Anaphoric expressions and Discourse structures

Laurence DANLOS
Université Paris 7, IUF, Lattice
laurence.danlos@linguist.jussieu.fr

1 Right Frontier Constraint in SDRT

In SDRT [AL03], the notion of "right frontier" based on discourse structure is well defined. It relies on the distinction between coordinating and subordinating discourse relations and it governs where new information can attach in an SDRS. For anaphoric expressions, the following constraint, called Right Frontier Constraint and noted RFC, is postulated: *"an antecedent for an anaphoric expression must be* DRS*-accessible on the right frontier"*. This constraint relies only on discourse structure (and DRS accessibility): it does not involve factors such as the definite/indefinite nature of the possible antecedents or their salience[1]. These factors are not ignored but are taken into account thanks to the following solution: discourse structure can include a bunch of nodes which do not correspond to discourse constituents. These nodes are mainly abstract topics inspired from the information structure domain. For example, discourse relations Narration and Background, which are both considered as coordinating, demand specific topics to be added during the construction of an SDRS. The addition of these topics explains the difference of acceptability between (1a) and (1b): (1a) with Background(π_1, π_2) is natural since the topic constructed with Background licenses the use of the pronoun *he* in the third sentence although its antecedent is not on the right frontier, whereas (1b) with Naration(π_1, π_2) is odd since the topic constructed with Narration does not license the use of *he*. However, enriching discourse structure is not the only means to explain this difference of acceptability. I will present a constraint explaining it, which takes into account discourse structure without enriching it.

(1)a. A burglar broke into Mary's apartment (π_1). Mary was asleep (π_2). He broke the silver (π_3).
b. # A burglar broke into Mary's apartment (π_1). A policeman visited her the next day (π_2). He broke the silver (π_3).

Along the same lines, [Ash05] has recently admitted that definite NPs put out a challenge to RFC. Let us illustrate it with (2) taken from [WG05]. How Elaboration(π_2, π_4) can be obtained, since π_2 is not on the right frontier? How the definite NP in π_4, *the basil*, could be linked to its antecedent, which is in π_2 and so not on the right frontier? Asher contemplates solving the challenge with definites again by "adding structure into the SDRS beyond the minimal number of constituents derived from the text so as to bring elements into prominence".

(2) Susan wanted to buy some tomatoes (π_1) and she also tried to find some basil (π_2) because her recipe asked for these ingredients (π_3). The basil would probably be quite expensive at this time of the year (π_4).

I see a methodological problem with this solution. Originally, RFC can be seen as stating that discourse structure constraints anaphora resolution (on top of DRS accessibility): this is an innovative position which can explain many data. However, it turns out that it is anaphora resolution which finally determines discourse structure. The same methodological shift is found with the distinction between the coordinating and subordinating discourse relations. There exist some criteria which determine the nature of a discourse relation[2], and so which determine, for example, that Result is coordinating. However, in [AV05], if an example of discourse involving Result includes an anaphoric expression which

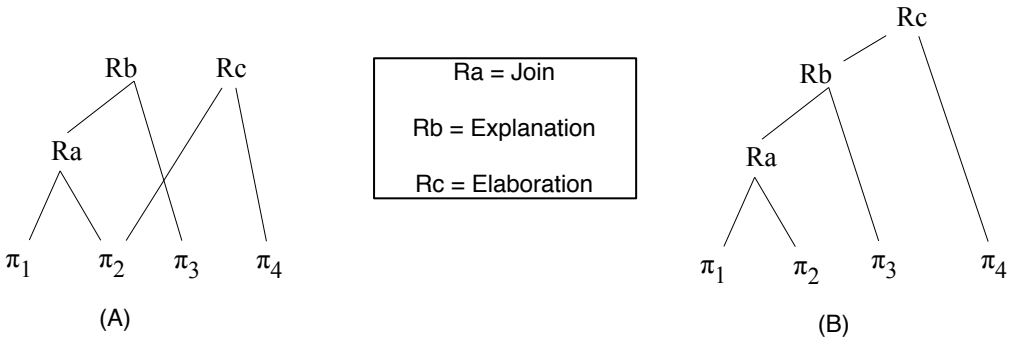
[1] These factors are well-known in the anaphora resolution community, who on the other hand ignores discourse structure factors.

[2] For example, in the temporal domain, there is a temporal progression of the events presented with coordination, whereas this progression is broken with subordination.

violates RFC, then Result is considered as subordinating. In conclusion, most of recent work in SDRT aims at saving RFC, event if it means adding nodes in discourse structures in an un-falsifiable way and/or changing the coordinating versus subordinating nature of discourse relations.

2 Treeness in RST

At the opposite of SDRT, RST proposes "surfacic" discourse structures, which contain only nodes corresponding to discourse constituents. It is postulated that discourse structure is tree shaped. Hence a problem with an example such as (2). Its discourse structure should be (A) below. (A) puts at stake crossing dependencies and is not tree shaped. To save the treeness of discourse structures, [ER06] claim that the discourse structure for (2) is (B) below. This is totally ad hoc and violates the basic principle to establish a discourse relation such as Elaboration[3]. This shows that RST is not descriptively adequate, a point already shown in [WG05] and [Dan06].



Conclusion

In the present state of knowledge, there exists no consensus on discourse structure, although the notion of discourse relation and the coordinating versus subordinating nature of discourse relations have reached a stable status. It seems more reasonable first to find a consensus on (surfacic) discourse structure and determine to which mathematical object it corresponds without a priori (contrarily to what is done in RST), and then to determine how discourse structure affects anaphora resolution, rather than establishing discourse structure with criteria based on anaphora resolution, as it is done in SDRT.

References

[AL03] Nicholas Asher and Alex Lascarides. *Logics of Conversation*. Cambridge University Press, Cambridge, 2003.

[Ash05] Nicholas Asher. Troubles on the right frontier. In *Proceedings of the First International Symposium on the Exploration and Modelling of Meaning, SEM-05*, Biarritz, 2005.

[AV05] Nicholas Asher and Laure Vieu. Subordinating and coordinating discourse relations. *Lingua*, 115(4):591–610, 2005.

[Cor95] Francis Corblin. *Les formes de reprise dans le discours, Anaphores et chaînes de référence*. Presses Universitaires de Rennes, Rennes, 1995.

[Dan06] Laurence Danlos. Comparing RST and SDRT discourse structures through dependency graphs. In Anton Benz and Peter Kuehnlein, editors, *Constraints in Discourse*. Academic Press, 2006.

[ER06] Markus Egg and Gisela Redeker. Underspecified discourse representation. In Anton Benz and Peter Kuehnlein, editors, *Constraints in Discourse*. Academic Press, 2006.

[GS86] Barbara Grosz and Candace Sidner. Attention, intention, and the structure of discourse. *Computational Linguistics*, 12(3), 1986.

[WG05] Florian Wolf and Edward Gibson. Representing discourse coherence: a corpus-based study. *Computational Linguistics*, 31(3):249–287, 2005.

[3] As shown by the fact that (B) would also be proposed if the last sentence of (2) would concern the price of tomatoes instead of that of basil.