Some considerations on the use of main verbs to express rhetorical relations

Pierre Lejeune, Amália Mendes and Nuno Martins
University of Lisbon - Centre of Linguistics / Faculty of Arts

1. Introduction
Rhetorical relations are typically expressed by discourse structuring devices that ensure textual cohesion and coherence. Resources such as the PDTB target specifically the annotation of these devices, while describing alternative lexicalizations of such relations (AltLex).

Based on our preparatory work to develop a discourse treebank for Portuguese in the PDTB framework, our goal is to:

- Analyse the status, in intra-sentential coherence, of main verbs that internally express a rhetorical relation
  - focusing on causative senses Reason, Result, Pragmatic_justification
  - and causal discourse verbs such as provocar ‘to cause’, obrigar ‘to force’ and reduzir ‘to reduce’
- Discuss to what extent these verbs have a cohesive function in texts
  - based on contexts extracted from the corpus CINTIL, a 1M word corpus annotated for part-of-speech and manually revised.

2. Verbalization of discourse relations in PDTB

- “DRMs [Discourse Relational Markers] are a lexically open-ended class of elements which may or may not belong to well-defined syntactic classes”.
- AltLex require that “A discourse relation can be inferred between adjacent sentences” → inter-sentential only.
- The verbalization of discourse relations at the intra-clausal level is mentioned as related work, but it is not clear whether the possibility of annotating instances of the verbs involved as AltLex is envisaged at all.

3. Conceptual Structure
The decomposition of lexical meaning into semantic primitives expressed by a conceptual structure or a lexical conceptual structure:

- provocar ‘to provoke’:
  - Internally complex event formed by a causative and an existential meaning [CAUSE [TO BE]]
  - The two arguments of the verb provocar are frequently nominalizations, “a by-product of explicit realization of the relations as verbs and propositions” (Kibble, 1999).
  - “Typically, nominalized forms denote a property, an event or process, or the state resulting from an event” (Kibble, 1999).

4. The Framenet Model
In order to annotate verbs like reduzir in the PTDB framework, we could rely on Framenet. In Frame Semantics:

- A frame is constituted by a lexical unit called the target and by frame elements that combine with it.
- Typically, but not necessarily, the target is a verb and the core frame elements are complements or adjuncts.
- Some frames semantically encapsulate discourse relations and, quite naturally, their expression through verbs (e. g. for contingency relations: [causation], [cause to X], [concessive], [conditional occurrence], [creating], [evidence]).
- A system of annotation layers allows lexical units to be annotated at the same time as targets and frame elements.

5. Final remarks
One could argue that verbs that mark discourse relations (discourse verbs) should be included in the annotation of discourse relations, provided they link “events, states and propositions”, whatever the grammatical realization of the arguments (nominalization, non-finite clause, etc.) is (Danlos, 2006).

A context such as those illustrated in (1) and (2) are at the crossroad between syntax and discourse and consequently challenge the limits of the annotation performed in the framework of the PDTB, namely the concepts that we explore in this paper of Alternative Lexicalization (taking into account the fact that they “convey more than just the meaning of the relation”) and nominalization.

Selected references: