The THIS-NPs Hypothesis: A Corpus-Based Investigation

Massimo Poesio† and **Natalia N. Modjeska**‡
†University of Essex, poesio@essex.ac.uk
‡University of Edinburgh, natalian@cogsci.ed.ac.uk

Abstract

We report on an analysis of the use of THIS-NPs, i.e., noun phrases with the determiner *this* and the demonstrative pronouns *this* and *these*. We test the THIS-NP hypothesis, a refinement and clarification of earlier proposals, such as (Linde, 1979; Gundel, Hedberg, and Zacharski, 1993; Passonneau, 1993), by way of a systematic analysis of the uses of these NPs in two different genres. In order to do this, we devised a reliable annotation scheme for classifying THIS-NPs in our corpus as *active* or not, in the sense of the hypothesis. 92% of THIS-NPs in our corpus were classified as referring to entities which are active in this sense. We tested three formalizations of the THIS-NP hypothesis. The version that received most empirical support is the following: THIS-NPs are used to refer to entities which are *active* but not the *backward-looking center of the previous utterance*.

1 The THIS-NPs Hypothesis

In formal semantics / pragmatics, noun phrases with the determiner *this* and the demonstrative pronouns *this* and *these* (THIS-NPs henceforth)[1] have mostly been studied for their deictic function—to refer to objects in the visual situation, and particularly to objects the speaker is pointing at (Kaplan, 1979; Jarvella and Klein, 1982; André, Poesio, and Rieser, 1999).

(1) A [pointing to his house]: *I have lived in <u>this house</u> for twenty years.*

It is, however, well-known that THIS-NPs can be used in other ways as well; and indeed, preliminary analyses of the corpus used in this study (discussed below) suggested that only about 39% of THIS-NPs were cases of visual deixis (Poesio, 2000). A second function of 'demonstrative' NPs was identified by authors such as Linde (1979), Gundel, Hedberg, and Zacharski (1993), and Passonneau (1993). These authors pointed out that *pronominal* THIS-NPs in particular[2] are often used to refer to a discourse entity other than the current discourse focus:

(2) *Dilbert arrived to work. He saw one of his colleagues. As he was trying to avoid <u>this person</u>, he quickly ducked into his cubicle.*

It is also known from work by, among others, Asher (1993) and Webber (1991) that THIS-NPs can be used to refer to abstract objects such as propositions or plans (Webber used the term DISCOURSE DEIXIS for these cases) as in the following example:

(3) *For example, binocular stereo fusion is known to take place in a specific area of the cortex near the back of the head. Patients with damage to this area of the cortex have visual handicaps but they show no obvious impairment in their ability to think. <u>This</u> suggests that stereo fusion is not necessary for thought.*
(Webber, 1991)

What the discourse and visual deixis cases, and the cases studied by Linde and Passonneau, have in common is that in all cases, the THIS-NP is used to refer to an entity which, while salient, is not the current 'topic' or 'discourse focus' (we are deliberately using these terms in a vague way here). This intuition was captured by Gundel, Hedberg, and Zacharski (1993), who developed a theory of the conditions under which referring expressions are used based on the notion of ACTIVATION HIERARCHY: a speaker's choice of expression depends on assumptions about the 'cognitive status' of the referent in the hearer's information state. Gundel *et al.*'s 'activation levels' range from TYPE IDENTIFIABILITY for indefinite NPs, to IN FOCUS for pronouns. Gundel *et al.* propose that the use of THIS-NPs, as well as of pronoun *that*[3] requires the referent to be ACTIVATED, i.e., to be represented in current short-term memory.[4]

We believe these proposals can be made at the same time more broad in their coverage and more precise by (i) specifying which entities are supposed to be 'in focus' and (ii) by being more explicit about the types of entities that can be 'in short term memory' without being 'in focus'. Our goal in this paper is to refine, clarify

[1] We will mostly avoid the use of the term 'demonstrative' as the starting point of this research is the realization that not all these uses are 'demonstrative' in Kaplan's sense (Kaplan, 1979). We are concentrating on THIS-NPs because our corpus contains very few cases of *that* noun phrases.

[2] Passonneau studied the use of *that* rather than *this*.

[3] But not of full *that* NPs, which only require the referent to have the lower 'familiar' status.

[4] In fact, for THIS-NPs, Gundel *et al.* claim that the referent has to be speaker-activated—introduced by the speaker.

and test the ideas just discussed, summarized as follows:

The THIS-NP **Hypothesis** : THIS-NPs are used to refer to entities which are ACTIVE but not IN FOCUS.

Notice that two notions to be made more precise are: what it means for an entity to be 'in focus' and what it means for it to be 'active'. We'll consider each below.

2 Background: Our previous corpus analysis work

Recent years has seen an increasing interest in corpora as a means to explore linguistic generalizations, and a correspondingly increased sophistication in the methods used. This includes better techniques for storing and annotating language corpora, based on annotation standards such as XML. It also includes techniques for measuring the RELIABILITY of a given annotation scheme (Passonneau and Litman, 1993; Carletta, 1996).

One of the major motivations for this work is that we felt that we could improve upon previous analyses of the uses of THIS-NPs by building on the results of our own previous corpus analyses of the uses of referring expressions in general and of salience (Poesio et al., 2000; Poesio, 2000). As a result of this work we had at our disposal the GNOME corpus (further discussed below) whose NPs, the anaphoric relations between them, and their visual deixis status, had been marked in a reliable way (Poesio, 2000). Secondly, we have developed methods for computing the BACKWARD-LOOKING CENTER, or CB (Grosz, Joshi, and Weinstein, 1995; Walker, Joshi, and Prince, 1998)—a well-known formalization of the notion of 'local focus'—automatically, instead of relying on hand-identification, which is notoriously problematic; and according to several definitions proposed in the literature, among which we were able to find the 'best' (i.e., those which resulted in fewer violations of the claims of Centering theory) (Poesio et al., 2000). These two previous pieces of work allowed us a more systematic exploration of the conditions under which the use of a THIS-NP was licensed, as discussed below.

2.1 Annotation Scheme

Our annotation followed a fairly systematic manual, available from the GNOME project's home page at `http://www.hcrc.ed.ac.uk/~gnome`; here, we discuss the most important details of the scheme. All units of text in the GNOME corpus that might be identified with utterances (in the Centering sense) are marked as ⟨unit⟩ elements; the attributes of such elements allow us to identify finite and non finite clauses, etc. Each NP is marked with a ⟨ne⟩ tag and with a variety of attributes capturing syntactic and semantic properties. Important attributes for our purposes are cat (specifying the type of an NP), gf specifying its grammatical function, deix (whether the object is a visual deictic reference or not) and generic (whether the NP denotes generically or not).

A separate ⟨ante⟩ element is used to mark anaphoric relations; the ⟨ante⟩ element itself specifies the index of the anaphoric expression and the type of semantic relation (e.g., identity), whereas one or more embedded ⟨anchor⟩ elements indicate possible antecedents (the presence of more than one ⟨anchor⟩ element indicates that the anaphoric expression is ambiguous). (See 4.)

(4)
```
<unit finite='finite-yes' id='u227'>
 <ne id='ne546' gf='subj'>  The drawing of
   <ne id='ne547' gf='np-compl'>the corner
   cupboard, </ne></ne>
 <unit finite='no-finite' id='u228'>or more
  probably
   <ne id='ne548' gf='no-gf'> an engraving of
      <ne id='ne549' gf='np-compl'>  it
      </ne></ne>
 </unit>,
...
</unit>
<ante current="ne549" rel="ident">
<anchor ID="ne547"></ante>
```

2.2 The Corpus

The GNOME corpus currently includes texts from three domains; texts from two domains were used in this study. The museum subcorpus consists of descriptions of museum objects and brief texts about the artists that produced them. The pharmaceutical subcorpus is a selection of leaflets providing the patients with mandatory information about their medicine. Each subcorpus contains about 6,000 NPs; in this study we used texts from the first two domains, for a total of about 3,000 NPs, including 112 THIS-NPs. As for utterances, the corpus includes about 500 sentences, and 900 finite clauses; the actual number of utterances used in the study is one of the parameters that we varied, as discussed below.

3 Methods

3.1 Clarifying 'In Focus'

Focusing and Centering Theory The notion of 'topic' or 'discourse focus' is notoriously difficult to formalize. We used as the basis for our investigation of this notion the terminology and ideas introduced in Centering Theory by Grosz, Joshi, and Weinstein (1995) and Walker, Joshi, and Prince (1998), in particular the notions of Backward-Looking Center (CB) and Preferred Center (CP). In the 'mainstream' version of Centering by Grosz, Joshi, and Weinstein (1995), it is assumed that each UTTERANCE introduces new discourse entities (or Forward-Looking Centers) into the discourse, and in so doing, updates the 'local focus'. It is further assumed that the discourse entities introduced (or better, REALIZED) by an utterance are ranked; the most highly ranked entity in an utterance is called the CP. The CB is Centering's equivalent of the notion of 'topic' or 'focus', and is defined as follows:

CB CB(U_i), the BACKWARD-LOOKING CENTER of utterance U_i, is the highest ranked element of CF(U_{i-1}) that is realized in U_i.

It should be noted that Centering provides no definition of the notions of 'ranking', 'utterance' and 'realization'; researchers using the theory have to specify their own. For her comparative study of *it* vs *that*, Passonneau developed her own notion of CB, that she called 'local center'. In previous work, we did a comparative analysis of several ways of 'filling in' Centering Theory's parameters (Poesio et al., 2000; Poesio et al., 2002). One of our results was that Passonneau's notion of local center, while predicting pronominalization well (in the sense that virtually all realizations of discourse entities that were 'local center' in Passonneau's sense were pronominalized), is very restrictive—only about 20 utterances out of 500 have a 'local center' in her sense. Two instantiations of the notion of CB gave the best (and pretty much equivalent) results. Both of these involved identifying utterances with sentences, and allowing for indirect realization of the CB; they differed in the ranking function: in one case, grammatical function (subjects rank more highly than objects that rank more highly than adjuncts) augmented with a linear disambiguation factor; in the other, Strube and Hahn's (1999) ranking function based on 'information status' (according to which hearer-old entities are ranked more highly than inferrables, which in turn are ranked more highly than hearer-new entities (Prince, 1992)).

Using Centering Theory to specify the notion of 'in focus' There are three natural way of using notions from Centering theory to formalize the idea that THIS-NPs are used to refer to entities that are not 'in focus':

1. THIS-NPs are used to refer to entities other than CB(U_i), the CB of the present utterance.
2. They are used to refer to entities other than CB(U_{i-1}), the CB of the *previous* utterance.
3. They are used to refer to entities other than CP(U_{i-1}), the most highly-ranked entity of the *previous* utterance.

We tested all three formalizations, considering all of the 'best' ways of specifying the parameters of Centering Theory found in our previous work (identifying utterances with either sentences or finite clauses; considering both ranking based on grammatical function and ranking based on 'information status'; and allowing for both direct and indirect realization).

3.2 Clarifying 'Active'

Types of Active Entities The second aspect of the THIS-NP hypothesis that needs clarification is what it means for an entity to be 'active'. The idea that we explored is that an entity is 'active' if it is in the GLOBAL FOCUS and is 'sufficiently salient'. The notion of global focus was introduced by Grosz and Sidner (1986), to characterize the entire set of entities which are in some sense part of the attentional state of the participants of a discourse. Grosz and Sidner argue that the global focus has a stack structure. In previous work, we argued that assuming a single global focus of attention is probably not right; separate structures are needed for the 'discourse' global focus and 'visual' global focus, the former having a stack-like structure and containing every discourse entity introduced by a construction algorithm similar to that of DRT[5], the latter having a situation-based structure and containing every entity in the visual scene (Poesio, 1993; Poesio, 1994; Poesio and Traum, 1997). We will make similar assumptions here.[6]

The crucial aspect of both the 'discoursal' part and the 'visual' part of the global focus for our purposes is that they contain all sort of entities beyond those explicitly mentioned or currently in the visual focus of attention; these entities are added to the global focus either as a result of immediate inference, or as a result of search procedures activated by the use of an anaphoric expressions. The term IMPLICIT FOCUS was introduced by Grosz (1977) and Sanford and Garrod (1981) to refer to this type of entities, but what is in the implicit focus has never been made explicit. Yet THIS-NPs are often used to introduce into the discourse such objects, so a detailed analysis of the possible cases is needed in order to carry out an empirical verification of the claim. Kamp and Reyle (1993) discuss in some detail one type of entity that can become 'active' or 'in the implicit focus' in this way, plural entities such as *they* in the following:

(5) *John met Mary at the movies. They had both gone to see an old French film.*

Webber (1991) and Asher (1993) analyzed in detail a second type of entity that can enter 'implicit focus' in this way, propositions, as seen in example (3). A third type of entity that can enter implicit focus, and not previously discussed (to our knowledge) in connection with uses of THIS-NPs, are *types*, in the broad sense, i.e., references to concepts whose instantiations are explicitly mentioned in the discourse. References to types take a variety of forms, the simplest among which is simply a full demonstrative like *this type* or *this kind*:

(6) *A great refinement among armorial signets was to reproduce not only the coat-of-arms but the correct tinctures; they were repeated in colour on the reverse side and the crystal would then be set in the gold bezel.*

Although the engraved surface could be used for impressions, the colours would not wear away.

The signet-ring of Mary, Queen of Scots (beheaded in 1587) is probably the most interesting example of this type; ....

[5] Similar proposals are also made in SDRT (Asher, 1993; Lascarides and Asher, 1993).

[6] (Walker, 1998) suggests that the global focus has a cache structure. We will not be concerned with this issue here, except to notice that a cache model doesn't automatically give us a definition of 'active': we still need to explain how entities in the visual situation enter in the cache.

More complex references to types refer to concepts introduced only very implicitly in the text; in this case, the process of adding the antecedent to the implicit focus, to the extent that it actually takes place (Poesio and Reyle, 2001), appears to be driven entirely by the use of the demonstrative:

(7) *The craftsmen also bent carefully over cheaper metals or glass to create the jewelry that would adorn the arm of the humble servant girl, or the ordinary, insignificant woman, and would accompany her to her final resting place. This yearning for embellishment, this special relationship between a woman and her jewelry emerges quite clearly here ....*

Certain types of ellipsis can also be considered as references to a type:

(8) *The cutouts of the first are decorated with griffins setin rectangular panels, and those of the second with large buttons ...*

We can now make more precise the THIS-NP hypothesis as follows. An entity is ACTIVE if that entity

1. is in the visual situation; or
2. is a CF of the previous utterance; or
3. can be CONSTRUCTED out of the previous utterance. An entity can be constructed out of an utterance if:
   (a) it is a plural object whose elements or subsets have been explicitly mentioned in that utterance; or
   (b) it is an abstract entity introduced by that utterance. There are two types of abstract entities:
      i. propositions
      ii. types

The Markup Scheme for Active Entities We tested our hypothesis by classifying the THIS-NPs in our corpus as active or not. In this section we discuss the markup scheme derived from the definition of 'active' above.

The annotation scheme developed in previous work (Poesio, 2000), together with the focus tracking methods developed in (Poesio et al., 2000), allowed us to to classify two of the uses of THIS-NPs identified in the literature: 'focus-shifting' uses and visual deixis. On the other hand, while developing the scheme we had found— as others before us Eckert and Strube (2001; Navarretta (2000)— that identifying the antecedents of 'discourse deictic' expressions in the broad sense (i.e., expressions referring to—typically, abstract—entities introduced in the discourse indirectly, such as propositions) is very hard, especially when the annotation produces something less than a full logical form in, say, the DRT sense (Kamp and Reyle, 1993). We had also seen in the case of visual deixis that while *identifying* the antecedent of such expressions is quite hard, *classifying* a NP as deictic is easier. This proved to be the case for discourse deixis, as well. As a result, we developed a scheme for classifying THIS-NPs that avoids these choices. The annotators are instructed to follow the decision tree below:

1. If (i) an ⟨ante⟩ elements has *not* been marked up specifying an anaphoric relation of type ident between a ⟨ne⟩ and a previous entity, and (ii) the ⟨ne⟩ is visually deictic (its deix attribute has value yes), classify it as **visual deixis**. (And thefore, active.)
2. Else, if the THIS-NP is connected by an ⟨ante⟩ elements to a previous ⟨ne⟩ by an identity relation, mark it as **anaphoric**. (This applies whether the entity is singular or plural.)
3. Else, if the THIS-NP is a plural entity which contains as elements entities previously introduced, mark it as **plural**;
4. Else, if the THIS-NP involves an elliptical reference to a previous entity (as in (8)), mark it as **ellipsis**;
5. Else, if it is a (non explicitly mentioned) temporal entity, mark it as **time**;
6. Else, if the ⟨ne⟩ is marked as generic, and its instances are concrete objects, mark it as **type**;
7. Else, if the NP refers to an abstract object 'introduced' only implicitly by the previous discourse, such a proposition or an abstract concept, mark it as **discourse deixis**;
8. Else, mark it as **problem**.

We tested the reliability of this scheme by measuring the agreement among ourselves on about 87 THIS-NPs in the corpus. We disagreed on 3 THIS-NPs and 5 were classified as problematic; with 6 possible values, we get a $\kappa = .82$, significant at the .01 level.

4 Results

4.1 The Distribution of THIS-NPs

All of the THIS-NPs in our corpus were active in the sense above. The observed distribution of THIS-NPs in our corpus is as follows:

Class	**Number (Percentage)**
Anaphora	45 (40%)
Visual Deix	28 (25%)
Discourse Deix	19 (17%)
Type	9 (8%)
Plurals	1
Ellipsis	1
Time	1
Unsure	5
Disagreement	3
Total	112

4.2 The Correlation Between Focus and THIS-NPs

For each instantiation of the notion of 'in focus', described in section 3.1, we observed some variation depending on the values of parameters, but the results were nevertheless clear-cut:

- We found between 8 and 11 violations to the hypothesis that a THIS-NP is used to refer to entities other than CB(U_{i-1}), which is therefore verified by 90-93% of THIS-NPs;
- the hypothesis that THIS-NPs are used to refer to entities other than CP(U_{i-1}) is verified by 75-80% of THIS-NPs;
- the hypothesis that a THIS-NP is used to refer to entities other than CB(U) is verified by 61-65% of THIS-NPs;

Interpreting 'not in focus' as 'not CB(U_{i-1})' leads to better empirical results. (All the differences are significant.)

4.3 Violations of the THIS-NP Hypothesis

We analyzed the 13 uses of THIS-NPs that were exceptions to the THIS-NP Hypothesis even under its best-performing version.[7] Since the entities refered to by THIS-NPs in the violation examples are IN FOCUS, we tested whether pronouns could be used in their place. These cases can be divided in three classes: (i) 5 cases in which pronominalization is possible, (ii) 3 cases in which a pronoun would be possible but awkward, and (iii) 4 cases in which a pronoun would seem rather infelicitous. We discuss some of these cases below.

In (9), a pronoun could be used instead of *this work*. (Though, a slight change in word order would make the example sound better; we would paraphrase *appears twice on this work* as *appears on it twice*.).

(9) *The fleurs-de-lis on the top two drawers indicate that the cabinet was made for Louis XIV. As it does not appear in inventories of his possessions, it may have served as a royal gift. The Sun King's portrait appears twice on this work.*

In (9), the last mention of the entity was made via a pronoun, but note that another entity is pronominalized in the same sentence, *his*, i.e., *Louis XIV*, and the focus subsequently shifts to that entity, *the Sun King* in sentence three. By using a demonstrative, rather than a pronoun, the speaker seems to prepare the listener for this shift.

In two examples, the pronominalization is possible, even if the referent is mentioned after a paragraph break:

(10) *Modeled in the form of three of laurel branches tied with a ribbon, these massive wall lights with their detailed chasing and burnishing reveals the extraordinary skill of their maker, a silversmith to Louis XV, King of France. Each wall light is slightly different, and no one model repeats another.*
These four wall lights are among eight made in 1756 ...

[7] The two instantiations of the ranking function—grammatical role and information status—both resulted in 11 violations, but they differed slightly as to which examples they produced.

In (11), the entity is also mentioned after a paragraph break. A pronoun would be possible but awkward:

(11) *Do not keep your patches if your doctor decides to stop treatment. Return them to your pharmacist who will arrange for their destruction.*
REMEMBER these patches are only for you.

This example seems to differ from (10) in that there is an implicit argument of the imperative (*you*), which perhaps is more salient than the referent of *these patches*.

Example (12)—note two THIS-NPs—is a quite interesting example for which we do not have a ready analysis:

(12) *This brooch is made of titanium ... It was made by Anne-Marie Shillitoe, an Edinburgh jeweller, in 1991. It's a good example of a modern material being used in jewelry. In fact, this piece is not one of the very earliest examples of titanium jewelry; The technique for colouring in this piece has already become quite sophisticated.*

We believe that a 'principle of variety' is at play here and interacts with the principle of always using the stronger form possible—for the first instance of *this piece* (see also (Poesio et al., 2002)). But it is not clear why a pronoun would appear awkward in the later occurrence.

In some cases, a pronoun would seem rather infelicitous, in particular if the antecedent of the THIS-NP occurs in a title:

(13) ***Has the cream any side effects?***
Most people find using this cream causes no problems when used in the right amount ...

Example (14) shows another infelicitous substitution:

(14) *This piece is also a brooch, but from an earlier period; around 1920. It's particularly interesting because of its colour scheme. Purple, white and green were the colours of the suffragette movement; women would wear a brooch like this to show solidarity or affiliation with the movement.*

We believe that the infelicity here arises from a reference to a type. As we showed in section 3.2, reference to types is one of the conditions that licenses THIS-NPs.

5 Discussion and Conclusions

We reported on an empirical investigation into the use of THIS-NPs in two genres: museum descriptions and pharmaceutical texts. The THIS-NP Hypothesis that we tested extends and clarifies previous proposals on the conditions that license the use of THIS-NPs. Specifically, we defined what it means for an entity to be 'in focus' and what it means to be 'active'. We devised and tested a reliable annotation scheme for classifying THIS-NPs as *active*. Three instantiations of the THIS-NP Hypothesis were tested on our data. The version that has the most empirical support is the following:

The THIS-NP **Hypothesis** : THIS-NPs are used to refer to entities which are ACTIVE (in the sense above) but not CB(U_{i-1}).

Acknowledgments

Part of this research was done while Massimo Poesio was at the University of Edinburgh, supported by an EPSRC Advanced Research fellowship. Natalia Modjeska is supported by the EPSRC grant No. GR/M75129.

References

André, E., M. Poesio, and H. Rieser, editors. 1999. *Proc. of the ESSLLI Workshop on Deixis, Demonstration and Deictic Belief in Multimedia Contexts*, Utrecht. University of Utrecht.

Asher, N. 1993. *Reference to Abstract Objects in English*. D. Reidel, Dordrecht.

Byron, D. and J. Allen. 1998. Resolving demonstrative anaphora in the trains-93 corpus. In *Proc. of the Second Colloquium on Discourse, Anaphora and Reference Resolution*. University of Lancaster.

Carletta, J. 1996. Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, 22(2):249–254.

Eckert, M. and M. Strube. 2001. Dialogue acts, synchronising units and anaphora resolution. *Journal of Semantics*.

Grosz, B. J. 1977. *The Representation and Use of Focus in Dialogue Understanding*. Ph.D. thesis, Stanford University.

Grosz, B. J., A. K. Joshi, and S. Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):202–225. (The paper originally appeared as an unpublished manuscript in 1986.).

Grosz, B. J. and C. L. Sidner. 1986. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3):175–204.

Gundel, J. K., N. Hedberg, and R. Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language*, 69(2):274–307.

Jarvella, R. J. and W. Klein, editors. 1982. *Speech, Place and Action - Studies in Deixis and Related Topics*. John Wiley, Chichester and New York.

Kamp, H. and U. Reyle. 1993. *From Discourse to Logic*. D. Reidel, Dordrecht.

Kaplan, D. 1979. Dthat. In P. Cole, editor, *Syntax and Semantics v. 9, Pragmatics*. Academic Press, New York, pages 221–243.

Lascarides, A. and N. Asher. 1993. Temporal interpretation, discourse relations and commonsense entailment. *Linguistics and Philosophy*, 16(5):437–493.

Linde, C. 1979. Focus of attention and the choice of pronouns in discourse. In T. Givon, editor, *Syntax and Semantics 12*. Academic Press.

Navarretta, C. 2000. Abstract anaphora resolution in Danish. In L. Dybkjaer, K. Hasida, and D. Traum, editors, *Proc. of the 1st SIGdial Workshop on Discourse and Dialogue*, pages 56–65. ACL.

Passonneau, R. and D. Litman. 1993. Feasibility of automated discourse segmentation. In *Proceedings of 31st Annual Meeting of the ACL*.

Passonneau, R. J. 1993. Getting and keeping the center of attention. In M. Bates and R. M. Weischedel, editors, *Challenges in Natural Language Processing*. Cambridge University Press, chapter 7, pages 179–227.

Poesio, M. 1993. A situation-theoretic formalization of definite description interpretation in plan elaboration dialogues. In P. Aczel, D. Israel, Y. Katagiri, and S. Peters, editors, *Situation Theory and its Applications, vol.3*. CSLI, Stanford, chapter 12, pages 339–374.

Poesio, M. 1994. Focusing in the interpretation of definite descriptions. In P. Bosch and R. van der Sandt, editors, *Proceedings of the Workshop on Focus in NLP*, Schloß Wolfsbrunnen, June. Available as Working Paper 6, IBM Deutschland.

Poesio, M. 2000. Annotating a corpus to develop and evaluate discourse entity realization algorithms: issues and preliminary results. In *Proc. of the 2nd LREC*, pages 211–218, Athens, May.

Poesio, M., H. Cheng, B. Di Eugenio, J. M. Hitzeman, and R. Stevenson. 2002. A corpus-based evaluation of centering theory. Submitted.

Poesio, M., H. Cheng, R. Henschel, J. M. Hitzeman, R. Kibble, and R. Stevenson. 2000. Specifying the parameters of Centering Theory: a corpus-based evaluation using text from application-oriented domains. In *Proc. of the 38th ACL*, Hong Kong, October.

Poesio, M. and U. Reyle. 2001. Underspecification in anaphoric reference. In E. Thijsse H. Bunt, I. van der Sluis, editor, *Proc. of the Fourth International Workshop on Computational Semantics*, pages 286–300, Tilburg, January.

Poesio, M. and D. Traum. 1997. Conversational actions and discourse situations. *Computational Intelligence*, 13(3):309–347.

Prince, E. F. 1992. The ZPG letter: subjects, definiteness, and information status. In S. Thompson and W. Mann, editors, *Discourse description: diverse analyses of a fund-raising text*. John Benjamins, pages 295–325.

Sanford, A. J. and S. C. Garrod. 1981. *Understanding Written Language*. Wiley, Chichester.

Strube, M. and U. Hahn. 1999. Functional centering–grounding referential coherence in information structure. *Computational Linguistics*, 25(3):309–344.

Walker, M. A. 1998. Centering, anaphora resolution, and discourse structure. In M. A. Walker, A. K. Joshi, and E. F. Prince, editors, *Centering in Discourse*. Oxford University Press, chapter 19, pages 401–435.

Walker, M. A., A. K. Joshi, and E. F. Prince, editors. 1998. *Centering Theory in Discourse*. Clarendon Press / Oxford.

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