The role of segment-internal elements in the signaling of coherence relations

Jet Hoek (The University of Edinburgh), Sandrine Zufferey (University of Bern), Jacqueline Evers-Vermeul (Utrecht University), & Ted J.M. Sanders (Utrecht University)

When readers or listeners are presented with a text, they do not treat the individual clauses and sentences in that text as independent and unrelated. Instead, they try to relate each part of the text, or each *discourse segment*, to the rest of the discourse by inferring coherence relations between the discourse segments. Language users can be facilitated in inferring coherence relations by the presence of connectives or cue phrases (from now on referred to as 'connectives') that provide explicit processing instructions on how to relate two discourse segments to each other (cf. Sanders & Spooren 2007), but many coherence relations have to be inferred without the help of a connective.

Traditionally, relations with a connective, as in (1), have been labeled 'explicit' coherence relations; relations without a connective, as in (2), as 'implicit' relations. Although this distinction seems very straightforward, it is not without its problems. Connectives can for instance signal a relation that is less specific than the relation that is constructed by language users, as in (3), where the relation is marked by *after*, a temporal connective, but the inferred relation is causal. The relation in (3) is therefore less explicitly signaled than the relation in (2). In addition, a relation without a connective may contain strong other cues that help language users infer the appropriate relation. The semantic opposition between *loves* and *despises*, for instance, appears to function as a signal for the contrastive coherence relation in (4). This relation could thus be argued to be more explicitly signaled than the relation in (2), even though neither fragment contains a connective.

(1) [Kate missed last night's dinner]$_{S1}$ **because** [she had to take all three of her dogs to the emergency vet.]$_{S2}$

(2) [The day after Christmas Mark always makes chocolate mousse.]$_{S1}$ **Ø** [It is a great way to get rid of all the leftover holiday candy.]$_{S2}$

(3) [Paul was banned for life from the bowling alley]$_{S1}$ **after** [getting drunk and hiding all the bowling pins.]$_{S2}$

(4) [Harry <u>loves</u> Easter.] **Ø** [His sister <u>despises</u> all big holidays.]

Connectives are the most prototypical linguistic elements that signal how discourse segments should be related to each other, which is why research on discourse coherence has mostly been focused on connectives as markers of coherence relations. However, while connectives are the only linguistic elements that by definition express relational meaning, that does not necessarily mean they are the only indicators for coherence relations. By limiting our attention to connectives, we are likely missing out on important other cues readers and listeners use when establishing coherence relations.

The most elaborate research effort to identify other signals for coherence relations has been the recently released RST Signalling Corpus (Das, Taboada, & McFetridge 2015), in which linguistic cues that signal coherence relations annotated in the RST Treebank (Carlson, Okurowski, & Marcu 2002) are identified. While the RST Signalling Corpus is an extremely valuable inventory of potential signals for coherence relations, it does not (yet) draw a systematic link between signals and specific relation types, and does not comment on *how* or *why* the indicated signals function as cues for coherence relations. In addition, since the annotation was mostly focused on relations without connectives, the RST Signalling Corpus does not identify potential additional signals in relations that contain a connective. In this presentation, we will explore the marking of coherence relations by connectives on the one hand, and other types of cues on the other. Specifically, we will investigate how linguistic elements within the segments of a coherence relation, i.e., *segment-internal elements*,

can contribute to the marking of the relation, and how the presence of segment-internal signals relates to the presence of connectives.

Within the field of discourse, there are several segment-internal features that have been linked to particular types of coherence relations. These segment-specific elements include a wide range of linguistic categories, such as complex phrases, lexical items, modal markers, and verbal inflection. The features can either occur in one of the segments or in both of the segments. It seems, however, that not all linguistic elements that have been associated with a specific type of coherence relation signal the relation in the same way, and there appear to be differences in the way in which the presence of a specific linguistic element in the segments of a relation can impact the marking of that relation by means of a connective. In this presentation, we will argue that there are, at least, three distinct ways in which segment-internal elements systematically interact with the connective that marks a coherence relation. We label these interactions *division of labor*, *agreement*, and *general collocation*. In division of labor types of interactions, the connective and the other signal overlap in the meaning they encode and the presence of one is likely to make (part of) the other redundant (e.g., *if*, a positive conditional connective, plus the presence of a negation element can express a meaning similar to *unless*, a negative conditional connective). In agreement types of interactions, the connective and the other signal overlap in the meaning they encode, but they are commonly used in addition to each other (e.g., in Dutch, subjective causal connectives often occur in addition to subjective adjectives or adverbs). In general collocation types of interactions, there is no overlap in the meaning signaled by the connective and the other signal (e.g., implicit causality verbs, for instance *fire*, *ridicule*, or *congratulate*, often co-occur with causal relations, but the implicit causality verb itself does not explicitly signal causality).

We base our categorization on several combinations of segment-internal elements and connectives that have been linked to each other on the basis of monolingual corpus data, theoretical explorations, or experimental work. We then use parallel corpus data to show that the three types of interactions we formulated are observable in translation. We opt for a translation corpus because even though monolingual corpora are extremely valuable resources for language research, when studying *meaning* they require researchers to rely on their own interpretations, since "meaning is not directly observable," (Noël 2003:758). When it comes to the interaction between segment-internal elements and connectives, it is not necessarily obvious what and how each element contributes to the overall interpretation of a relation. We therefore make use of parallel corpora, which consist of a source text and one or multiple translations (e.g., Noël 2003). In this approach, the translator is treated as a naive 'annotator,' whose main purpose was to accurately convey the meaning of the source text fragment in the target language. The three types of interactions between segment-internal elements and connectives we identify make different predictions for translation; our corpus data show that we can indeed distinguish distinct translation patterns for each type of interaction. Identifying a way in which linguistic elements other than connectives can be systematically linked to coherence relations is an important step toward fully understanding the marking of coherence relations.

References

Carlson, Lynn, Mary Ellen Okurowski, & Daniel Marcu, 2002. *RST Discourse Treebank*. Philadelphia: Linguistic Data Consortium.

Das, Debopam, Taboada, Maite, and McFetridge, Paul, 2015. RST Signalling Corpus LDC2015T10. Web Download. Philadelphia: Linguistic Data Consortium.

Noël, Dirk, 2003. Translations as evidence for semantics: An illustration. *Linguistics* 41(4), 757-785.

Sanders, Ted J.M. & Wilbert P.M.S. Spooren, 2007. Discourse and text structure. In: Dirk Geeraerts & Hubert Cuyckens (eds.), *The Oxford Handbook of cognitive linguistics*, 916-941. Oxford: Oxford University Press.