



# Discourse markers as signals (or not) of rhetorical relations



Maite Taboada  
Simon Fraser University  
Vancouver, Canada  
mtaboada@sfu.ca



## Abstract



Ever since the formulation of Rhetorical Structure Theory (RST) by Mann and Thompson, researchers have debated about what is the “right” number of relations. The original 1988 paper contained 24 relations. Later proposals range in number from two to over 100. One proposal is based on the discourse markers (connectives) signaling the presence of a particular relationship.

In this paper I discuss the adequacy of such a proposal, in the light of two different corpus studies: a cross-linguistic study of conversations, and a study of English newspaper articles. In the conversation study, I report the results of examining 60 English and Spanish conversations, analyzed in terms of RST relations. I discuss the identification of relations both through lexical discourse markers and through position in the conversation, concluding that there exists a class of rhetorical relations that are never signaled in the corpus. In the newspaper articles study a similar analysis is carried out. A comparison between the two studies suggests that genre-specific factors may affect which relations are signaled.

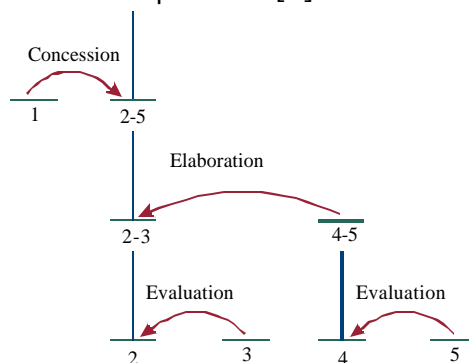
# Rhetorical relations

- ◆ Text coherence through text relations: *coherence* (Bateman & Rondhuis 1997), *discourse* (Hobbs 1985, Grosz & Sidner 1986) or *conjunctive relations* (Martin 1992)
- ◆ Rhetorical Structure Theory (RST) (Mann & Thompson 1988)
- ◆ Text organization: relations between parts of a text
- ◆ Every part has a function with respect to other parts in the text
  - Text divided in spans (nuclei and satellites)
  - Hierarchical (represented in trees)
  - Some relations: Condition, Concession, Evidence, Contrast, Solutionhood, etc.

3

# RST Example

[1] No, the eighth doesn't look good at all, [2] but the ninth, [3] that sounds perfect. [4] Before two, [5] that'd be good



4

## Discourse markers

- ◆ Pesky little particles (Grimes 1975)
- ◆ Conjunctions, adverbs, sentence adverbs, 'fillers' (Schiffrin 1987, Fraser 1997, Jucker & Ziv 1998)
- ◆ Multiple functions: indicate new topics, maintain the floor, express attitude, establish rapport...
- ◆ DMs explicitly signal relations between text spans:
  - Tom quit his job. He was tired of the long hours.
  - Tom quit his job **because** he was tired of the long hours.

5

## Corpus analysis 1: task-oriented conversations

- ◆ Collected by the Interactive Systems Lab at Carnegie Mellon University
- ◆ Task-oriented conversations: two participants set up an appointment
- ◆ 30 in English (6,804 words), 30 in Spanish (9,112 words)

6

# Method

- ◆ Issues in dialogue
  - RST not conceived for dialogue. Some RST constraints had to be relaxed
  - For this analysis, only RRs within the turn
- ◆ RST analysis first, then looked for DMs
- ◆ Only conjunctions
  - Coordinating
  - Subordinating

7

## Results, conversations

	English		Spanish	
	Tokens relation	% marked	Tokens relation	% marked
Antithesis	1	-	1	100.00
Background	13	7.69	16	-
Circumstance	5	-	9	55.56
Concession	71	53.52	70	54.29
Contrast	9	22.22	8	75.00
Condition	66	50.00	89	26.97
Elaboration	166	1.20	140	6.43
Enablement	10	-	4	-
Evaluation	8	-	11	-
Evidence	0	-	4	-
Interpretation	1	100.00	11	100.00
Joint	29	55.17	26	42.31
Justify	34	8.82	22	77.27
Motivation	3	-	0	-
Non-volitional Cause	37	62.16	54	57.41
Non-volitional Result	43	62.79	29	72.41
Otherwise	4	50.00	4	100.00
Purpose	10	90.00	26	96.15
Restatement	28	10.71	47	36.17
Sequence	7	57.14	20	85.00
Solutionhood	7	-	0	-
Summary	4	25.00	6	16.67
Volitional Cause	7	14.29	11	100.00
Volitional Result	17	82.35	48	100.00
<b>Total</b>	<b>580</b>	<b>30.86</b> <i>(n=179)</i>	<b>656</b>	<b>45.88</b> <i>(n=301)</i>

## Summary, conversations

- ◆ Most marked: Concession, Condition, Cause, Result
- ◆ Least: Enablement, Evaluation, Solutionhood
- ◆ Some relations marked more often in Spanish: Justify, Vol. Cause and Vol. Result
- ◆ Overall, relations are marked less than half the time:
  - English, 30.86%
  - Spanish, 45.88%

9

## Corpus analysis 2: WSJ articles

- ◆ RST-annotated corpus of Wall Street Journal articles (Carlson et al. 2001)
- ◆ Study of 6 relations:
  - Frequently marked: Concession, Circumstance, Result (45,000 words)
  - Rarely marked: Elaboration, Summary, Background (113,000 words)

10

## Results, articles ('typically marked' relations)

	<b>Tokens relation</b>	<b>% marked</b>
Concession	228	90%
Circumstance	539	66%
Result	117	68%
<b>Total</b>	<b>884</b>	<b>72.4%</b> <i>(n=640)</i>

11

## Summary, articles

- ◆ In progress ('typically unmarked' relations)
- ◆ Heterogeneous: articles, letters to the editor, editorials
- ◆ Even within the relations that were chosen as 'typically marked', marking is not always present
  - 72.4% for Concession, Circumstance and Result overall

12

## A taxonomy of RRs

- ◆ Active area of research
- ◆ Original proposal for RST: 24 relations. Other: from two relations (Grosz & Sidner 1986) to 100 (Hovy & Maier 1992, Martin 1992)
- ◆ How to create a taxonomy:
  - Cognitive approach (Sanders et al. 1992, 1993)
  - Use discourse markers as indicators of the relation (Knott and Dale 1994)

13

## Do discourse markers indicate the presence of a RR?

- ◆ Cognitively plausible: speakers/writers signal explicitly the relationship between two pieces of text
- ◆ Method used in discourse parsing (Marcu 2000)
- ◆ Answer: **sometimes**
  - *But* (Eng.) and *pero* (Spa.) consistently indicate Concession
  - *As a result (of)* used very often in Result
  - Etc.

14

## Can DMs be used to taxonomize relations?

- ◆ A method for a taxonomy: we can look for DMs, and then establish relations based on the DMs found

15

## Answer: No

- ◆ DMs, given a relation: from 0 to 90%
- ◆ Even RRs with the highest signaling use a number of markers
  - Purpose, in the spoken corpus is marked about 90% of the time
  - Markers: *and, so that, that, to*
  - *And*: used also in Elaboration, Joint, Cause, Sequence and Result

16



---

## The signaling problem

---



- ◆ Speakers and hearers presumably identify these relations
- ◆ How does the signaling happen,
  - in general, with relations that are rarely marked, and
  - for individual cases, with relations that are frequently marked, but that occur in unmarked contexts?

17



---

## The signaling problem

---



- ◆ Marking a process of grammaticalization: markers used to signal intra-clausal relations are used also for inter-clausal ones
- ◆ Other cues: genre expectations, punctuation (Dale 1991), mood, modality, tense, aspect

18

## Wall Street Journal corpus

- ◆ Non-finite clause (-ing) marks a Circumstance:

Net Income at Sears's Dean Witter Financials Services group, meanwhile, rose nearly 32% to \$35.7 million, **reflecting** improvements in its basic stock brokerage and Discover credit card businesses

19

## DMs, RRs, and genre

- ◆ Genre: purposeful, goal-oriented activity. It develops in stages (Martin 1985)
- ◆ Stages show different lexico-grammatical and discourse patterns
- ◆ WSJ: articles include quoted speech
  - Circumstance is more marked when it appears in the body of a quote than in the expository part of the article

20

## Task-oriented conversations

Stages	Most frequent RRs
Opening	Solutionhood
Task Performance	
Proposal	Solutionhood, Motivation, Background
Rejection	Non-Vol. Cause/Result, Concession
Acceptance	Volitional Cause or Result
New Proposal	Result (Vol. or Non-Vol.)
Details	Elaboration
Closing	Evaluation, Restatement, Summary

21

## Conclusions

- ◆ Rhetorical relations are sometimes marked by discourse markers
- ◆ DMs cannot be used to establish a taxonomy of relations
- ◆ There are differences in DM use in rhetorical relations
  - across different genres
  - within a genre, across its different stages

22

## References

- Bateman, J., & Rondhuis, K. J. (1997). Coherence Relations: Towards a General Specification. *Discourse Processes*, 24, 3-49.
- Carlson, L., Marcu, D. & Okurowski, M. E. (2001) Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory. *Proceedings, 2nd SIGDIAL Workshop in Discourse and Dialogue, Eurospeech 2001*.
- Dale, R. (1991). The Role of Punctuation in Discourse Structure. Paper presented at the *AAAI Fall Symposium on Discourse Structure in Natural Language Understanding and Generation*, Asilomar, California.
- Fraser, B. (1997) What are discourse markers? *Journal of Pragmatics*, 31, 931-952.
- Grimes, J. (1975) *The Thread of Discourse*. Mouton: The Hague.
- Grosz, B. J., & Sidner, C. L. (1986). Attention, Intentions, and the Structure of Discourse. *Computational Linguistics*, 12 (3), 175-204.
- Hobbs, J. (1985). *On the Coherence and Structure of Discourse* (Research Report). Stanford University.
- Hovy, E. & Maier, E. (1992) *Parsimonious or Profligate: How Many and Which Discourse Structure Relations*. ISI Technical Report.

23

## References (cont.)

- Jucker, A. & Ziv, J. (eds.) (1998) *Discourse Markers: Description and Theory*. Amsterdam: Benjamins.
- Knott, A., & Dale, R. (1994). Using Linguistic Phenomena to Motivate a set of Coherence Relations. *Discourse Processes*, 18 (1), 35-62.
- Mann, W. C. & Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8 (3). 243-281.
- Marcu, D. (2000) *The Theory and Practice of Discourse Parsing and Summarization*. Cambridge, Ma: MIT Press.
- Martin, J. R. (1985) Language, register and genre. In Frances Christie (ed.) *Children Writing: Reader*. Geelong, Victoria: Deakin U. Press. 21-30.
- Martin, J. R. (1992) *English Text: System and Structure*. Amsterdam: Benjamins.
- Sanders, T., Spooren, W., & Noordman, L. (1992). Toward a Taxonomy of Coherence Relations. *Discourse Processes*, 15, 1-35.
- Sanders, T., Spooren, W., & Noordman, L. (1993). Coherence Relations in a Cognitive Theory of Discourse Representation. *Cognitive Linguistics*, 4(2), 93-133.
- Schiffrin, D. (1987) *Discourse Markers*. Cambridge: CUP.

24