

Backwards anaphora marking in coherence relations

Radoslava Trnavac and Maite Taboada

Cognitive Science Program/Department of Linguistics
Simon Fraser University
Burnaby, BC V5A 1S6, Canada

radoslava_trnavac@sfu.ca, mtaboada@sfu.ca

This paper investigates the distribution of backwards anaphora in nucleus/satellite coherence relations as defined within Rhetorical Structure Theory (Mann & Thompson 1988). In most approaches backwards anaphora is claimed to require the dependency of the pronoun clause on the antecedent clause (Carden 1978, van Hoek 1997, among others). However, consider the following example from Bosch 1983/Ariel 1990:

- (1) *He lied to me and John betrayed me.
- (2) He lied to me, and John was my friend.

In sentence (1), the conjuncts are symmetrical, and for that reason backwards anaphora is blocked. However in (2), the antecedent seems to occur in the background clause, which goes against the requirement on the pronoun clause being the dependent one. The frequent explanation for sentences like (2) is that the backwards anaphora is possible because of a process of pragmatic subordination where the pronoun occurs in the pragmatically non-Dominant clause, while the antecedent forms part of the Dominant clause (see McCray 1980, Harris & Bates 2002). Contrary to this argument, Ariel (1990:158) suggests that dependency and pragmatic subordination are crucial when the antecedent is a New entity, but when the entities form part of the discourse already, as in (2), dependency is not needed at all. According to her, usual Accessibility parameters such as distance and high/low cohesion are determining factors for the acceptability of backwards anaphora. We explore both views (the “pragmatic subordination” approach and the “new versus continuous discourse referent” approach) within the context of coherence relations.

We take as initial position that discourse structure is parallel to syntactic structure, and adopt the proposal of Matthiessen & Thompson (1988), who argue for a direct mapping between subordinate/matrix clauses at the syntactic level and satellites/nuclei of coherence relations at the discourse level. We analyze text spans with third person singular pronouns (*he*, *she*, *it*) in terms of the nucleus/satellite distinction in four different corpora: the American National Corpus (Reppen et al. 2005), the RST Discourse Treebank (Carlson et al. 2002), the Broadcast News corpus (Alabiso et al. 1998) and the New York Times corpus (Sandhaus 2008). Based on the preliminary analysis of instances with cataphoric *it*, we suggest the following hypotheses: (i) Along the lines of Ariel (1990), we claim that pragmatic subordination is not a sufficient factor for the backwards anaphora „violations“ since backwards anaphora appears within the nucleus (Dominant) relations; and (ii) contrary to what Ariel suggests, the correlation between dependency (high cohesion) and backwards anaphora as a New entity is not always a reliable

factor for the detection of backwards anaphora, since many satellite relations contain instances of continuous discourse referents. However, our data suggests that Distance as Accessibility marker may give us a slightly different picture in nucleus/satellite relations – nucleus relations as unmarked for the position of the continuous discourse referent have more instances of further distance between the antecedent and anaphora which, at least partially, confirms the predictions of Accessibility Theory.

References

- Alabiso, Jennifer, Robert MacIntyre and David Graff (1998). 1997 English Broadcast News Transcripts (HUB4) [Corpus]. Philadelphia, PA: Linguistic Data Consortium.
- Ariel, Mira (1990). *Accessing Noun-Phrase Antecedents*. London/New York: Routledge.
- Carden, Guy (1978). Backwards anaphora in discourse context. *Journal of Linguistics* 18 (2): 36-387.
- Carlson, Lynn, Daniel Marcu and Mary Ellen Okurowski (2002). RST Discourse Treebank, LDC2002T07 [Corpus]. Philadelphia, PA: Linguistic Data Consortium.
- Harris, Catherine L. and Elizabeth A. Bates (2002). Clausal backgrounding and pronominal reference: A functionalist approach to C-command. *Language and Cognitive Processes* 17: 237 - 269.
- Mann, William, and Thompson, Sandra (1988). Rhetorical structure theory: Towards a functional theory of text organization. *Text* 8 (3): 243-281.
- Matthiessen, Christian and Thompson, Sandra (1988). The structure of discourse and “subordination”. In *Clause Combining in Discourse and Grammar*, John Haiman and Sandra Thompson (Eds), 275-329. Amsterdam: Benjamins.
- McCray, Alexa (1980). The semantics of backward anaphora. *Cahiers Linguistiques d’Ottawa*, 8:329-343.
- Reppen, Randi, Nancy Ide and Keith Suderman (2005). American National Corpus (ANC) Second Release, LDC2005T35 [Corpus]. Philadelphia, PA: Linguistic Data Consortium.
- Sandhaus, Evan (2008). The New York Times Annotated Corpus [Corpus]. Philadelphia, PA: Linguistic Data Consortium.
- van Hoek, Karen (1997). *Anaphora and conceptual structure*. Chicago, University of Chicago Press.