Specifying the Parameters of Centering Theory: a Corpus-Based Evaluation using Text from Application-Oriented Domains

M. Poesio, H. Cheng, R. Henschel, J. Hitzeman,[†] **R. Kibble,**[§] and **R. Stevenson**[*]

University of Edinburgh, ICCS and HCRC,
`{poesio,huac,henschel}@cogsci.ed.ac.uk`
[†] The MITRE Corporation, `hitz@linus.mitre.org`
[§]University of Brighton, ITRI, `Rodger.Kibble@itri.bton.ac.uk`
[*]University of Durham, Psychology and HCRC, `Rosemary.Stevenson@durham.ac.uk`

Abstract

The definitions of the basic concepts, rules, and constraints of centering theory involve underspecified notions such as 'previous utterance', 'realization', and 'ranking'. We attempted to find the best way of defining each such notion among those that can be annotated reliably, and using a corpus of texts in two domains of practical interest. Our main result is that trying to reduce the number of utterances without a backward-looking center (CB) results in an increased number of cases in which some discourse entity, but not the CB, gets pronominalized, and viceversa.

1 MOTIVATION

Centering Theory (Grosz et al., 1995; Walker et al., 1998b) is best characterized as a 'parametric' theory: its key definitions and claims involve notions such as 'utterance', 'realization', and 'ranking' which are not completely specified; their precise definition is left as a matter for empirical research, and may vary from language to language. A first goal of the work presented in this paper was to find which way of specifying these parameters, among the many proposed in the literature, would make the claims of centering theory most accurate as predictors of coherence and pronominalization for English. We did this by annotating a corpus of English texts with the sort of information required to implement some of the most popular variants of centering theory, and using this corpus to automatically check two central claims of the theory, the claim that all utterances have a backward looking center (CB) (Constraint 1), and

the claim that if any discourse entity is pronominalized, the CB is (Rule 1). In doing this, we tried to make sure we would only use information that could be annotated reliably.

Our second goal was to evaluate the predictions of the theory in domains of interest for real applications–natural language generation, in our case. For this reason, we used texts in two genres not yet studied, but of interest to developers of NLG systems: instructional texts and descriptions of museum objects to be displayed on Web pages.

The paper is organized as follows. We first review the basic notions of the theory. We then discuss the methods we used: our annotation method and how the annotation was used. In Section 4 we present the results of the study. A discussion of these results follows.

2 FUNDAMENTALS OF CENTERING THEORY

Centering theory (Grosz et al., 1995; Walker et al., 1998b) is an 'object-centered' theory of text coherence: it attempts to characterize the texts that can be considered coherent on the basis of the way discourse entities are introduced and discussed.[1] At the same time, it is also meant to be a theory of *salience*: i.e., it attempts to predict which entities will be most salient at any given time (which should be useful for a natural language generator, since it is these entities that are most typically pronominalized (Gundel et al., 1993)).

According to the theory, every UTTERANCE in a spoken dialogue or written text introduces into the discourse a number of FORWARD-LOOKING CENTERS (CFs). CFs correspond more or less

---

[1]For a discussion of 'object-centered' vs. 'relation-centered' notions of coherence, see (Stevenson et al., 2000).

to discourse entities in the sense of (Karttunen, 1976; Webber, 1978; Heim, 1982), and can be linked to CFs introduced by previous or successive utterances. Forward-looking centers are RANKED, and because of this ranking, some CFs acquire particular prominence. Among them, the so-called BACKWARD-LOOKING CENTER (CB), defined as follows:

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

Utterance U_{i+1} is classified as a CONTINUE if $CB(U_{i+1}) = CB(U_i)$ and $CB(U_{i+1})$ is the most highly ranked CF of U_{i+1}; as a RETAIN if the CB remains the same, but it's not any longer the most highly-ranked CF; and as a SHIFT if $CB(U_{i+1}) \neq CB(U_i)$.

The main claims of the theory are articulated in terms of constraints and rules on CFs and CB.

Constraint 1: All utterances of a segment except for the 1st have exactly one CB.

Rule 1: if any CF is pronominalized, the CB is.

Rule 2: (sequences of) continuations are preferred over (sequences of) retains, which are preferred over (sequences of) shifts

Constraint 1 and Rule 2 express a preference for utterances in a text to talk about the same objects; Rule 1 is the main claim of the theory about pronominalization. In this paper we concentrate on Constraint 1 and Rule 1.

One of the most unusual features of centering theory is that the notions of utterance, previous utterance, ranking, and realization used in the definitions above are left unspecified, to be appropriately defined on the basis of empirical evidence, and possibly in a different way for each language. As a result, centering theory is best viewed as a cluster of theories, each of which specifies the parameters in a different ways: e.g., ranking has been claimed to depend on grammatical function (Kameyama, 1985; Brennan et al., 1987), on thematic roles (Cote, 1998), and on the discourse status of the CFs (Strube and Hahn, 1999); there are at least two definitions of what counts as 'previous utterance' (Kameyama, 1998; Suri and McCoy, 1994); and 'realization' can be interpreted either in a strict sense, i.e., by taking a CF to be realized in an utterance only if an NP in that utterance denotes that CF, or in a looser sense, by also counting a CF as 'realized' if it is referred to indirectly by means of a bridging reference (Clark, 1977), i.e., an anaphoric expression that refers to an object which wasn't mentioned before but is somehow related to an object that already has, as in *the vase …the handle* (see, e.g., the discussion in (Grosz et al., 1995; Walker et al., 1998b)).

3 METHODS

The fact that so many basic notions of centering theory do not have a completely specified definition makes empirical verification of the theory rather difficult. Because any attempt at directly annotating a corpus for 'utterances' and their CBs is bound to force the annotators to adopt some specification of the basic notions of the theory, previous studies have tended to study a particular variant of the theory (Di Eugenio, 1998; Kameyama, 1998; Passonneau, 1993; Strube and Hahn, 1999; Walker, 1989). A notable exception is (Tetreault, 1999), which used an annotated corpus to compare the performance of two variants of centering theory.

The work discussed here, like Tetreault's, is an attempt at using corpora to compare different versions of centering theory, but considering also parameters of centering theory not studied in this earlier work. In particular, we looked at different ways of defining the notion of utterance, we studied the definition of realization, and more generally the role of semantic information. We did this by annotating a corpus with information that has been claimed by one or the other version of centering theory to play a role in the definitions of its basic notions - e.g., the grammatical function of an NP, anaphoric relations (including information about bridging references) and how sentences break up into clauses and subclausal units– and then tried to find out the best way of specifying these notions automatically, by trying out different configurations of parameters, and counting the number of violations of the constraints and rules that would result by adopting a particular

parameter configuration.

The Data

The aim of our project, which is called GNOME and whose home page is at `http://www.hcrc.ed.ac.uk/˜gnome`, is to develop NP generation algorithms whose generality is to be verified by incorporating them in two distinct systems: the ILEX system developed at the University of Edinburgh, that generates Web pages describing museum objects on the basis of the perceived status of its user's knowledge and of the objects she previously looked at (Oberlander et al., 1998); and the ICONOCLAST system, developed at the University of Brighton, that supports the creation of patient information leaflets (Scott et al., 1998). The corpus we collected includes texts from both the domains we are studying. The texts in the museum domain consist of descriptions of museum objects and brief texts about the artists that produced them; the texts in the pharmaceutical domain are leaflets providing the patients with the legally mandatory information about their medicine. The total size of the corpus is of about 6,000 NPs. For this study we used about half of each subset, for a total number of about 3,000 NPs, of which 103 are third person pronouns (72 in the museum domain, 31 in the pharmaceutical domain) and 61 are third-person possessive pronouns (58 in the museum domain, 3 in the pharmaceutical domain).

Annotation

Previous empirical studies of centering theory typically involved a single annotator annotating her corpus according to her own subjective judgment (Passonneau, 1993; Kameyama, 1998; Strube and Hahn, 1999). One of our goals was to use for our study only information that could be annotated reliably (Passonneau and Litman, 1993; Carletta, 1996), as we believe this will make our results easier to replicate. The price we paid to achieve replicability is that we couldn't test all hypotheses proposed in the literature, especially about segmentation and about ranking. We discuss some of the problems in what follows. (The latest version of the annotation manual is available from the GNOME project's home page.)

We used eight annotators for the reliability study and the annotation.

Utterances Kameyama (1998) noted that identifying utterances with sentences is problematic in the case of multiclausal sentences: e.g., grammatical function ranking becomes difficult to measure, as there may be more than one subject. She proposed to use all and only tensed clauses instead of sentences as utterance units, and then classified finite clauses into (i) utterance units that constitute a 'permanent' update of the local focus: these include coordinated clauses and adjuncts) and (ii) utterance units that result in updates that are then erased, much as in the way the information provided by subordinated discourse segments is erased when they are popped. Kameyama called these EMBEDDED utterance units, and proposed that clauses that serve as verbal complements behave this way. Suri and McCoy (1994) did a study that led them to propose that some types of adjuncts–in particular, clauses headed by *after* and *before*–should be treated as 'embedded' rather than as 'permanent updates' as suggested by Kameyama; these results were subsequently confirmed by more controlled experiments Pearson et al. (2000). Neither Kameyama nor Suri and McCoy discuss parentheticals; Kameyama only briefly mentions relative clauses, but doesn't analyze them in detail.

In order to evaluate these definitions of utterance (sentences versus finite clauses), as well as the different ways of defining 'previous utterance', we marked up in our corpus what we called (DISCOURSE) UNITS. These include clauses, as well as other sentence subconstituents which may be treated as separate utterances, including parentheticals, preposed PPs, and (the second element of) coordinated VPs. The instructions for marking up units were in part derived from (Marcu, 1999); for each unit, the following attributes were marked:

- **utype**: whether the unit is a main clause, a relative clause, appositive, a parenthetical, etc.

- **verbed**: whether the unit contains a verb or not.

- **finite**: for verbed units, whether the verb is finite or not.

- **subject**: for verbed units, whether they have a full subject, an empty subject (expletive, as in *there* sentences), or no subject (e.g., for infinitival clauses).

The agreement on identifying the boundaries of units, using the κ statistic discussed in (Carletta, 1996), was $\kappa = .9$ (for two annotators and 500 units); the agreement on features(2 annotators and at least 200 units) was follows:

Attribute	κ Value
utype	.76
verbed	.9
finite	.81
subject	.86

NPs Our instructions for identifying NP markables derive from those proposed in the MATE project scheme for annotating anaphoric relations (Poesio et al., 1999). We annotated attributes of NPs which could be used to define their ranking, including:

- The NP type, **cat** (pronoun, proper name, etc.)

- A few other 'basic' syntactic features, **num**, **per**, and **gen**, that could be used to identify contexts in which the antecedent of a pronoun could be identified unambiguously;

- The grammatical function, **gf**;

- **ani**: whether the object denoted is animate or inanimate

- **deix**: whether the object is a deictic reference or not

The agreement values for these attributes are as follows:

Attribute	κ Value
ani	.81
cat	.9
deix	.81
gen	.89
gf	.85
num	.84
per	.9

one of the features of NPs claimed to affect ranking (Sidner, 1979; Cote, 1998) that we haven't so far been able to annotate because of failure to reach acceptable agreement is thematic roles ($\kappa = .35$).

Anaphoric information Finally, in order to compute whether a CF from an utterance was realized directly or indirectly in the following utterance, we marked up anaphoric relations between NPs, again using a variant of the MATE scheme. Theories of focusing such as (Sidner, 1979; Strube and Hahn, 1999), as well as our own early experiments with centering, suggested that indirect realization can play quite a crucial role in maintaining the CB; however, previous work, particularly in the context of the MUC initiative, suggested that while it's fairly easy to achieve agreement on identity relations, marking up bridging references is quite hard; this was confirmed by, e.g., Poesio and Vieira (1998). As a result we did annotate this type of relations, but to achieve a reasonable agreement, and to contain somehow the annotators' work, we limited the types of relations annotators were supposed to mark up, and we specified priorities. Thus, besides identity (IDENT) we only marked up three non-identity ('bridging' (Clark, 1977)) relations, and only relations between objects. The relations we mark up are a subset of those proposed in the 'extended relations' version of the MATE scheme (Poesio et al., 1999) and include set membership (ELE-MENT), subset (SUBSET), and 'generalized possession' (POSS), which includes part-of relations as well as more traditional ownership relations.

As expected, we achieved a rather good agreement on identity relations. In our most recent analysis (two annotators looking at the anaphoric relations between 200 NPs) we observed no real disagreements; 79.4% of these relations were marked up by both annotators; 12.8% by only one of them; and in 7.7% of the cases, one of the annotators marked up a closer antecedent than the other. Concerning bridges, limiting the relations did limit the disagreements among annotators (only 4.8% of the relations are actually marked differently) but only 22% of bridging references were marked in the same way by both annotators; 73.17% of relations are marked by only one or the other annotator. So reaching agreement on this information involved several discussions between annotators and more than one pass over the corpus.

Segmentation Segmenting text in a reliable fashion is still an open problem, and in addition the relation between centering (i.e., local focus shifts) and segmentation (i.e., global focus shifts) is still not clear: some see them as independent aspects of attentional structure, whereas other researchers define centering transitions with respect to segments (see, e.g., the discussion in the introduction to (Walker et al., 1998b)). Our preliminary experiments at annotating discourse structure didn't give good results, either. Therefore, we only used the layout structure of the texts as a rough indication of discourse structure. In the museum domain, each object description was treated as a separate segment; in the pharmaceutical domain, each subsection of a leaflet was treated as a separate segment. We then identified by hand those violations of Constraint 1 that appeared to be motivated by too broad a segmentation of the text.[2]

Automatic computation of centering information

The annotation thus produced was used to automatically compute utterances according to the particular configuration of parameters chosen, and then to compute the CFs and the CB (if any) of each utterance on the basis of the anaphoric information and according to the notion of ranking specified. This information was the used to find violations of Constraint 1 and Rule 1. The behavior of the script that computes this information depends on the following parameters:

utterance: whether sentences, finite clauses, or verbed clauses should be treated as utterances.

previous utterance: whether adjunct clauses should be treated Kameyama-style or Suri-style.

rank: whether CFs should be ranked according to grammatical function or discourse status in Strube and Hahn's sense

---

[2](Cristea et al., 2000) showed that it is indeed possible to achieve good agreement on discourse segmentation, but that it requires intensive training and repeated iterations; we intend to take advantage of a corpus already annotated in this way in future work.

realization: whether only direct realization should be counted, or also indirect realization via bridging references.

4 MAIN RESULTS

The principle we used to evaluate the different configurations of the theory was that the best definition of the parameters was the one that would lead to the fewest violations of Constraint 1 and Rule 1. We discuss the results for each principle.

Constraint 1: All utterances of a segment except for the 1st have precisely one CB

Our first set of figures concerns Constraint 1: how many utterances have a CB. This constraint can be used to evaluate how well centering theory predicts coherence, in the following sense: assuming that all our texts are coherent, if centering were the only factor behind coherence, all utterances should verify this constraint. The first table shows the results obtained by choosing the configuration that comes closest to the one suggested by Kameyama (1998): utterance=finite, prev=kameyama, rank=gf, realization=direct. The first column lists the number of utterances that satisfy Constraint 1; the second those that do not satisfy it, but are segment-initial; the third those that do not satisfy it and are not segment-initial.

	CB	Segment Initial	NO CB	Total Number
Museum	132	35	245	412
Pharmacy	158	13	198	369
Total	290	48	443	791

The previous table shows that with this configuration of parameters, most utterances do not satisfy Constraint 1 in the strict sense even if we take into account text segmentation (admittedly, a very rough one). If we take sentences as utterances, instead of finite clauses, we get fewer violations, although about 25% of the total number of utterances are violations:

	CB	Segment Initial	NO CB	Total Number
Museum	120	22	85	227
Pharmacy	152	8	51	211
Total	272	30	136	438

Using Suri and McCoy's definition of previous utterance, instead of Kameyama's (i.e., treating adjuncts as embedded utterances) leads to a slight improvement over Kameyama's proposal but still not as good as using sentences:

	CB	Segment Initial	NO CB	Total Number
Museum	140	35	237	412
Pharmacy	167	14	188	369
Total	307	49	425	791

What about the finite clause types not considered by Kameyama or Suri and McCoy? It turns out that we get better results if we do not treat as utterances relative clauses (which anyway always have a CB, under standard syntactic assumptions about the presence of traces referring to the modified noun phrase), parentheticals, clauses that occur in subject position; and if we treat as a single utterance matrix clauses with empty subjects and their complements (as in *it is possible that John will arrive tomorrow*).

	CB	Segment Initial	NO CB	Total Number
Museum	143	35	153	331
Pharmacy	161	14	159	334
Total	304	49	312	665

But by far the most significant improvement to the percentage of utterances that satisfy Constraint 1 comes by adopting a looser definition of 'realizes', i.e., by allowing a discourse entity to serve as CB of an utterance even if it's only referred to indirectly in that utterance by means of a bridging reference, as originally proposed by Sidner (1979) for her discourse focus. The following sequence of utterances explains why this could lead to fewer violations of Constraint 1:

(1)   (u1) These "egg vases" are of exceptional quality: (u2) basketwork bases support egg-shaped bodies (u3) and bundles of straw form the handles, (u4) while small eggs resting in straw nests serve as the finial for each lid. (u5) Each vase is decorated with inlaid decoration: …

In (1), u1 is followed by four utterances. Only the last of these directly refers to the set of egg vases introduced in u1, while they all contain implicit references to these objects. If we adopt this looser notion of realization, the figures improve dramatically, even with the rather restricted set of relations on which our annotators agree. Now the majority of utterances satisfy Constraint 1:

	CB	Segment Initial	NO CB	Total Number
Museum	225	35	71	331
Pharmacy	174	14	146	334
Total	399	49	217	665

And of course we get even better results by treating sentences as utterances:

	CB	Segment Initial	NO CB	Total Number
Museum	171	17	39	227
Pharmacy	168	7	36	211
Total	339	24	75	438

It is important, however, to notice that even under the best configuration, at least 17% of utterances violate the constraint. The (possibly, obvious) explanation is that although coherence is often achieved by means of links between objects, this is not the only way to make texts coherent. So, in the museum domain, we find utterances that do not refer to any of the previous CFs because they express generic statements about the class of objects of which the object under discussion is an instance, or viceversa utterances that make a generic point that will then be illustrated by a specific object. In the following example, the second utterance gives some background concerning the decoration of a particular object.

(2)   (u1) On the drawer above the door, gilt-bronze military trophies flank a medallion portrait of Louis XIV. (u2) In the Dutch Wars of 1672 - 1678, France fought simultaneously against the Dutch, Spanish, and Imperial armies, defeating them all. (u3) This cabinet celebrates the Treaty of Nijmegen, which concluded the war.

Coherence can also be achieved by explicit coherence relations, such as **EXEMPLIFICATION** in the following example:

(3)   (u1) Jewelry is often worn to signal membership of a particular social group. (u2) The Beatles brooch shown previously is another case in point:

Rule 1: if any NP is pronominalized, the CB is

In the previous section we saw that allowing bridging references to maintain the CB leads to fewer violations of Constraint 1. One should not, however, immediately conclude that it would be a good idea to replace the strict definition of 'realizes' with a looser one, because there is, unfortunately, a side effect: adopting an indirect notion of realizes leads to more violations of Rule 1. Figures are as follows. Using utterance=s, rank=gf, realizes=direct 22 pronouns violating Rule 1 (9 museum, 13 pharmacy) (13.4%), whereas with realizes=indirect we have 38 violations (25, 13) (23%); if we choose utterance=finite, prev=suri, we have 23 violations of rule 1 with realizes=direct (13 + 10) (14%), 32 with realizes=indirect (21 + 11) (19.5%). Using functional centering (Strube and Hahn, 1999) to rank the CFs led to no improvements, because of

the almost perfect correlation in our domain between subjecthood and being discourse-old. One reason for these problems is illustrated by (4).

(4)   (u1) A great refinement among armorial signets was to reproduce not only the coat-of-arms but the correct tinctures; (u2) they were repeated in colour on the reverse side (u3) and the crystal would then be set in the gold bezel.

They in u2 refers back to *the correct tinctures* (or, possibly, *the coat-of-arms*), which however only occurs in object position in a (non-finite) complement clause in (u1), and therefore has lower ranking than *armorial signets*, which is realized in (u2) by the bridge *the reverse side* and therefore becomes the CB having higher rank in (u1), but is not pronominalized.

In the pharmaceutical leaflets we found a number of violations of Rule 1 towards the end of texts, when the product is referred to. A possible explanation is that after the product has been mentioned sentence after sentence in the text, by the end of the text it is salient enough that there is no need to put it again in the local focus by mentioning it explicitly. E.g., *it* in the following example refers to the cream, not mentioned in any of the previous two utterances.

(5)   (u1) A child of 4 years needs about a third of the adult amount. (u2) A course of treatment for a child should not normally last more than five days (u3) unless your doctor has told you to use it for longer.

5   DISCUSSION

Our main result is that there seems to be a trade-off between Constraint 1 and Rule 1. Allowing for a definition of 'realizes' that makes the CB behave more like Sidner's Discourse Focus (Sidner, 1979) leads to a very significant reduction in the number of violations of Constraint 1.[3] We also noted, however, that interpreting 'realizes' in this way results in more violations of Rule 1. (No differences were found when functional centering was used to rank CFs instead of grammati-

---

[3]Footnote 2, page 3 of the intro to (Walker et al., 1998b) suggests a weaker interpretation for the Constraint: 'there is no more than one CB for utterance'. This weaker form of the Constraint does hold for most utterances, but it's almost vacuous, especially for grammatical function ranking, given that utterances have at most one subject.

cal function.) The problem raised by these results is that whereas centering is intended as an account of both coherence and local salience, different concepts may have to be used in Constraint 1 and Rule 1, as in Sidner's theory. E.g., we might have a 'Center of Coherence', analogous to Sidner's discourse focus, and that can be realized indirectly; and a 'Center of Salience', similar to her actor focus, and that can only be realized directly. Constraint 1 would be about the Center of Coherence, whereas Rule 1 would be about the Center of Salience. Indeed, many versions of centering theory have elevated the CP to the rank of a second center.[4]

We also saw that texts can be coherent even when Constraint 1 is violated, as coherence can be ensured by other means (e.g., by rhetorical relations). This, again, suggests possible revisions to Constraint 1, requiring every utterance either to have a center of coherence, or to be linked by a rhetorical relation to the previous utterance.

Finally, we saw that we get fewer violations of Constraint 1 by adopting sentences as our notion of utterance; however, again, this results in more violations of Rule 1. If finite clauses are used as utterances, we found that certain types of finite clauses not previously discussed, including relative clauses and matrix clauses with empty subjects, are best not treated as utterances. We didn't find significant differences between Kameyama and Suri and McCoy's definition of 'previous utterance'. We believe however more work is still needed to identify a completely satisfactory way of breaking up sentences in utterance units.

---

[4]This separation among a 'center of coherence' and a 'center of salience' is independently motivated by considerations about the division of labor between the text planner and the sentence planner in a generation system; see, e.g., (Kibble, 1999).

References

S.E. Brennan, M.W. Friedman, and C.J. Pollard. 1987. A centering approach to pronouns. In *Proc. of the 25th ACL*, pages 155–162, June.

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

H. H. Clark. 1977. Inferences in comprehension. In D. Laberge and S. J. Samuels, editors, *Basic Process in Reading: Perception and Comprehension*. Lawrence Erlbaum.

S. Cote. 1998. Ranking forward-looking centers. In M. A. Walker, A. K. Joshi, and E. F. Prince, editors, *Centering Theory in Discourse*, chapter 4, pages 55–70. Oxford.

D. Cristea, N. Ide, D. Marcu, and V. Tablan. 2000. Discourse structure and co-reference: An empirical study. In *Proc. of COLING*.

B. Di Eugenio. 1998. Centering in italian. In M. A. Walker, A. K. Joshi, and E. F. Prince, editors, *Centering Theory in Discourse*, chapter 7, pages 115–138. Oxford.

B. J. Grosz, A. K. Joshi, and S. Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):202–225.

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

I. Heim. 1982. *The Semantics of Definite and Indefinite Noun Phrases*. Ph.D. thesis, University of Massachusetts at Amherst.

M. Kameyama. 1985. *Zero Anaphora: The case of Japanese*. Ph.D. thesis, Stanford University.

M. Kameyama. 1998. Intra-sentential centering: A case study. In M. A. Walker, A. K. Joshi, and E. F. Prince, editors, *Centering Theory in Discourse*, chapter 6, pages 89–112. Oxford.

L. Karttunen. 1976. Discourse referents. In J. McCawley, editor, *Syntax and Semantics 7 - Notes from the Linguistic Underground*. Academic Press.

R. Kibble. 1999. Cb or not Cb? centering applied to NLG. In *Proc. of the ACL Workshop on discourse and reference*.

D. Marcu. 1999. Instructions for manually annotating the discourse structures of texts. Unpublished manuscript, USC/ISI, May.

J. Oberlander, M. O'Donnell, A. Knott, and C. Mellish. 1998. Conversation in the museum: Experiments in dynamic hypermedia with the intelligent labelling explorer. *New Review of Hypermedia and Multimedia*, 4:11–32.

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

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

J. Pearson, R. Stevenson, and M. Poesio. 2000. Pronoun resolution in complex sentences. In *Proc. of AMLAP*, Leiden.

M. Poesio and R. Vieira. 1998. A corpus-based investigation of definite description use. *Computational Linguistics*, 24(2):183–216, June.

M. Poesio, F. Bruneseaux, and L. Romary. 1999. The MATE meta-scheme for coreference in dialogues in multiple languages. In M. Walker, editor, *Proc. of the ACL Workshop on Standards and Tools for Discourse Tagging*, pages 65–74.

D. Scott, R. Power, and R. Evans. 1998. Generation as a solution to its own problem. In *Proc. of the 9th International Workshop on Natural Language Generation*, Niagara-on-the-Lake, CA.

C. L. Sidner. 1979. *Towards a computational theory of definite anaphora comprehension in English discourse*. Ph.D. thesis, MIT.

R. Stevenson, A. Knott, J. Oberlander, and S McDonald. 2000. Interpreting pronouns and connectives. *Language and Cognitive Processes*, 15.

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

L. Z. Suri and K. F. McCoy. 1994. RAFT/RAPR and centering: A comparison and discussion of problems related to processing complex sentences. *Computational Linguistics*, 20(2):301–317.

J. R. Tetreault. 1999. Analysis of syntax-based pronoun resolution methods. In *Proc. of the 37th ACL*, pages 602–605, University of Marylan, June. ACL.

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

M. A. Walker. 1989. Evaluating discourse processing algorithms. In *Proc. ACL-89*, pages 251–261, Vancouver, CA, June.

B. L. Webber. 1978. A formal approach to discourse anaphora. Report 3761, BBN, Cambridge, MA.