

Discourse Markers and the Structure of Discourse

Livia Polanyi and Martin van den Berg

In discourse processing systems, computational linguists relied upon the presence of discourse markers (“cue words”) such as *so*, *well*, and *anyway* to signal discourse phenomena such as a move to or return from an embedded discourse units even before developing rigorous definitions of what elementary or complex discourse units might be (Marcu [1998]. Asher and Lascarides [2003], Polanyi et al [2004]). In the paper proposed here, we intend to demonstrate how the Linguistic Discourse Model (LDM) provides an account of how some markers in English (*and* and *that’s it*) and Japanese (*yo*, *kudasai*) signal how the propositional content within their scope continues, completes, or interrupts an ongoing discourse activity.

Surprisingly, within the growing literature in the fields of sociolinguistics, pragmatics and, increasingly, formal semantics in discussions of the meaning and function of these terms, their discourse structuring functions are all but ignored.. What little has been published in the linguistics literature concerning the distribution of markers relative to the structural configuration of the discourses in which they occur is vague, i.e. Levinson (1983, 87), quoted in Lenk (1998, 39), states that *many words in English, and no doubt most languages, ... indicate the relationship ... the utterance that contains them is a response to ... some portion of the prior discourse*, while Schiffrin(1987) asserts that discourse markers are *sequentially dependent elements that bracket units of talk*” (1987 , 31) [and that] *“occur at the boundaries of units”* (1987, 37) and Davis suggests that the placement of *yo* at the end of the texts of Japanese online reviews containing sequences of sentiment marked sentences *rhetorically makes sense* (113) .

Without adequate theoretical mechanisms that provide an account of how language above the sentence functions, little progress in understanding the discourse structural functioning of discourse markers can be expected. Therefore, in this paper, we propose that the LDM, a theory of discourse syntax and semantics developed to explain phenomena quite distinct from those posed by discourse markers can provide a framework in which these functions of discourse markers as well as other phenomena of general linguistic interest can be insightfully addressed.

The LDM (Polanyi et al 2004a, b) is a formal theory of discourse syntax and semantics consisting of (1) a principled method of decomposing discourse into basic constituent units (BDUs, Basic Discourse Units) (2) a set of construction rules that uses information from sentential syntax and semantics and the lexicon to determine where on an Open Right discourse parse tree the next BDU is to be attached and what the structural type and semantics content of the constructed node is and (3) a dynamic semantic interpretation mechanism (Polanyi et al 2003). The structure of the discourse tree determines which previously processed discourse units remain available for continuation. In the paper, example discourses will be analyzed using the LDM and the resulting placement of the markers in the tree will illuminate both the boundaries and nature of the “units” bracketed by the markers. We show that they signal prominent structural constraints on discourse parse tree construction.

REFERENCES

- Asher, Nicholas and Alex Lascarides (2003). *Logics of Conversation*. Studies in Natural Language Processing. Cambridge University Press
- Davis, Christopher (2010). Decision theory and discourse particles: a case study from a large Japanese sentiment corpus. 24th Pacific Asia Conference on Language, Information, and Computation (PACLIC 24)
- Lenk, Uta (1998) *Marking Discourse Coherence: Functions of Discourse Markers in Spoken English*. Tübingen: Gunter Naar Verlag.
- Levinson, Stephen C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- Marcu, Daniel (1998). A surface-based approach to identifying discourse markers and elementary textual units in unrestricted texts. *The COLING/ACL'98 Workshop on Discourse Relations and Discourse Markers*, pages 1-7, Montreal, Canada, August 1998.
- Polanyi, Livia, Martin H. van den Berg and David Ahn (2003). Discourse Structure and Sentential Information Structure An Initial Proposal." *Journal of Logic, Language and Information*, Dordrecht, The Netherlands Kluwer Academic Publishers.
- Polanyi, Livia, Christopher Culy, Martin H. van den Berg, Gian Lorenzo Thione, and David Ahn (2004a). "Sentential Structure and Discourse Parsing". *Proceedings of the ACL2004 Workshop on Discourse Annotation*, Barcelona, Spain, July 25-26, 2004.
- Polanyi, Livia, Christopher Culy, Martin H. van den Berg, and Gian Lorenzo Thione (2004b). "A Rule Based Approach to Discourse Parsing". in: *Proceedings of the 5th SIGdial Workshop in Discourse And Dialogue*. Cambridge, MA USA pp. 108-117, May 1, 2004