# Extracts from Thesis

Daniel Marcu

## Rhetorical relations used in the corpus analysis

| Rhetorical relation | Number of occurrences in the first 2100 text fragments of the corpus |
|---|---:|
| **additive-emphasis** | 17 |
| **alternative** | 4 |
| **anti-sequence** | 18 |
| **antithesis** | 67 |
| **antithesis-sequence** | 2 |
| **argumentation** | 36 |
| **background** | 70 |
| **broken-intention** | 1 |
| **circumstance** | 156 |
| **comparison** | 36 |
| **concession** | 76 |
| **conclusion** | 14 |
| **concurrency** | 6 |
| **condition** | 41 |
| **continuation** | 6 |
| **contrast** | 120 |
| **counter-evidence** | 1 |
| **detail** | 5 |
| **duration** | 1 |
| **elaboration** | 236 |
| **enablement** | 4 |
| **evaluation** | 2 |
| **evidence** | 108 |
| **example** | 2 |
| **explanation** | 48 |

| Rhetorical relation | Number of occurrences in the first 2100 text fragments of the corpus |
|---|---|
| **forward-reference** | 2 |
| **final-step** | 1 |
| **interpretation** | 70 |
| **introduction** | 1 |
| **joint** | 214 |
| **justification** | 34 |
| **means** | 2 |
| **motivation** | 11 |
| **narration** | 2 |
| **non-evidence** | 1 |
| **non-explanation** | 1 |
| **nonvolitional-cause** | 74 |
| **nonvolitional-result** | 14 |
| **nonvolitional-cause-result** | 5 |
| **or** | 2 |
| **otherwise** | 20 |
| **outcome** | 2 |
| **parenthetical** | 60 |
| **problem-solution (solutionhood)** | 1 |
| **purpose** | 10 |
| **question-answer** | 5 |
| **reason** | 10 |
| **refutation** | 1 |
| **restatement** | 9 |
| **sequence** | 160 |
| **summary** | 6 |
| **topic-shift** | 25 |
| **volitional-cause** | 28 |
| **volitional-result** | 2 |

During my corpus analysis, I noticed a set of other problems that I consider worthy of being brought to the reader's attention. These problems stem from the lack of objective definitions for the notions of elementary textual unit, nuclearity, and rhetorical relation. Below, I discuss each of these problems in turn.

### Problems with identifying the elementary units of text

My initial intent was to take clauses as the elementary units of discourse. Consider, however, the text shown in (1), below.

(1) [*Because of* light leakage from one ultraviolet source to another,][ the lights are switched by a commutator-like assembly rotated by a synchronous motor.]

If I had taken my initial intent literally, I would have not broken sentence (1) into two units, because "light leakage from one ultraviolet source to another" does not contain a verb, and therefore, is not a clause. However, the marker *Because of* clearly signals a causal relation between the textual spans "light leakage from one ultraviolet source to another," and "the lights are switched by a commutator-like assembly rotated by a synchronous motor". Uncovering this relation can be only beneficial from a text understanding perspective. However, how far should one go in this attempt of using phrases rather than clauses as the elementary units of discourse?

As I have already discussed in section ??, in the texts that I analyzed, I did not use an objective definition of elementary textual unit. Rather, I relied on a more intuitive one: whenever I found that a cue phrase signalled a rhetorical relation between two spans of text of significant sizes, I assigned those spans an elementary unit status, although, in some cases, they were not fully fleshed clauses.

### Problems with identifying the rhetorical status of the textual units involved in a discourse relation

As we have seen, nuclearity plays a major role in the formalization of text structures that I proposed in chapter ??. One of the main assumptions that this formalization relies upon is that the rhetorical statuses of the units involved in a rhetorical relation can be determined unambiguously. However, in few cases in the corpus, although the rhetorical relation that held between two units was easy to label, it was ambiguous as to which unit was the nucleus and which was the satellite. Consider the following example:

(2) [It is not enough for man to be an ontological esse.] [He needs existential completion,] [he needs, that is, to move in the direction of completion.] [And the direction of that movement is determined by his perception of the truth about himself.] [He must, *consequently*, exist as a self-perceived substantive, developing agent,] [or he does not exist as man.] [Thus, it is no mystical intuition,] [but an analyzable conception to say that man and his tradition can "fall out of existence".] [This happens at the moment man loses the perception of moral substance in himself,] [of a nature that, in Maritain's words, is perceived as a "locus of intelligible necessities".] [An existentialist is a man who perceives himself only as "esse",] [as existence without substance.]

The cue phrase *consequently* clearly marks a causal relation between units "And the direction of that movement is determined by his perception of the truth about himself" and "He must, *consequently*, exist as a self-perceived substantive, developing agent". It is, however, not obvious which unit should be assigned the status of nucleus and which that of satellite. In fact, in general, causal relations are difficult to assign a nuclear status: in some cases, the context provides enough evidence with respect to whether the writer intended to assign a more important role to the cause or to the result. In some cases, however, it seems that the nuclearity assignment can go either way.

More precisely, it is not that the taxonomy of relations does not distinguish between causal relations in which the cause is the nucleus and causal relations in which the result is the nucleus, but rather that we lack an objective definition that would allow us to determine which of these relations to use.

### Problems with identifying the rhetorical relations that holds between two textual units

During the corpus analysis, it was sometimes difficult to determine one rhetorical relation that would most adequately characterize the relation between two units. Consider, as an example the text shown in (3) below.

(3) [Certain badly disillusioned market critics are often apt to feel that there is something somehow unfair, dirty, or even thoroughly criminal about this interplay of competitive forces.] [*But after all*, can anyone imagine a market wherein the reverse of these things were true?] [Try to imagine a market in which only a minority of traders would lose,] [and the majority would make consistent profits.] [How much and how many profits could a majority take out of the losses of a few?]

What is the rhetorical relation that best describes the relationship between the first two sentences? To a certain degree is a **contrast** between the features of a real market and the features of an imaginary one. But at the same time, an **interpretation** relation can be considered to hold between the two sentences as well. Which one should we choose? And if we choose both, how do we objectively assign a strength or preference to one of the relations? In my analysis, I chose to label a relation between two textual spans with *all* the names of the rhetorical relations whose definitions seemed to apply.