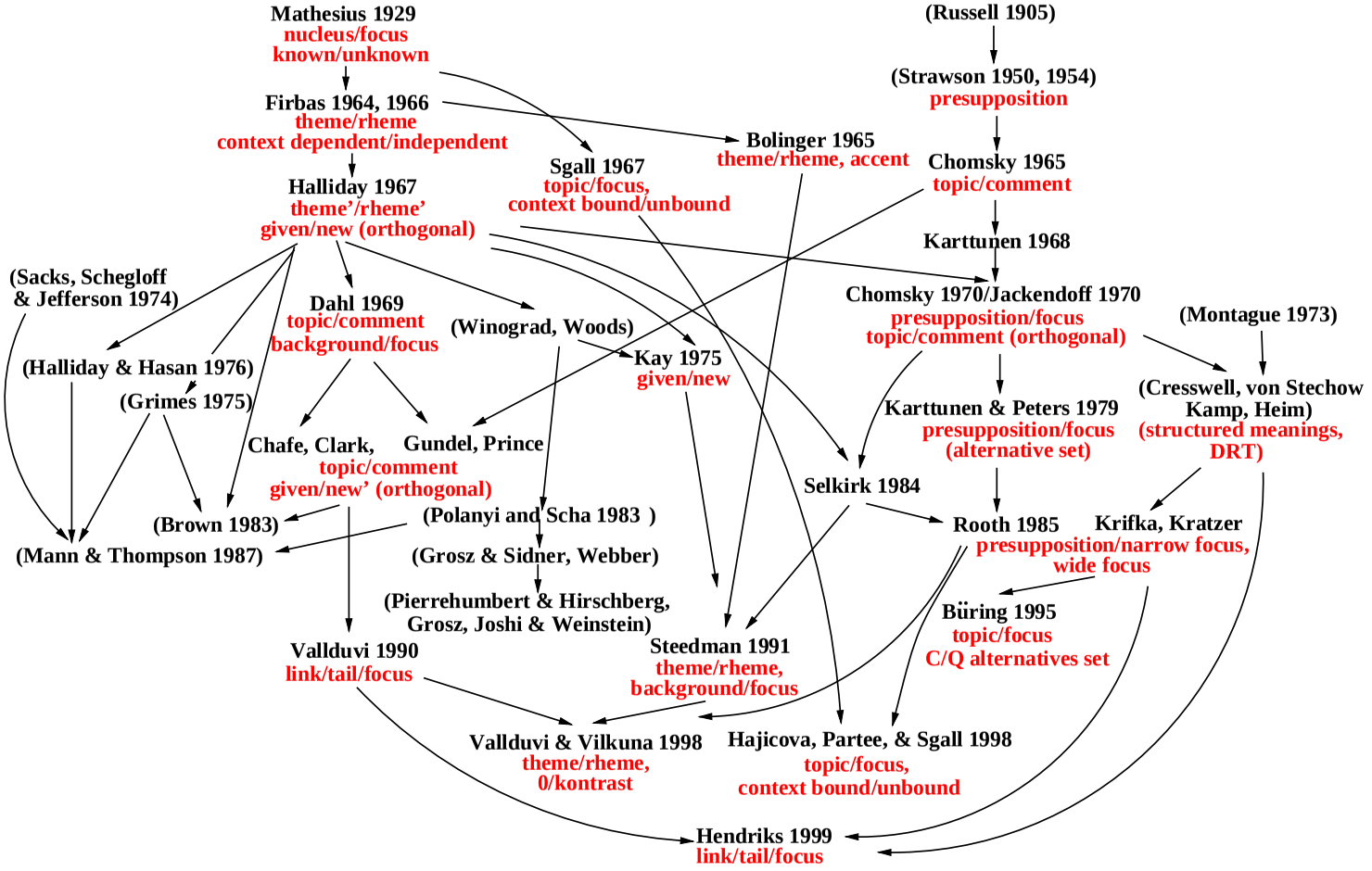
A: (Mary ADMIRES) (the woman who DIRECTED the musical)

$$\underbrace{\underbrace{\qquad\qquad}_{\textit{Background}}\ \underbrace{\mathrm{L}+\mathrm{H}*\mathrm{LH\%}}_{\textit{Focus}}}_{\textit{Theme}}\ \underbrace{\underbrace{\qquad\qquad}_{\textit{Background}}\ \underbrace{\mathrm{H}*\qquad}_{\textit{Focus}}\ \underbrace{\mathrm{LL\%}}_{\textit{Background}}}_{\textit{Rheme}}$$

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.


Mathesius 1929
nucleus/focus
known/unknown
Firbas 1964, 1966
theme/rheme
context dependent/independent
Halliday 1967
theme'/rheme'
given/new (orthogonal)
Sgall 1967
topic/focus,
context bound/unbound
Bolinger 1965
theme/rheme, accent
(Russell 1905)
(Strawson 1950, 1954)
presupposition
Chomsky 1965
topic/comment
Karttunen 1968
(Sacks, Schegloff
& Jefferson 1974)
Dahl 1969
topic/comment
background/focus
(Winograd, Woods)
Kay 1975
given/new
Chomsky 1970/Jackendoff 1970
presupposition/focus
topic/comment (orthogonal)
(Montague 1973)
(Halliday & Hasan 1976)
(Grimes 1975)
(Cresswell, von Stechow
Kamp, Heim)
(structured meanings,
DRT)
Karttunen & Peters 1979
presupposition/focus
(alternative set)
Chafe, Clark,
topic/comment
given/new' (orthogonal)
Gundel, Prince
Selkirk 1984
(Brown 1983)
(Polanyi and Scha 1983)
Rooth 1985
Krifka, Kratzer
presupposition/narrow focus,
wide focus
(Mann & Thompson 1987)
(Grosz & Sidner, Webber)
(Pierrehumbert & Hirschberg,
Grosz, Joshi & Weinstein)
Büring 1995
topic/focus
C/Q alternatives set
Vallduvi 1990
link/tail/focus
Steedman 1991
theme/rheme,
background/focus
Vallduvi & Vilkuna 1998
theme/rheme,
0/kontrast
Hajicova, Partee, & Sgall 1998
topic/focus,
context bound/unbound
Hendriks 1999
link/tail/focus

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 Infomation 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 Vallduvíand Vilkuna'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 *quæstio* 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 dicourse 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 alterning 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

Bolinger, Dwight. 1965. *Forms of English*. Cambridge, MA: Harvard University Press.

Brown, Gillian. 1983. "Prosodic Structure and the Given/New Distinction." In Anne Cutler, D. Robert Ladd, and Gillian Brown, eds., *Prosody: Models and Measurements*, 67–77. Berlin: Springer-Verlag.

Büring, Daniel. 1995. *The 59th Street Bridge Accent: On the Meaning of Topic and Focus*. Ph.D. thesis, Universität Tübingen. Publ. as (**?**).

Chafe, Wallace. 1976. "Givenness, Contrastiveness, Definiteness, Subjects, Topics and Points of View." In Charles Li, ed., *Subject and Topic*, 25–55. New Tork: Academic Press.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1971. "Deep Structure, Surface Structure, and Semantic Interpretation." In Danny Steinberg and Leon Jakobovits, eds., *Semantics*, 183–216. Cambridge: Cambridge University Press.

Clark, Herbert, and Susan Haviland. 1977. "Comprehension and the Given-New Contract." In R. Freedle, ed., *Discourse production and comprehensio*, 1–40. Norwood, NJ: Ablex.

Cresswell, M.J. 1973. *Logics and Languages*. London: Methuen.

Dahl, Östen. 1969. *Topic and Focus: a Study in Russian and General Transformational Grammar*. Göteborg: Elandres Botryckeri.

Firbas, Jan. 1964. "On Defining the Theme in Functional Sentence Analysis." *Travaux Linguistiques de Prague*, 1, 267–280.

Firbas, Jan. 1966. "Non-thematic Subjects in Contemporary English." *Travaux Linguistiques de Prague*, 2, 229–236.

Grimes, Joseph. 1975. *The Thread of Discourse*. Mouton.

Grosz, Barbara, Aravind Joshi, and Scott Weinstein. 1995. "Centering: A Framework for Modeling the Local Coherence of Discourse." *Computational Linguistics*, 2, 203–225.

Grosz, Barbara, and Candace Sidner. 1986. "Attention, Intention and the Structure of Discourse." *Computational Linguistics*, 12, 175–204.

Gundel, Janet. 1974. *The Role of Topic and Comment in Linguistic Theory*. Ph.D. thesis, University of Texas, Austin.

Hajičová, Eva, Barbara H. Partee, and Petr Sgall. 1998. *Topic-Focus Articulation, Tripartite Structures, and Semantic Content*. Dordrecht: Kluwer.

Halliday, Michael. 1967. "Notes on Transitivity and Theme in English, Part II." *Journal of Linguistics*, 3, 199–244.

Halliday, Michael, and Ruqaiya Hasan. 1976. *Cohesion in English*. Longman.

Heim, Irene. 1982. *The semantics of definites and indefinite noun phrases in English.*. Ph.D. thesis, University of Massachussetts, Amherst.

Hendriks, H. 1998. "Links without Locations." In F. Hamm and E. Zimmermann, eds., *Linguistische Berichte, Sonderheft 9/1998: SEMANTIK*. Opladen: Westdeutscher Verlag.

Jackendoff, Ray. 1972. *Semantic Interpretation in Generative Grammar*. Cambridge, MA: MIT Press.

Kamp, Hans. 1984. "A Theory of Truth and Semantic Representation." In Jeroen Groenendijk, Theo Janssen, and Martin Stokhof, eds., *Truth, Interpretation, and Information*, 1–41. Dordrecht: Foris.

Karttunen, Lauri. 1968. "What Makes Definite Noun Phrases Definite?" Tech. rep., Rand Corporation.

Karttunen, Lauri, and Stanley Peters. 1979. "Conventional Implicature." In Choon-Kyu Oh and David Dinneen, eds., *Syntax and Semantics 11: Presupposition*, 1–56. New York: Academic Press.

Kay, Martin. 1975. "Syntactic Processing and Functional Sentence Perspective." In Roger Schank and Bonnie Nash-Webber, eds., *Theoretical Issues in Natural Language Processing, (Supplement)*, 12–15. MIT, June.

Kratzer, Angelika. 1991. "The Representation of Focus." In Arnim von Stechow and Dieter Wunderlich, eds., *Semantics: An International Handbook of Contemporary Research*, 825–834. Berlin: de Gruyter.

Krifka, Manfred. 1991. "A Compositional Semantics for Multiple Focus Constructions." *Cornell Working Papers in Linguistics*, 10, 127–158.

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.

Mann, William, and Sandra Thompson. 1987. "Rhetorical structure theory: A framework for the analysis of texts." Tech. Rep. RS-87-185, Information Science Institute, Marina del Rey, CA.

Mathesius, Vilem. 1929. "Functional Linguistics." In Josef Vachek, ed., *Praguiana: Some Basic and Less Well-known Aspects of the Prague Linguistics School*, 121–142. Amsterdam: John Benjamins. Publ. 1983.

Montague, Richard. 1973. "The Proper Treatment of Quantification in Ordinary English." In Jaakko Hintikka, J. M. E. Moravcsik, and Patrick Suppes, eds., *Approaches to Natural Language: Proceedings of the 1970 Stanford Workshop on Grammar and Semantics*, 221–242. Dordrecht: Reidel. Reprinted in Montague 1974, 247-279.

Pierrehumbert, Janet. 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. thesis, MIT. Distributed by Indiana University Linguistics Club, Bloomington.

Pierrehumbert, Janet, and Julia Hirschberg. 1990. "The Meaning of Intonational Contours in the Interpretation of Discourse." In Philip Cohen, Jerry Morgan, and Martha Pollack, eds., *Intentions in Communication*, 271–312. Cambridge, MA: MIT Press.

Polanyi, Livia, and Remko Scha. 1983. “The Syntax of Discourse.” *Text*, 3, 261–270.

Prince, Ellen. 1981. “Towards a Taxonomy of Given-New Information.” In Peter Cole, ed., *Radical Pragmatics*, 223–256. New York: Academic Press.

Rooth, Mats. 1985. *Association with Focus*. Ph.D. thesis, University of Massachusetts, Amherst.

Rooth, Mats. 1992. “A Theory of Focus Interpretation.” *Natural Language Semantics*, 1, 75–116.

Russell, Bertrand. 1905. “On Denoting.” *Mind*, 14, 479–493.

Sacks, Harvey, Emanuel Schegloff, and Gail Jefferson. 1974. “A Simplest Semantics for the Organization of Turn–Taking for Conversation.” *Language*, 50, 696–735.

Selkirk, Elisabeth. 1984. *Phonology and Syntax*. Cambridge, MA: MIT Press.

Sgall, Petr. 1967. “Functional Sentence Perspective in a Generative Description.” *Prague Studies in Mathematical Linguistics*, 2, 203–225.

von Stechow, Arnim. 1991. “Focusing and Backgrounding Operators.” In Werner Abraham, ed., *Discourse Particles: Descriptive and Theoretical Investigations on the Logical, Syntactic and Pragmatic Properties of Discourse Particles in German*, 37–84. Amsterdam: John Benjamins.

Steedman, Mark. 1991. “Structure and Intonation.” *Language*, 67, 262–296.

Steedman, Mark. 2000. “Information Structure and the Syntax-Phonology Interface.” *Linguistic Inquiry*, 34, 649–689.

Strawson, Peter. 1950. “On Referring.” *Mind*, 59, 320–344.

Strawson, Peter. 1954. “A Reply to Mr. Sellars.” *Philosophical Review*, 63, 216–231.

Vallduví, Enric. 1990. *The Information Component*. Ph.D. thesis, University of Pennsylvania.

Vallduví, Enric, and Maria Vilkuna. 1998. “On Rheme and Kontrast.” In Peter Culicover and Louise McNally, eds., *Syntax and Semantics, Vol. 29: The Limits of Syntax*, 79–108. San Diego, CA: Academic Press.

Webber, Bonnie. 1988. “Tense as Discourse Anaphor.” *Computational Linguistics*, 14, 61–73.

Winograd, Terry. 1972. *Understanding Natural Language*. Edinburgh: Edinburgh University Press.