

# Towards a Literary Machine Translation: The Role of Referential Cohesion

**Rob Voigt**

Center for East Asian Studies  
Stanford University  
robvoigt@stanford.edu

**Dan Jurafsky**

Department of Linguistics  
Stanford University  
jurafsky@stanford.edu

## Abstract

What is the role of textual features above the sentence level in advancing the machine translation of literature? This paper examines how referential cohesion is expressed in literary and non-literary texts and how this cohesion affects translation. We first show in a corpus study on English that literary texts use more dense reference chains to express greater referential cohesion than news. We then compare the referential cohesion of machine versus human translations of Chinese literature and news. While human translators capture the greater referential cohesion of literature, Google translations perform less well at capturing literary cohesion. Our results suggest that incorporating discourse features above the sentence level is an important direction for MT research if it is to be applied to literature.

## Introduction

The concept of literary machine translation might seem at first to be a near-contradiction in terms. The field of machine translation has traditionally aimed its sights at the translation of technical or otherwise informative texts, with the strongest focus on newswire and other informative texts relevant to the goals of government funders.

Nevertheless, the prospect of literary MT is appealing. Human translation of literary texts is an extremely time- and money-intensive task, but one that is a crucial element of the global system of transcultural literary exchange. From a technical standpoint, since “by definition, literature is the art that uses language” (Chapman 1973), literary translation represents perhaps the strongest formulation of the machine translation problem. Jonathan Slocum, writing in 1985, essentially rejects the idea of literary MT altogether, noting

that it is serendipitous for technical MT that emphasis is placed on semantic fidelity to the source text, whereas literary translation must take into account larger considerations such as style with which “computers do not fare well.” Given the explosion of statistical methodologies in MT, are we now at a point where we can hope to begin tackling some of the questions associated with a potential literary machine translation?

This problem is severely understudied. Regardless of the plausibility (or even desirability) of eventually using MT to produce full-fledged translations of literary texts, a serious consideration of the unique difficulties posed by literary translation may well serve to push forward our computational understanding of literature and the language of translation.

In particular, literary translation seems to demand that we address larger-scale textual features beyond the sentence-level approach commonly employed by contemporary MT systems. There is a substantial body of work by scholars in the field of translation studies addressing greater-than-sentence-level textual features from a linguistic and literary-theoretical perspective, and this existing work can offer conceptual understanding and a parallel vocabulary with which to discuss progress in this regard in machine translation.

Eugene Nida (1964), for example, used the terms “formal equivalence” and “dynamic equivalence” to differentiate between translations aiming to replicate the form of their source and those aiming to replicate the source text's effects on its readers. Hatim and Mason (1995) brought the “seven standards of textuality” set forth by Beaugrande and Dressler (1981) into the translation studies context as metrics for evaluating the “expectation-fulfilling” or “expectation-defying” outcome of a translated text.

Cohesion is defined by Beaugrande and Dressler as “concern[ing] the ways in which the components of the textual world, i.e., the configuration of concepts and relations which underlie the surface text, are mutually accessible and relevant.” Cohesion considers the limited human capacity for storing the “surface materials” of a text long enough to relate them semantically during the act of reading.

We therefore propose to study referential cohesion (Halliday and Hasan 1976), the relation between co-referring entities in a narrative, as an important component of cohesion. Referential cohesion has a significant literature in natural language processing (Grosz et al. 1995, Mani et al. 1998, Marcu 2000, Karamanis et al. 2004, Kibble and Power 2004, Elsner and Charniak 2008, Barzilay and Lapata 2008, inter alia) as does automatic coreference resolution, which has significantly increased in accuracy in recent years (Bengston and Roth 2008, Haghighi and Klein 2009, Haghighi and Klein 2010, Rahman and Ng 2011, Pradhan et al. 2011, Lee et al. 2011).

We formulate and test two hypotheses in this position paper: First, we anticipate that given stylistic considerations and their fundamental narrative function, prose literary texts are inherently “more cohesive” than news. Second, in light of the aforementioned necessity for “dynamic equivalence” in the literary translation, we anticipate that current machine translation systems, built with newswire texts in mind, will be less successful at conveying cohesion for literary texts than for news.

## 2. Investigating Literary Cohesion

Our first preliminary experiment examines how referential cohesion in literary texts differs from news text by examining coreference in a monolingual English-language corpus, without considering machine-translated texts.

We created a small corpus of twelve short stories for comparison with twelve recent long-form news stories from the New York Times, Wall Street Journal, The Atlantic, and the news blog The Daily Beast. The stories chosen were written by a variety of authors: Isaac Asimov, J.D. Salinger, Edgar Allen Poe, Tobias Wolff, Vladimir Nabokov, Sir Arthur Conan Doyle, Shirley Jackson, Jack London, Mark Twain, Willa Cather, Ambrose

Bierce, and Stephen Crane – in the interest of avoiding over-specificity to any particular genre or style. The corpus thus included 12 short stories with 76,260 words and 12 news articles with 23,490 words, for a total corpus size of 24 documents and 99,750 words.

We used standard publicly-available NLP tools to process the corpus. We used the Stanford CoreNLP suite<sup>1</sup> to tokenize and sentence-split both the human and MT versions of each text and then to run the multi-pass sieve coreference resolution system described in Lee et al. (2011).

This system works by making multiple passes over the text, first doing recall-oriented mention extraction, then resolving coreference through a series of sieves moving from highest to lowest precision. This system is state-of-the-art, with a B<sup>3</sup> F1 score of 68.9 with no gold mention boundaries on the CoNLL 2011 shared task test set. Nevertheless, it is likely to introduce some measure of noise into our results.

For the rest of the paper we use the term “cluster” to refer to clusters agglomerated by the system that co-refer to the same entity, and “mention” to refer to individual instances of each entity in the text.

	Clusters per 100 Tokens	Mentions per 100 Tokens	Density: Mentions per Cluster
Short Stories	3.6	19.3	5.4
News Text	3.9	15.0	3.9

Table 1. Cohesion as measured by coreference in literary vs. non-literary texts. Figures given are the overall average across all documents.

Table 1 reports the numbers of clusters and mentions (normalized per 100 tokens). The literary texts had the same number of clusters (entities) as the news texts (one-tailed t-test,  $p = 0.080$ ), albeit with a trend towards fewer clusters in literature. But literary text had more mentions ( $p < 0.001$ ), and a higher number of mentions per cluster ( $p < 0.001$ ) than the news texts.

The results of this preliminary study suggest that the literary text tended to discuss the same number of entities as the non-fiction, but to

<sup>1</sup> Available online at [nlp.stanford.edu/software/corenlp.shtml](http://nlp.stanford.edu/software/corenlp.shtml)

Suddenly, the nurse resorted to direct measures. She seized the boy's upper arm in one hand and dipped the other in the milk. She dashed the milk across his lips, so that it dripped down cheeks and receding chin.

...

Always, his frightened eyes were on her, watching, watching for the one false move. She found herself soothing him, trying to move her hand very slowly toward his hair, letting him see it every inch of the way, see there was no harm in it. And she succeeded in stroking his hair for an instant.

...

Instead, she turned on the night light and moved the bed. The poor thing was huddled in the corner, knees up against his chin, looking up at her with blurred and apprehensive eyes.

...

She looked down at those eager brown eyes turned up to hers and passed her hands softly through his thick, curly hair.

Figure 1. Human markup of cohesion throughout Asimov's "The Ugly Little Boy." Recurring entities are color-coded: red is the character Edith Fellowes, grey is her hands, blue is the character Timmie, light green is his eyes, dark green is his chin, yellow is his hair, and magenta is the milk. This sample contains 149 words and 7 recurring entities with a total of 29 mentions.

mention each entity more often. In other words, literary text uses more dense reference chains as a way of creating a higher level of cohesion.

Figures 1 and 2 provide representative examples, hand-labeled for coreference, to offer a qualitative intuition for this difference in cohesion. In the literary example in Figure 1 we find seven recurring entities with an average of 4.1 mentions each. In the news example in Figure 2 we find seven recurring entities but only 3.0 average mentions, resulting in qualitatively less dense reference chains in the news sample.

Our results are consistent with Biber (1988), whose factor analysis study found that fiction tended to have a high frequency of third-person personal pronouns. This is true in our corpus; third-person pronouns occur 57.7% more in the fiction as opposed to the non-fiction texts (16.9 vs 10.7 occurrences per 100 words). But even when we count ignoring third-person pronouns, we found a greater density of mentions per cluster for literature than for news (4.0 vs 3.3,  $p = 0.015$ ). The result that literature seems to have more to say about each entity thus extends and

Two studies have found that weight-loss operations worked much better than the standard therapies for Type 2 diabetes in obese and overweight people whose blood sugar was out of control. Those who had surgery, which stapled the stomach and rerouted the small intestine, were much more likely to have a complete remission of diabetes, or to need less medicine, than people who were given the typical regimen of drugs, diet and exercise.

...

The new studies, published on Monday by The New England Journal of Medicine, are the first to rigorously compare medical treatment with these particular stomach and intestinal operations as ways to control diabetes. Doctors had been noticing for years that weight-loss operations, also called bariatric surgery, could sometimes get rid of Type 2 diabetes. But they had no hard data.

...

One of the studies, conducted at the Catholic University in Rome, compared two types of surgery with usual medical treatment.

Figure 2. Human markup of cohesion throughout a NYT news article. Recurring entities are color-coded, similar to the above. This sample contains 152 words and 7 recurring entities with a total of 21 mentions.

explains Biber's finding that literature has more third-person pronouns.

While our results are suggestive, they remain preliminary. A more detailed follow-up will need to look at the specific realization of the mentions and the kind of local coherence relations that link them (Althaus et al. 2004, Poesio et al. 2004, Barzilay and Lapata 2008, Elsnér and Charniak 2008), and to investigate the different aspects of referential chains with larger corpora and more varying genres.

### 3. MT Success at Conveying Cohesion

To evaluate the impact of this difference in expressed cohesion on machine translation systems, we compared coreference output between human and machine translations of literary and informative texts from Chinese. For this task we chose a small dataset of sixteen short stories in Chinese by the early 20th-century author Lu Xun (鲁迅) and their corresponding English translations by Gladys Yang. We chose Lu Xun for his prominence as the "father of modern Chinese literature" and vernacular style, and because Yang's English translations are widely accepted as being

of high quality by the literary community. For comparison to news text, we chose a series of six long-form articles from the magazine *Sinorama* and their corresponding English reference translations in the LDC's "Chinese English News Magazine Parallel Text" corpus (LDC2005T10). These magazine texts were chosen because the brief newswire texts often used in MT evaluation are too short to allow for meaningful textual-level comparisons of this sort. Thus our corpus contained 16 human-translated short stories with 90,712 words, 16 machine-translated short stories with 82,475 words, 6 human-translated magazine articles with 45,310 words, and 6 machine-translated magazine articles with 39,743 words, for a total size of 44 documents and 258,240 words.

We used Google Translate as our MT translation engine, first because the large web-based resources behind that system might help to mitigate the inevitable complication of domain specificity in the training data, and second because of its social position internationally as the most likely way average readers might encounter machine translation.

We first used Google Translate to produce machine translations of both the literary and magazine texts, and then used the Lee et al. (2011) coreference system in Stanford CoreNLP as described above to evaluate cohesion on both the human and machine English translations. As acknowledged in the prior section, automatic coreference is likely to introduce some amount of noise, but there is no reason to think that this noise would be biased in any particular direction for MT.

Results from the coreference analysis of the literary and magazine texts are shown in Table 2. The results in the two rows labeled "Human" substantiate our findings from the previous section. The human translations of the short stories have a significantly ( $p = 0.003$ ) higher referential chain density (5.2) than the human translations of the magazine pieces (4.2). Translators, or at least Gladys Yang in these translations, seem to act similarly to source-text writers in creating more dense referential chains in literature than in non-fiction genres.

In order to study the success of machine translation in dealing with cohesion, we took the human translations as a gold standard in each case, using this translation to normalize the number of clusters and mentions to the length of the reference

	Clusters per 100 Tokens	Mentions per 100 Tokens	Density: Mentions per Cluster
<i>Short Story</i>			
Human	3.7	19.0	5.2
Machine	4.1	16.4	3.8
<i>Magazine</i>			
Human	3.9	16.0	4.2
Machine	3.9	14.0	3.7

Table 2. Cohesion as measured by coreference in human and machine translations of Lu Xun short stories and *Sinorama* magazine articles. The first two columns are normalized to the length of the human "gold" translations, and figures given are the overall average across all documents.

documents to address the length variance caused by the MT system.

The results in Table 2 show little underclustering for the MT output. The number of clusters (entities) in the machine translations (4.1 and 3.9) do not differ from the human translations (3.7 and 3.9), ( $p = 0.074$ ), although there is a trend toward underclustering for literature.

The main difference we see is in referential chain density (mentions per cluster). Whereas these experiments reconfirm the trend towards more mentions per cluster in literature than informative text, referential chains in the MT output do not differ between the two genres. The machine translation only captures 79.4% (13,846 vs. 17,438) of the human-translated mentions in the literary texts.

In the literary genre the automatic coreference system finds more than one additional mention per cluster in the human translations as compared to MT ( $p < 0.001$ ), while in the magazine case the human and MT translations are the same, though there is a similar trend towards less dense referential chains in MT output ( $p = 0.055$ ).

#### 4. Examples and Discussion

It is worth first acknowledging the somewhat surprising ability of MT to maintain cohesion in both domains. The fact that a system operating almost exclusively on a sentence-by-sentence basis is able to maintain upwards of three-quarters of the mentions in the difficult and linguistically distant context of Chinese-to-English



MT is remarkable in and of itself, and speaks to the relative success of modern MT. There is, of course, no guarantee that these mentions found by the coreference system are in fact all the correct ones, so the true figure is likely somewhat lower, but a qualitative examination of the system's output shows that they are largely accurate.

What is actually causing the discrepancies in cohesion noted above as regards our two domains? Below we look at some specific cases of reduced cohesion in our results from the Lu Xun story "Flight to the Moon." In these examples the human translator was forced to rely on greater-than-sentence-level features of the text to effect an appropriately cohesive translation that the MT system was unable to convey.

#### Zero Anaphora

Zero anaphora is a well-documented and common linguistic phenomena in Chinese (Li and Thompson 1979, Huang 1989). Kim (2000) investigated subject drop in Chinese and English, finding that English overtly specifies subjects in 96% of cases, while the figure for Chinese is only 64%, and a significant amount of prior work has focused on the computational identification and resolution of zero anaphora in Chinese (see Yeh and Chen 2001, Converse 2006, Zhao and Ng 2007, Kong and Zhou 2010). The following example sentences demonstrate this difficulty.

<p><u>Human Translation</u></p> <p>When the big game was finished <b>they</b> ate wild boars, rabbits and pheasants. <b>He</b> was such a fine archer, <b>he</b> could shoot as much as <b>he</b> pleased.</p> <p><u>Machine Translation</u></p> <p>Later large animal shot down, ate wild boar, rabbit pheasant; shooting method and high strength, many as <b>you</b> want.</p> <p><u>Original Chinese</u></p> <p>后来大动物射完了，就吃野猪兔山鸡射法又高强，要多少有多少。</p>
---

Figure 3. Reduced cohesion via zero anaphora in MT output. Relevant mentions are hand-annotated in bold.

In a qualitative analysis of our results, problems such as these were by far the most common cause of cohesion errors, and as the reader will notice, they often lead to an output that loses crucial elements for maintaining the cohesion

of the narrative, such as in this case the distinction between the husband/wife couple, "they," and the husband individually, "he."

#### Inconsistent Reference

Having no process for maintaining consistency of reference to entities in the narrative, the following non-consecutive coreferencing sentences illustrate how in the MT version of the text the cohesiveness of the "hen" cluster in the original is lost.

<p><u>Human Translation</u></p> <p>"Who are you? Why have you shot my best black laying <b>hen</b>?"</p> <p>"What! A <b>hen</b>?" he echoed nervously. "I thought it was a <b>wood pigeon</b>."</p> <p>"Imagine mistaking a <b>hen</b> for a <b>wood pigeon</b>!"</p> <p>"I am Yi." While saying this he saw that his arrow had pierced the <b>hen's</b> heart, killing it outright.</p> <p>"What about this <b>hen</b>?"</p> <p>"<b>She</b> was my best: <b>she</b> laid me an egg every day."</p> <p>"I'll give you these for your <b>hen</b>"</p> <p><u>Machine Translation</u></p> <p>"Who are you what? How good black <b>hen</b> shot to the top of my house?"</p> <p>"Ah! <b>Chicken</b>? I only said a <b>wood pigeon partridge</b>," he said in dismay.</p> <p>"<b>hens</b> do not know, will be treated as the <b>wood pigeon partridge</b>"</p> <p>"I Yi Yi." He said, to see his shot arrows, is being consistently the heart of the <b>hen</b>, of course, died</p> <p>"<b>Chicken</b> how to do it?"</p> <p>"Lost my best <b>hen</b> every day to lay eggs."</p> <p>"they brought lost your <b>chicken</b>."</p> <p><u>Original Chinese</u></p> <p>"你是谁哪?怎么把我家的顶好的黑母鸡射死了?"</p> <p>"阿呀!鸡么?我只道是一只鹌鹑。"他惶恐地说。</p> <p>"连母鸡也不认识,会当作鹌鹑!"</p> <p>"我就是夷羿。"他说着,看看自己所射的箭,是正贯了母鸡的心,当然死了</p> <p>"这鸡怎么办呢?"</p> <p>"这是我家最好的母鸡,天天生蛋。"</p> <p>"就拿来赔了你的鸡"</p>
---

Figure 4. Reduced cohesion via inconsistent reference in MT output. Relevant mentions are hand-annotated in bold.

The reader will notice that in the original Chinese, *ji* (鸡, lit. "chicken") is used here as a

shortened version of *muji* (母鸡, lit. “hen”) in colloquial speech, which the human translator clearly notes and translates each mention consistently to maintain cohesion. Similarly, being that number is not explicitly marked in Chinese, the MT system translates *lian muji* (连母鸡, lit. “even hen”) as “hens” instead of catching that here 母鸡 refers back to the entity being discussed.

### *De* (的) Drops

It is common in Chinese for the noun head of a nominalization formed by the particle *de* (的) to be implicit, yet in many cases the human translator will add it for clarity and, presumably, to maintain cohesion.

<p><u>Human Translation</u> "There are those who know <b>my name</b>."</p> <p><u>Machine Translation</u> "Some people is one to know."</p> <p><u>Original Chinese</u> “有些人是一听就知道的。” Exist some people be one hear then know NOM</p>
---

Figure 5. Reduced cohesion via *de* dropping in MT output. Relevant mentions are hand-annotated in bold.

This phenomenon reminds of translation theorist Mona Baker's (1996) concept of “explicitation”: “an overall tendency to spell things out rather than leave them implicit in translation.” Indeed, Olohan and Baker (2000) demonstrate this empirically using the Translational English Corpus, finding a strong tendency in translated texts to explicitly mark the “that”-connective following words such as “say,” “tell,” “promise,” and so on where it could have been omitted.

## 5. Implications and Future Research

We found in two separate analyses that literary texts had more dense reference chains than informative texts. This result supports our hypothesis that literary texts are indeed more cohesive in general than informative texts; that is to say, the stylistic and narrative demands of literature lead to prose being more cohesively “about” its subjects than news. It remains to replicate this experiment on a large, carefully sampled cross-genre corpus to confirm these preliminary findings, perhaps integrating a more

complex measure of cohesion as in Barzilay and Lapata (2008).

We also found that MT systems had difficulty in conveying the cohesion in literary texts. Of course these results are preliminary and may be confounded by the nature of the training data used by modern MT systems. The uses of Google Translate as an MT system and longer-form magazine articles as our informative texts were aimed at mitigating these concerns to some extent, but for now these results primarily serve as indicative of the need for further research in this area.

Cohesion, as well, is only one of the seven “standards of textuality” put forth by Beaugrande and Dressler (1981) and taken up by Hatim and Mason (1997) in the translation context. Some of these have an existing literature addressing their computational identification and analysis (eg. Morris and Hirst 1991), in which cases we might apply existing methods to identify genre effects in literary text. For others, such as situationality, it remains to investigate appropriate computational analogues for large-scale automatic analysis and application to literary text. Studies addressing relevant textual-level concerns in literature show increasing promise, such as Elson et al. (2010)'s work in automatically extracting social networks from fiction.

Once these sorts of genre effects in literature are more clearly understood, they can be addressed on a large scale for comparisons between machine- and human-translated literary texts in the manner carried out in this paper, in order to identify further potential stumbling blocks for machine translation on the textual level as regards literary texts. Our preliminary work as presented here suggests, at the very least, the potential value and necessity of such analyses if we are to make progress towards a true literary machine translation.

### Acknowledgements

Thanks to Heeyoung Lee for help with the coreference system, three anonymous reviewers for their careful reading and helpful comments, and the U.S. Department of Education for the Foreign Language and Area Studies grant that helped fund this research.

## References

- Althaus, Ernst, Nikiforos Karamanis, and Alexander Koller. 2004. Computing locallycoherent discourses. In *ACL*.
- Baker, Mona. 1996. Corpus-based translation studies: The challenges that lie ahead. In *Terminology, LSP and Translation: Studies in language engineering*. John Benjamins, Amsterdam.
- Barzilay, Regina and Mirella Lapata. 2008. Modeling Local Coherence: An Entity-based Approach. *Computational Linguistics*, 34(1).
- Beaugrande, Robert and Wolfgang Dressler. 1981. *Introduction to Text Linguistics*. Longman, London.
- Bengston, E. and Dan Roth. 2008. Understanding the value of features for coreference resolution. In *EMNLP*.
- Biber, Douglas. 1988. *Variation across speech and writing*. Cambridge University Press, Cambridge.
- Chapman, Raymond. 1973. *Linguistics and Literature*. Edward Arnold, London.
- Converse, Susan. 2006. *Pronominal anaphora resolution for Chinese*. Ph.D. thesis.
- Elsner, Micha and Eugene Charniak. 2008. Coreference-inspired Coherence Modeling. In *Proceedings of ACL 2008*.
- Elson, David, Nicholas Dames, and Kathleen McKeown. 2010. Extracting social networks from literary fiction. In *ACL*.
- Grosz, Barbara, Aravind K. Joshi, and Scott Weinstein. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2).
- Haghighi, Aria and Dan Klein. 2009. Simple coreference resolution with rich syntactic and semantic features. In *EMNLP*.
- Haghighi, Aria and Dan Klein. 2010. Coreference resolution in a modular, entity-centered model. In *HLT-NAACL*.
- Halliday, M. A. K. and Ruqaiya Hasan. 1976. *Cohesion in English*. Longman, London.
- Hatim, Basil and Ian Mason. 1997. *The Translator as Communicator*. Routledge, London.
- Huang, James C.-T. 1989. Pro drop in Chinese, a generalized control approach. In O, Jaeggli and K. Safir, editors, *The Null Subject Parameter*. D. Reidel Dordrecht.
- Karamanis, Nikiforos, Massimo Poesio, Chris Mellish, and Jon Oberlander. 2004. Evaluating centering-based metrics of coherence for text structuring using a reliably annotated corpus. In *ACL*.
- Kibble, Rodger and Richard Power. 2004. Optimizing Referential Coherence in Text Generation. *Computational Linguistics* 30(4).
- Kim, Young-Joo. 2000. Subject/object drop in the acquisition of Korean: A cross-linguistic comparison. *Journal of East Asian Linguistics*, 9(4).
- Kong, Fang and Guodong Zhou, 2010. A Tree Kernel-based Unified Framework for Chinese Zero Anaphora Resolution. In *EMNLP*.
- Lee, Heeyoung, Yves Peirsman, Angel Chang, Nathanael Chambers, Mihai Surdeanu, Dan Jurafsky. Stanford's Multi-Pass Sieve Coreference Resolution System at the CoNLL-2011 Shared Task. 2011. In *Proceedings of the CoNLL-2011 Shared Task*.
- Li, Charles and Sandra Thompson. 1979. Third-person pronouns and zero-anaphora in Chinese discourse. *Syntax and Semantics*, 12:311-335.
- Ma, Xiaoyi. 2005. Chinese English News Magazine Parallel Text. LDC2005T10.
- Mani, Inderjeet, Barbara Gates, and Eric Bloedorn. 1998. Using Cohesion and Coherence Models for Text Summarization. In *AAAI*.
- Marcu, Daniel. 2000. *The Theory and Practice of Discourse Parsing and Summarization*. MIT Press, Cambridge, MA.
- Morris, Jane and Graeme Hirst. 1991. Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text. *Computational Linguistics*, 17(1).
- Nida, Eugene. 1964. *Towards a Science of Translating*. Brill, Leiden.
- Olohan, Maeve and Mona Baker. 2000. Reporting that in translated English: Evidence for subconscious processes of explicitation? *Across Languages and Cultures* 1.
- Poesio, Massimo, Rosemary Stevenson, Barbara di Eugenio, and Janet Hitzeman, 2004. Centering: A Parametric theory and its instantiations. *Computational Linguistics*, 30(3).
- Pradhan, Sameer, Lance Ramshaw, Mitchell Marcus, Martha Palmer, Ralph Weischedel, and Nianwen Xue. 2011. CoNLL-2011 Shared Task: Modeling Unrestricted Coreference in OntoNotes. In *CoNLL*.

Rahman, Altaf and Vincent Ng. 2011. Coreference resolution with world knowledge. In *ACL*.

Slocum, Jonathan. 1985. A Survey of Machine Translation: its History, Current Status, and Future Prospects. *Computational Linguistics*, 11(1).

Zhao, Shanheng and Hwee Tou Ng. 2007. Identification and Resolution of Chinese Zero Pronouns: A Machine Learning Approach. In *Proceedings of EMNLP CoNLL Joint Conference*.