

## Discourse Modeling and the Pragmatics of Reference

Small SSHRC Proposal (1991)

Nancy Hedberg, Department of Linguistics

### SUMMARY

A cooperative speaker's choice of form of referring expression is constrained by assumptions s/he can reasonably make regarding the hearer's knowledge and attention state in the particular context in which the expression is used. Previous research has defined six such 'cognitive statuses' (e.g. in focus, activated, familiar, identifiable, etc.) relevant to referring and proposition-denoting expressions of particular types (e.g. pronominal, demonstrative, indefinite, etc., cleft presuppositions), and a theory explaining the logical and empirical relations among them has been developed.

This project aims (1) to extend the domain of the theory to explore the role of cognitive status in constraining the use of plural and quantified expressions, (2) to more rigorously define the elements of the theory, adapting techniques of formal discourse modelling developed by logical semanticists and computational linguists, critically comparing definitions of similar concepts, (3) to test the predictions of the theory against an existing corpus of natural language discourse, and (4) to explore implications of the theory for the semantics of quantification and for computational models of reference resolution.

### PROJECT DESCRIPTION

#### A. CONCEPTUAL FRAMEWORK

The form of referring expressions, like other aspects of language such as word order and sentence intonation, depends on assumptions that a cooperative speaker can reasonably make regarding the hearer's knowledge and attention state in the particular context in which the expression is used—the 'cognitive status' of the referent. In a series of papers, Gundel, Hedberg & Zacharski (1988, 1989, 1990, under preparation) have developed a theory aimed at explaining the nature of such statuses and the logical and empirical relations among them. We propose that there are six cognitive statuses relevant to the form of referring expressions in natural language discourse, and that these statuses are implicationally related in the Givenness Hierarchy shown in (1):

#### (1) The Givenness Hierarchy

in			uniquely		type
focus	>. activated	>. Familiar.	> identifiable	>. referential	> identifiable
<i>it</i>	{ <i>this, that,</i> <i>this N</i> }	<i>that N</i>	<i>the N</i>	indef <i>this N.</i>	<i>a N</i>

Each status on the hierarchy is a necessary and sufficient condition for the appropriate use of the pronominal or determiner form or forms associated with it. Use of a particular form thus signals that the associated cognitive status is met and, since each status implies all lower

statuses by definition, it also signals that all lower statuses (i.e. statuses to the right) have been met. Compare, for example, the contexts in which the referring expressions in (2) can be appropriately used:

- (2) I couldn't sleep last night.
- |                                      |                                      |
|--------------------------------------|--------------------------------------|
| a. It kept me awake.                 | d. The dog next door kept me awake.  |
| b. This dog/this/that kept me awake. | e. This dog next door kept me awake. |
| c. That dog next door kept me awake. | f. A dog kept me awake.              |

A personal pronoun tells the hearer that the referent is already in the focus of attention; a demonstrative pronoun or proximal determiner tells the hearer that the referent is readily accessible to consciousness; a distal determiner, that the hearer is familiar with the referent; a definite article, that the hearer can identify the referent given the way it is described; a specific indefinite determiner, that the hearer will be able to identify the referent given what is said about it; and an indefinite article, that the hearer can identify the type of referent described.

In my own research (Hedberg 1990, 1991, in press), I have extended the empirical domain of the theory to explore the cognitive status of proposition-denoting expressions, identifying pragmatic similarities between definite referring expressions and various types of presupposed clauses. For example, the propositional information expressed by the relative clause of a cleft sentence (*It was BUSH who won* presupposes that someone won) may already be in the focus of attention, in which case it need not be expressed at all, as in (3); it may be activated in the context, in which case it may be introduced by a proximal dummy pronoun, as in (4); it may be familiar to the hearer, in which case it receives nuclear stress but may be introduced by a distal dummy pronoun, as in (5), or it may be new but uncontroversial (the propositional analog of uniquely identifiable), as in (6), in which case it receives nuclear stress and can only be introduced by the ordinary dummy pronoun *it*:

(3) My heart beat fast, for I had thought that as the discoverer of the body I would be the first to be called; but to my surprise, **it was Marcel**. [*mystery novel*]

(4) M: There **IS** something that happened here that you might not know about.  
N: Fred Lukermann resigned.  
M: Oh, that's right—you talked to Karen.  
N: No, **this was Jeanette who told me**—I talked to her last Sunday. [*telephone conversation*]

(5) ...Mr. and Mrs. Nev Barnes. She bakes bread and pies and sells them, and he snitches some of the proceeds and buys hooch from a bootlegger named Henrietta..."  
**'Was that her bread at breakfast?'**  
'Yes. Salt-rising. You ate four slices.' [*mystery novel*]

(6) **It was just about 50 years ago that Henry Ford gave us the weekend**. On September 25, 1926, in a somewhat shocking move for that time, he decided to establish a 40-hour work week, giving his employees two days off instead of one. [*newspaper filler, from E. Prince*]

## B. PLAN OF RESEARCH.

The primary empirical goal of the proposed project is to extend the domain of the theory still further and explore the role of cognitive status in constraining the use of plural and quantified expressions, to explain, for example, why the students who are upset about their grade in (7) are understood to be students from the activated syntax course:

- (7) I just posted the final grades for my syntax course. **Several students** are upset about their grade.

The primary methodological goal is to devise ways to more rigorously and explicitly define the crucial theoretical concepts, by adapting techniques of formal discourse modelling developed by logical semanticists and computational linguists, and testing the predictions of the more explicit theory against a corpus of natural discourse texts that were collected, transcribed, and informally analyzed during previous phases of the research.

I propose to adopt the formal indexing procedures of file-change semantics (Heim on definite and indefinite article phrases; Enç on plurals and partitives). It will be necessary to modify these procedures to distinguish the four additional cognitive statuses needed for pronominal and demonstrative reference from the two statuses already assumed by file-change semanticists (familiarity and novelty).

One theoretical goal is to critically compare our discourse-pragmatic definitions of topic, focus, definiteness, inferrability and referentiality with recently proposed discourse-semantic definitions (Partee, Pesetsky, Cinque, Ludlow & Neale), and another is to explore the implications of the extended theory for the theory of quantification.

It is also important to make our distinction between 'in-focus' and 'activated' entities more explicit. We claim that the in-focus entities are those activated entities which are likely to be continued as topics of subsequent utterances, and thus include the topic of the preceding sentence, any still relevant higher-level topics, and certain syntactically or prosodically foregrounded referents. These claims need to be more explicitly formulated and subjected to empirical verification.

To satisfy this goal, I propose to return to the field which originally inspired me to write the 1987 conference abstract that triggered the series of coauthored papers on cognitive status—computational linguistics (e.g. Grosz & Sidner; Grosz, Joshi & Weinstein; Webber, Hajicova). It will first be necessary to examine recent developments in computational discourse modelling, and then to compare the most successful approaches with our specific claims. Issues include: topic vs. cognitive status, sentence topic vs. discourse topic, hierarchy of grammatical relations, topic continuation vs. shift, role of syntactic structure and prosodic form.

## C. EXPECTED RESULTS

This project should result in a greater degree of integration of the discourse-pragmatic approach to topic, focus and cognitive status developed in my own previous research and in the collaborative research with Jeanette Gundel (my dissertation advisor at the University of Minnesota) and Ron Zacharski (a computational linguist, now at the University of Edinburgh) with current developments in theoretical semantics and computational linguistics. Progress will

be made in developing formally explicit foundations for the cross-linguistic work on forms of reference reported in Gundel, Hedberg, and Zacharski (1990, under preparation), and for the cross-linguistic research program I am currently planning on the syntactic encoding of topic and focus. The immediate outcome should be publications in the fields of semantics, discourse-pragmatics, and computational linguistics.

## REFERENCES (see also publications list)

- Cinque, G. 1990. *Types of A' Dependencies*. MIT Press, Cambridge, Mass.
- Enç, M. 1991. "The semantics of specificity." *Linguistic Inquiry* 22.1-25.
- Hajičová, E. 1987. "Focussing: A meeting point of linguistics and artificial intelligence." In P. Jorrand and V. Sgurev (eds.) *Artificial Intelligence II*. Amsterdam: North Holland, pp. 311-321.
- Heim, I. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. University of Massachusetts dissertation.
- Ludlow, P., and S. Neale. 1991. "Indefinite descriptions: In defense of Russell." *Linguistics and Philosophy* 14.171-202.
- Partee, B. 1991. "Topic, focus, and quantification." Unpublished manuscript.
- Pesetsky, D. 1987. "Wh-in-situ: movement and unselective binding." In E. Reuland and A. ter Meulen (eds.), *The Representation of (In)definiteness*. Cambridge, Mass.: MIT Press.
- Grosz, B., A. Joshi, and S. Weinstein. 1983. "Towards a unified account of definite noun phrases in discourse." In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, pp. 44-50.
- Grosz, B. and C. Sidner. 1986. "Attentions, intentions and the structure of discourse." *Computational Linguistics* 12, pp. 175-204.
- Webber, B. 1988. "Discourse deixis: reference to discourse segments". In *Proceedings of the 26th Annual Meeting of the Association for Computational Linguistics*, pp. 113-122.

## PUBLICATIONS

- \* Gundel, Jeanette, Nancy Hedberg, and Ron Zacharski. 1988. "On the generation and interpretation of demonstrative expressions," in D. Vargha (ed.), *Proceedings of the Twelfth International Conference on Computational Linguistics*, John von Neumann Society for Computing Sciences, Budapest, Hungary, pp. 216-221.
  - \* Gundel, Jeanette, Nancy Hedberg, and Ron Zacharski. 1989. "Givenness, implicature and demonstrative expressions in English discourse," B. Music, et al. (eds.), *Papers from the 25th Annual Regional Meeting of the Chicago Linguistic Society: Parasession on Language in Context*, The University of Chicago, Chicago, Illinois, pp. 89-103.
- Hedberg, Nancy. 1990. *Discourse Pragmatics and Cleft Sentences in English*, Ph.D. dissertation, University of Minnesota.
- \* Gundel, Jeanette, Nancy Hedberg, and Ron Zacharski. 1990. "Givenness, implicature, and the form of referring expressions in discourse," in M. Meacham, et al. (eds.), *Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society: Parasession on the Legacy of Grice*, University of California, Berkeley, California, pp. 442-453

Hedberg, Nancy. 1991. "Referential expletives and the extraposition analysis of clefts," in P. McFetridge (ed.) Simon Fraser University Working Papers in Linguistics, pp. 53-67.

\* Hedberg, Nancy. In Press. "On the subject-predicate structure of clefts," in M. Eid and G. Iverson (eds), Festschrift for Gerald A. Sanders. John Benjamins.

\* Gundel, Jeanette, Nancy Hedberg, and Ron Zacharski. Under preparation. "Givenness, implicature and the form of referring expressions." 76 page manuscript to be submitted to *Language* in April 1992.