Centering Theory in natural text: a large-scale corpus study

Annemarie Friedrich **Alexis Palmer**
Department of Computational Linguistics
Saarland University, Saarbrücken, Germany
{afried,apalmer}@coli.uni-saarland.de

Abstract

We present an extensive corpus study of Centering Theory (CT), examining how adequately CT models coherence in a large body of natural text. A novel analysis of transition bigrams provides strong empirical support for several CT-related linguistic claims which so far have been investigated only on various small data sets. The study also reveals genre-based differences in texts' degrees of entity coherence. Previous work has shown unsupervised CT-based coherence metrics to be unable to outperform a simple baseline. We identify two reasons: 1) these metrics assume that some transition types are more coherent and that they occur more frequently than others, but in our corpus the latter is not the case; and 2) the original sentence order of a document and a random permutation of its sentences differ mostly in the fraction of entity-sharing sentence pairs, exactly the factor measured by the baseline.

1 Introduction

Centering Theory (CT) models the degree of local coherence between adjacent utterances within paragraphs with respect to patterns of entity mentions and the choice of referring expressions (Grosz et al., 1995). CT regards a text as a sequence of utterances $U_1, U_2, \ldots, U_n$. The entities mentioned (*realized*) in an *utterance* U_i are referred to as *centers* and make up its set of *forward-looking centers* $CF(U_i)$, which are ranked according to their salience, i.e., how likely they are to be mentioned in the following utterance. Each utterance is assigned a single *backward-looking center* $CB(U_i)$, defined as the highest-ranked element of $CF(U_{i-1})$ also realized in U_i, and a *preferred center* $CP(U_i)$, the highest-ranked center of U_i. CT identifies different types of transitions between adjacent utterances and assumes that the types have different degrees of coherence. We define these as in Table 1 (following Brennan et al. (1987) and Kameyama (1986)), with NOCB as the case U_i and U_{i-1} have no shared center, so U_i has no CB.

Contributions. We present the largest corpus study of CT to date, confirming and consolidating previous results by investigating multiple predictions of the theory using a uniform implementation of CT over a large amount (14096 sentences) of natural text. CT has inspired various automatic methods for measuring coherence (Lapata and Barzilay (2005), Elsner and Charniak (2011), among others). In this paper we aim not to improve upon these methods, but rather to better understand when and why they work and what the reasons are for their limitations. Our main finding is that analysis of natural text, which can be assumed to be coherent, fails to support some of the predictions of CT which inform automatic coherence evaluation methods. Many adjacent sentences do not mention the same entities, and there is no clear preference for certain

	COHERENCE $CB(U_i) = CB(U_{i-1})$	¬ COHERENCE $CB(U_i) \neq CB(U_{i-1})$
SALIENCE $CB(U_i) = CP(U_i)$	CONTINUE	SMOOTH-SHIFT
¬ SALIENCE $CB(U_i) \neq CP(U_i)$	RETAIN	ROUGH-SHIFT

$CB(U_i) = undef.$	NOCB
$CB(U_{i-1}) = undef$ and $CB(U_i) = def.$	ESTABLISH

Table 1: **Definitions of Centering Theory transitions** used in this study.

CT transition types. The coherence experiments we study compare documents in their original orderings to randomly sentence-permuted texts; our analysis shows that the main difference is an increased number of NoCB transitions. This explains why no simple CT-based coherence metric outperforms a baseline that simply considers whether two adjacent sentences mention the same entity. However, some linguistic claims made by CT hold up when treated as patterns observable in large amounts of data rather than single texts: we show that transitions have different preferences for the transitions that follow them, supporting the assumptions of the RETAIN-SHIFT pattern (Brennan et al., 1987), and that cheapness and salience are the most important factors for transition preferences (Strube and Hahn, 1999; Kibble, 1999).

Related work. Previous empirical studies of CT use small corpora of limited domains; for example, Poesio et al. (2000) and Poesio et al. (2004) inspect the effect of various parameter settings on the percentage of utterances that obey the constraints and rules of CT, using about 500 sentences from pharmaceutical leaflets and descriptions of museum objects. While this study sheds light on many aspects of CT, pharmaceutical leaflets exhibit a special structure, and museum object descriptions belong to a limited domain. Karamanis et al. (2009) extend this corpus with news and other texts and report results on about 4500 sentences of natural text. Similar but smaller quantitative studies on various aspects of CT have been conducted by Hurewitz (1998), on about 400 spoken and written transitions, by Di Eugenio (1998) for Italian, by Strube and Hahn (1999) in order to evaluate functional information structure as a ranking function for centers, and more recently by Maat and Sanders (2009) for Dutch and by Taboada (2008) for spoken text.

documents (total)	535
news (479), essay (41), letters (15)	
sentences (total)	14,096
paragraphs (total)	5,605
one-sentence paragraphs	1,405
avg. # of sentences per par.	3.02
all CT transitions	13,561
transitions within paragraphs	8,491

Table 2: **Corpus statistics.**

2 Data and implementation of CT

This section describes the data our corpus study is based on, and the decisions we made when implementing our version of CT.

Data. Our corpus is the portion of the Wall Street Journal for which OntoNotes 4.0 (Hovy et al., 2006) provides manual coreference annotations.[1] For syntactic information, constituent parses from Penn TreeBank 2.0 (Marcus et al., 1993; Vadas and Curran, 2007) are automatically converted to dependency parses using the tool from Johansson and Nugues (2007).

OntoNotes annotates both *identical* coreference as in ‘She had a *good suggestion* and *it* was accepted’ and *appositive* coreference, as in ‘*Washington*, *the capital city*’. Additionally, we assume coreference between two nouns if they share a lemma.

We use only documents labeled as *news*, *essay* or *letters* by Webber (2009), omitting the other genres due to low frequency. Table 2 gives a statistical overview of the corpus.

Implementation. Implementing CT requires some parameter-setting; we follow the findings of Poesio et al. (2000), taking sentences as the unit

[1] Coreference information is necessary to appropriately link entities across utterances; the same data set (using OntoNotes 2.9) is used in (Louis and Nenkova, 2010).

of **utterances**, and identify **paragraphs** by empty lines in the source data. We consider nouns and personal and possessive pronouns to **realize entities**. Elements of $CF(U_n)$ are ranked by *grammatical function*, with SUBJ > OBJ > OTHER. After ranking subject and object of the main clause, remaining entity mentions are ranked according to their *surface order*. Nouns modifying other nouns directly follow their heads.

3 Corpus analysis

We investigate several aspects of CT on our corpus and implementation; here we describe these aspects and the results of our analysis.

3.1 Rule 1: pronominalization

Our first finding is strong support for **Rule 1** of CT (Grosz et al., 1995), which expresses the intuition that only the most salient entities of an utterance are pronominalized. According to this rule, if the CB of an utterance is not pronominalized, neither should any other entity in the utterance. The corpus contains 5907 utterances with non-pronominal CBs. 64.7% of these contain no pronouns at all. 4.9% contain expletive pronouns, and 26.4% contain pronouns that have antecedents in the same sentence such as in example (1). We do not regard these cases as violations. Only 4.0% of all utterances with a non-pronominal CB have pronouns with antecedents outside the sentence, violating Rule 1.

(1) More broadly, $[_{CB}$Mr. Boren$]$ hopes that Panama will shock *Washington* out of *its* fear of using military power. (wsj0771)

3.2 Preferences for transition types

It has been proposed that different CT transitions contribute differently to the perceived degree of coherence of a text. In their algorithm for centering and pronoun binding, Brennan et al. (1987) assume a simple ranking of transitions with respect to their assumed degree of coherence: CONTINUE > RETAIN > SMOOTH-SHIFT > ROUGH-SHIFT. Figure 1 shows that these four transitions occur with similar frequencies in our corpus, both within and between paragraphs. Hence, it is not the case that the transitions that are more coherent according to Brennan are in fact used more often by authors, even in perfectly coherent texts.

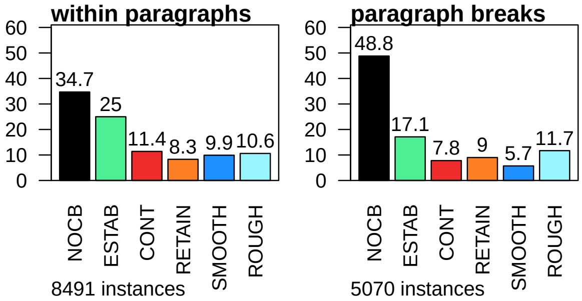

Figure 1: **Distribution of CT transitions** in percent.

The percentage of NOCB transitions is much higher at paragraph boundaries than within paragraphs. However, more than 50% of paragraph-initial sentences mention an entity realized at the end of the previous paragraph, with the salient transitions (see Table 1) being less likely than the non-salient transitions. This indicates that new paragraphs usually change focus when they relate to previous centers. The relatively high percentage of ESTABLISH is due to the high frequency of NOCB, after which only NOCB or ESTABLISH can follow. The *essay+letters* subset of documents has more NOCB transitions than *news* (within paragraphs 43.4% versus 32.6%), indicating that entity coherence matters more in news text, and that *essay+letters* more often reference entities indirectly (not shown in Figure 1).

3.3 Kibble (2001): reformulation of Rule 2

Kibble (2001) suggests that the standard preference ordering of transitions is unmotivated and suggests ranking transition types by considering the interaction of several criteria. Our analysis supports his claims that *cheapness* and *salience* are most important in determining transition preferences, and *cohesion* is of least importance. His proposal draws motivation from natural language generation work (Kibble, 1999), but no corpus study has previously been done. Here we consider only within-paragraph transitions, under the assumption that they do not contain topic changes. Of these transitions, 65.3% have a CB. Of those with a CB: 52.1% have *salient* CBs (i.e., the CB is also the CP of the utterance); 53.9% are *cheap* (the CB of an utterance matches the CP of the *previous* utterance); and 30.2% have the same CB as the previous utterance (*cohesion*).

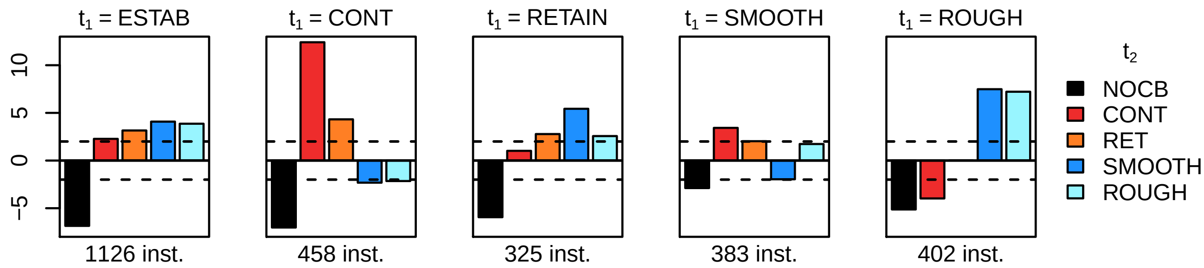

Figure 2: **Residuals of χ^2-tests**: based on 4291 within-paragraph pairs. In 1597 pairs, t_1 is NoCB (not shown).

3.4 Transition bigram distributions

Rule 2 of CT as originally formulated by Grosz et al. (1995) states: *"Sequences of continuation are preferred over sequences of retaining, which are in turn preferred over sequences of shifting."* Thus, in this part of the study, we ask: are there any patterns regarding sequences of transitions?

We first compute $P_{bigram} = P(t_2|t_1)$, the distributions of transitions t_2 conditioned on their previous transition t_1, using the within-paragraph subset. We want to find out whether some transition pairs occur more often than expected. As some transitions are much more frequent than others, it is hard to draw conclusions directly from looking at P_{bigram}. Instead, we apply a statistical test: we compare each P_{bigram} to $P_{unigram}$, the overall distribution of transitions that follow some other transition. We compute Pearson's χ^2-test and plot the residuals in Figure 2. Residuals with absolute value > 2 are considered major contributors to significance, indicated by the dashed lines. We find significant differences between P_{bigram} and $P_{unigram}$ for each t_1 ($p < 0.01$).

We conclude the following: (a) Although NoCB is by far the most frequent transition type overall, it occurs less often than expected after any of the other five transition types. This indicates that there are entity-coherent portions of texts, where multiple utterances share and develop centers. A similar intuition has been proposed, but not tested, in the framework of Rhetorical Structure Theory (Knott et al., 2001). (b) There is a strong tendency that after a Continue transition, there will be another Continue or a Retain. Once a segment strongly focuses on a center, it is likely that the center will be kept. Shifting is less likely after Continue than expected by the overall distribution of transitions. (c) After Retain, there are not more Continues than expected, but many more Smooth-Shifts. This supports the assumption of the Retain-Shift pattern, which may signal introduction of a new discourse topic (Brennan et al., 1987; Strube and Hahn, 1999).

The transition following a Smooth-Shift tends to continue with or retain the new center, and after Rough-Shifts, more shifts and far fewer Continues than expected occur. From this, we can conclude that salience influences the author's choice for the next transition: when the current center is salient (as in Continue and Smooth-Shift), there is a tendency to keep the center. When the current center is not salient (as in Retain and Rough-Shift), there is a tendency to shift to a new center. This observation again supports the principle of *cheapness* (Strube and Hahn, 1999). In fact, all transition pairs with the largest positive residual are classified as cheap by Strube and Hahn (1999), while most other pairs are considered expensive. The only exception is Rough-Rough, which is considered expensive by Strube and Hahn but has a large positive residual. The most frequently occurring bigrams excluding NoCB-bigrams are Estab-Cont (213), Estab-Rough (190), Cont-Cont (181), Estab-Smooth (180) and Rough-Rough (103).

4 Centering-based coherence metrics

Using a small corpus from a limited domain, Karamanis et al. (2009) find that CT-based metrics have no success in improving upon a baseline dubbed M.NoCB, which simply uses whether two sentences share a center or not. In order to shed light on the utility of CT for coherence assessment, we replicate their *information ordering* experiment using our corpus data. The assumption underlying this experimental method is that the original sentence order (OSO) of a text should be scored higher by coherence metrics than any permutation of the text's sentences. We exhaustively enumerate all permutations for texts with fewer than 10 sentences and use a random sample of 1,000,000 permutations for each longer text (only the first 30 sentences of each text are considered in this case). CT transitions are computed for the OSO and for each permutation. We use noun lemma matching as well as gold-standard coreference chains. This oracle style of entity reference resolution has also been applied by Lapata and Barzilay (2005), among others.

We compare the following CT-based metrics described by Karamanis et al. (2009): M.NOCB counts NOCB transitions; M.KP counts NOCBs as well as all violations of cheapness, coherence and salience (following Kibble and Power (2000)); M.BFP prefers the ordering with the most CONTINUEs; if equal, the one with most RETAINs etc.[2] (following Brennan et al. (1987)); and M.CHEAP sums up violations of cheapness (following Strube and Hahn (1999)). Karamanis et al. (2009) do not consider NOCBs to be violations of cheapness. As the permutations in general contain more NOCBs than the OSO, they contain fewer violations of cheapness. Using absolute counts of violations of cheapness hence leads to classification error rates worse than chance. We count NOCBs also as violations of cheapness, and hence actually test a combination of continuity and cheapness.

We score the OSO and the permutations with each CT-based metric. In order to evaluate the performance of metric M, the *classification error rate* is computed as $better(M, OSO) + 0.5 * equal(M, OSO)$ where $better(M, OSO)$ is the percentage of permutations scored higher than the OSO, and $equal(M, OSO)$ is the percentage of permutations achieving the same score as the OSO. The lower the classification error rate of M, the better its performance. A rate greater than 50% means that the metric scores the permutation higher than the OSO in the majority of cases.

Table 3 shows the classification error rates we obtained on our data set, with the results of Karamanis et al. (2009) for comparison. The texts in our data set contain 26 sentences on average (Karamanis et al.: 8 sentences per text on average). Similar to their findings, M.NOCB is among the best-performing metrics, but in contrast to their results, we find that M.KP performs best, though not significantly differently from M.NOCB, and M.BFP performs worst in our experiments. This is in line with the results presented in Section 3.1, and indicates that a feature-based approach to CT-based coherence metrics, using indicators such as coherence, salience and cheapness, works better than the more coarse-grained transition-based approach.

METRIC	Our corpus	Karamanis
M.KP†	0.219∗	0.561
M.NOCB	0.226∗	0.217
M.CHEAP†	0.265	0.698
M.BFP	0.285	0.280
documents	535	542
sentences	14,096	4,380

Table 3: **Classification error rates**. ∗ Rates do **not** differ significantly ($p < 0.01$) according to a two-sided binomial test. † Considers NOCB to be a violation of cheapness.

	NOCB	ESTAB	CONT	RET	SMOOTH	ROUGH
OSO	7.0	3.7	2.3	2.5	2.3	4.0
permutations	10.2	3.4	0.8	1.3	0.7	2.1
difference	*+3.2*	*-0.3*	*-1.5*	*-1.2*	*-1.6*	*-1.9*

Table 4: **Average frequencies** of transition types per document.

Table 4 shows the average frequencies of the transition types per document both for the original documents and for their permutations. When comparing the numbers for OSOs and permutations, the numbers of the other transition types are all reduced to approximately the same extent. The major difference between OSOs and permutations is that the latter have more NOCBs, which explains the fact that M.NOCB could not be outperformed by the CT-based coherence metrics proposed in the literature to date.

On the 56 documents of *letters+essay*, lower

[2]This metric doesn't make use of ESTAB.

classification error rates are achieved (0.055 for M.NoCB). This is surprising given that the original documents contain more NoCBs than *news* text. A possible explanation is that these texts change their focus on different entities as they progress, while news texts keep referring to the same set of entities, and hence a larger number of acceptable orderings is possible.

We conclude that CT-based coherence metrics are attractive as they are completely unsupervised and domain-independent, but they seem to reach their upper bound at a classification error rate of around 20% on our corpus. However, other CT-inspired coherence metrics such as the entity-grid model (Lapata and Barzilay, 2005; Barzilay and Lapata, 2008) achieve much better performance by means of a supervised training step.

5 Conclusion, discussion, future work

We have presented the largest study of CT based on natural text to date. While CT adequately describes some linguistic patterns according to our study, these can only be found by analysing collections of texts, not single texts. We show that the different transition types are used in natural text with no clear preference and that genre may play a role in choice of coherence device. We find strong empirical support for CT's claims regarding pronominalization of entity mentions, as well as for the claim that cheapness and salience play a greater role than cohesion.

Our replication of previous information ordering experiments indicates that it is not possible to leverage CT transitions to design unsupervised domain-independent metrics measuring the coherence of normal-length texts due to sparsity.[3] No metric significantly outperforms a baseline that uses only the number of NoCB transitions.

Miltsakaki and Kukich (2000) find ROUGH-SHIFTS to be a predictor of incoherence for student essays, but these are a domain very different from our corpus of financial news written by professional journalists. We suggest that if it is clear to the reader which entity is referred to in an utterance, it may even be easy to process a large number of shifts, as example (2) shows.

[3] Initial experiments trying to leverage the bigram patterns found in Section 3.4 were not successful as bigram distributions suffer even more from sparsity than unigrams.

(2) (a) Two dozen scientists reported results with variations of the *experiments* [...] by Fleischmann and Pons.
(b) The $[_{CB}$*experiments*$]$ involve plunging the two *electrodes* into "heavy" water. (ESTABLISHMENT)
(c) When an electric current is applied to the $[_{CB}$*electrodes*$]$, the heavy *water* did begin to break up, or dissociate. (ROUGH-SHIFT)
(d) Ordinarily the breakup of the $[_{CB}$*water*$]$ would consume almost all of the electrical energy. (ROUGH-SHIFT)
(wsj1550, shortened)

This kind of discourse organization, in which an element introduced in an utterance (*rheme*) is used as the *theme* (known information) in the next utterance, has been described as *simple linear textual progression* (Danes, 1974) or *focus-topic chaining* (Smith, 2003). We argue that shifting centers may be what makes a text interesting to readers.

CT focuses on entity-based coherence. However, in many perfectly coherent text passages no direct coreference links are found. Consider example (3):

(3) (a) Competition has glutted the market with both skins and coats, driving prices down.
(b) The animal-rights movement hasn't helped sales. (NoCB)
(c) Warm winters over the past two years have trimmed demand, too, furriers complain. (NoCB) (wsj1586)

Some utterance pairs are instead connected via reference to the same situations or events, which is one direction for future research; Christensen et al. (2013) and Hou et al. (2013) propose promising approaches to identifying mentions referring to the same situtation or event. Other interesting directions include investigating relationships between entity coherence and other coherence devices such as discourse relations (Louis and Nenkova, 2010); and combining CT-based features with, e.g., features reflecting semantic content or licensing particular syntactic realizations. Finally, further analysis of CT on a greater variety of genres is warranted.

Acknowledgments

We thank the anonymous reviewers, Manfred Pinkal, Michaela Regneri and Caroline Sporleder for helpful comments. This research was supported in part by the MMCI Cluster of Excellence, and the first author is supported by an IBM PhD Fellowship.

References

Regina Barzilay and Mirella Lapata. 2008. Modeling local coherence: An entity-based approach. *Computational Linguistics*, 34(1):1–34.

Susan E Brennan, Marilyn W Friedman, and Carl J Pollard. 1987. A centering approach to pronouns. In *Proceedings of the 25th annual meeting on Association for Computational Linguistics*, pages 155–162. Association for Computational Linguistics.

Janara Christensen, Stephen Soderland Mausam, and Oren Etzioni. 2013. Towards coherent multi-document summarization. In *Proceedings of NAACL-HLT*, pages 1163–1173.

Frantisek Danes. 1974. Functional sentence perspective and the organization of the text. *Papers on functional sentence perspective*, 113.

Barbara Di Eugenio. 1998. Centering in Italian. In Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors, *Centering Theory in Discourse*, chapter 7, pages 115–137. Clarendon Press, Oxford.

Micha Elsner and Eugene Charniak. 2011. Extending the Entity Grid with Entity-Specific Features. In *HLT*, pages 125–129.

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

Yufang Hou, Katja Markert, and Michael Strube. 2013. Global inference for bridging anaphora resolution. In *HLT-NAACL*, pages 907–917.

Eduard Hovy, Mitchell Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. 2006. OntoNotes: the 90% solution. In *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, NAACL-Short '06, pages 57–60, Stroudsburg, PA, USA. Association for Computational Linguistics.

Felicia Hurewitz. 1998. A quantitive look at discourse coherence. In Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors, *Centering Theory in Discourse*, chapter 14, pages 273–292. Clarendon Press, Oxford.

Richard Johansson and Pierre Nugues. 2007. Extended constituent-to-dependency conversion for english. In *Proc. of the 16th Nordic Conference on Computational Linguistics (NODALIDA)*, pages 105–112.

Megumi Kameyama. 1986. A property-sharing constraint in centering. In *Proceedings of the 24th annual meeting on Association for Computational Linguistics*, pages 200–206. Association for Computational Linguistics.

Nikiforos Karamanis, Chris Mellish, Massimo Poesio, and Jon Oberlander. 2009. Evaluating centering for information ordering using corpora. *Computational Linguistics*.

Rodger Kibble and Richard Power. 2000. An integrated framework for text planning and pronominalisation. In *Proceedings of the first international conference on Natural language generation-Volume 14*, pages 77–84. Association for Computational Linguistics.

Rodger Kibble. 1999. Cb or not Cb? Centering Theory applied to NLP. In *Proceedings of the ACL workshop on Discourse and Reference Structure*.

Rodger Kibble. 2001. A reformulation of Rule 2 of centering theory. *Computational Linguistics*, 27(4):579–587, December.

Alistair Knott, Jon Oberlander, Mick ODonnell, and Chris Mellish. 2001. Beyond elaboration: The interaction of relations and focus in coherent text. *Text representation: linguistic and psycholinguistic aspects*, pages 181–196.

M Lapata and R Barzilay. 2005. Automatic evaluation of text coherence: Models and representations. *International Joint Conference On Artificial ...*.

Annie Louis and Ani Nenkova. 2010. Creating local coherence: An empirical assessment. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 313–316. Association for Computational Linguistics.

Henk Pander Maat and Ted Sanders. 2009. How grammatical and discourse factors may predict the forward prominence of referents: two corpus studies. *Linguistics*, 47(6):1273–1319.

Mitchell P. Marcus, Mary Ann Marcinkiewicz, and Beatrice Santorini. 1993. Building a large annotated corpus of English: the Penn TreeBank. *Comput. Linguist.*, 19(2):313–330, June.

Eleni Miltsakaki and Karen Kukich. 2000. The role of centering theory's rough-shift in the teaching and evaluation of writing skills. In *Proceedings of the 38th Annual Meeting on Association for Computational Linguistics*, pages 408–415, Stroudsburg, PA, USA.

Massimo Poesio, Hua Cheng, Renate Henschel, Janet Hitzemann, Rodger Kibble, and Rosemary Stevenson. 2000. Specifying the parameters of Centering Theory: a corpus-based evaluation using text from application-oriented domains. In *Proceedings of the 38th ACL*.

Massimo Poesio, Barbara Di Eugenio, Rosemary Stevenson, and Janet Hitzeman. 2004. Centering: A Parametric Theory and Its Instantiations. *Computational Linguistics*, 30 (3)(2000).

Carlota S Smith. 2003. *Modes of discourse: The local structure of texts*, volume 103. Cambridge University Press.

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

Maite Taboada. 2008. Deciding on units of analysis within Centering Theory. *Corpus Linguistics and Linguistic Theory*, 4(1):63–108.

David Vadas and James Curran. 2007. Adding noun phrase structure to the penn treebank. In *ANNUAL MEETING-ASSOCIATION FOR COMPUTATIONAL LINGUISTICS*, volume 45, page 240.

Bonnie Webber. 2009. Genre distinctions for discourse in the Penn TreeBank. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2*, pages 674–682. Association for Computational Linguistics.