Centering in Mandarin Chinese – A Preliminary Study

As is well known by those who have sought to employ Centering Theory [2, 7] in an examination of some aspect of discourse analysis (e.g., topic tracking, textual coherence, anaphora resolution/choice of referring expression), there are parameters of the theory for which appropriate settings must be found.

In a recent paper, Poesio [5] even characterized Centering Theory as not one theory but a cluster of theories, in which each theory had its own particular value settings for each of the parameters.

These open parameters include the questions of:

1. the definition of utterance

2. how entities in the forward-looking centers (C_f) list

   for an utterance should be ranked

3. when an entity in an utterance can be said to be "realized" (e.g., do "bridging" references count?)

4. what can be counted as a "previous utterance"

Centering was initially examined with respect to a non-pro-drop, "western" language, namely English. But the theory has also been examined cross-linguistically, for example in the pro-drop languages Italian [1], Japanese [3, 4], and Turkish [6]. Each of these languages has a relatively rich system of inflectional morphology.

In contrast, Mandarin Chinese, another non-western, pro-drop language, has little morphology, and no inflectional morphology (other than pluralizers for pronouns) on the various forms of nominal referring expressions. In addition to the characterization of Chinese as pro-drop and as having little morphology, it has also been called a topic-prominent language.

We propose in this paper to present the results of some preliminary investigations into the proper settings for one of Centering Theory's parameters: the question of how entities on the forward-looking center list (the C_f list) should be ranked.

Since the question of ranking depends, necessarily, on what size has been chosen for an utterance unit, that parameter "setting" will be addressed indirectly as a separate condition. But the primary focus of the experiments will be on the C_f list ordering.

The data for this study are taken from the Chinese TreeBank[1]. Although this corpus is admittedly a biased sample of the Chinese language, it has the virtue of being already annotated with not only labels on phrasal consitutents, but also functional labels such as topic and subject. In addition the nulls are explicity expressed as either *pro* or as *PRO*.

References

[1] Barbara Di Eugenio. Centering in italian. In Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors, *Centering Theory in Discourse*, pages 115–137. Clarendon Press, New York, 1998.

[1] See **http://www.ldc.upenn.edu/ctb**

[2] Barbara J. Grosz, Aravind K. Joshi, and Scott Weinstein. Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21(2):203–225, 1995.

[3] Masayo Iida. Discourse coherence and shifting centers in japanese texts. In Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors, *Centering Theory in Discourse*, pages 161–180. Clarendon Press, New York, 1998.

[4] Megumi Kameyama. Intrasentential centering: A case study. In Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors, *Centering Theory in Discourse*, pages 89–112. Clarendon Press, New York, 1998.

[5] M. Poesio, H. Cheng, R. Henschel, J. Hitzeman, R. Kibble, and R. Stevenson. Specifying the parameters of centering theory: a corpus-based evaluation using text from application-oriented domains. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, pages 400–407, Hong Kong, 1-8 October 2000. Association for Computational Linguistics.

[6] Ümit Deniz Turan. Ranking forward-looking centers in turkish: Universal and language-specific properties. In Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince, editors, *Centering Theory in Discourse*, pages 139–160. Clarendon Press, New York, 1998.

[7] Marilyn A. Walker, Aravind K. Joshi, and Ellen F. Prince. *Centering Theory in Discourse*. Clarendon Press, New York, 1998.