

Ontological Proxies to Augment the Expressiveness of Discourse Analysis

Cesar Gonzalez-Perez

Incipit CSIC

Santiago de Compostela, Spain

cesar.gonzalez-perez@incipit.csic.es

Abstract

Discourse analysis usually focuses on what is being said, following the text very literally, and pays little or no attention to the world to which the text refers. We argue that considering the ontology to which a discourse refers as well as the text itself provides a richer and more useful representation of the discourse. To this end, we propose the notion of ontological proxies, i.e. conceptual artefacts that connect elements in the argumentation structure to the associated ontology elements.

1 PROBLEM AND MOTIVATION

Discourse analysis techniques such as RST (Mann and Thompson 1988) or IAT (Reed and Budzynska 2010) help us understand the argumentation structure and content of texts, contributing to better insights into how people justify their claims and how we construct knowledge. Discourse analysis usually proceeds by breaking down a text into meaningful chunks, such as sentences, locutions, or utterances, and then constructing a model of how these chunks are related to each other in terms of argumentation schemes or coherence relations. This allows us to identify argumentation devices such as inferences, conflicts or rephrasings.

Naturally, discourse analysis focuses on what is being said, and follows the text very literally. This is a desirable property, as being faithful to the text being analysed avoids unwanted biases and spurious information. However, this also means that little or no attention is paid to the world to which the text refers. However, the analyst must necessarily develop a mental model of what entities are being referred to by the text in order to understand it, resolve references, construct meaning and, in general, make sense of the words. But this mental model is rarely captured as a product of the analysis process and is lost forever. Readers or users of the analysis products must re-create the mental model in their heads, possibly diverging from the interpretation adopted by the analyst, and thus hindering the communication.

When the text being analysed involves a dialogical situation in which two or more agents exchange arguments, this issue becomes even more important. When working with dialogical texts, the analyst discovers the common ontology shared by the speakers and interprets their utterances in relation to it. A shared ontology between speakers must exist; otherwise, no communication would be possible. But, again, this ontology is rarely documented, and the products of the discourse analysis do not refer to it. In this manner, the argumentation relations are left unanchored, leaving to each reader or user the task of re-creating this ontology in their heads and re-interpreting the analysis products in relation to it.

2 PROPOSED APPROACH

In this talk I argue that considering the ontology to which a discourse refers, together with the text itself, provides a richer and more useful representation of the discourse. In other words, I propose that a model of a part of the world must accompany every model of a discourse, and that relations between elements in these models must be also described. In this manner, the joint model will describe a discourse plus the entities of the world that this discourse refers to in an interconnected manner.

Models of the world are sometimes called conceptual models, sometimes ontologies, depending on the disciplinary tradition. Their commonalities are much greater than their differences (Atkinson, Gutheil, and Kiko 2006; Henderson-Sellers 2011), so I will use these terms interchangeably in this talk.

There is an extensive body of literature on conceptual models (or ontologies), and conceptual modelling is practised today through the use of many techniques, languages and tools, such as ConML (Gonzalez-Perez 2018), OWL (World Wide Web Consortium 2012) or UML (OMG 2017). Inventing a new modelling language to cater for the need described in the previous section would not make sense, as it would be redundant and most likely inferior to those that already exist. For this reason, I propose the notion of *ontological proxies*. An ontological proxy is an element in a discourse model that stands for another element in an associated conceptual model. In other words, ontological proxies allow us to point to the right model elements in a fully-fledged conceptual model from a discourse model, thus connecting the two (Figure 1).

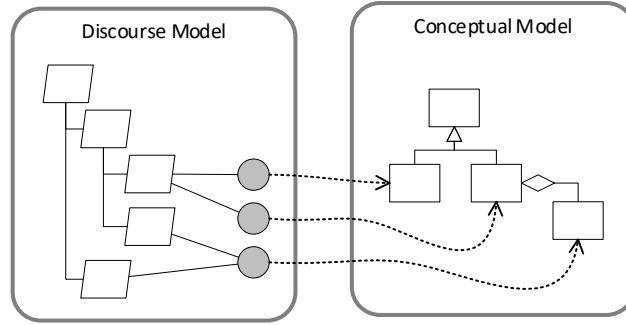


Figure 1. Ontological proxies in the context of discourse modelling. Trapezoids under “Discourse Model” represent the propositions or phrases modelled from the discourse. Grey circles represent ontological proxies. Rectangles under “Conceptual Model” represent elements in the conceptual model. Dotted arrows represent the mappings between the two models.

In this manner, ontological proxies are part of the basic building blocks from which we can construct discourse models, together with other discourse-oriented primitives. But, given the fact that they are, by definition, proxies for something else, they must be kept lightweight and minimal. Otherwise, we would be missing the point of having them in the first place. For this reason, we propose that ontological proxies are defined in terms of a minimal set of concepts, properties, and relationships that can easily map to the most common conceptual modelling languages. Specifically, we propose the following primitives:

- **Entity**, which represents particulars in the world such as the computer I am using or the 5/2016 Act on Cultural Heritage, for example.
- **Value**, which represent atomic qualities or quantities of particulars, such as `me.Age = 52`.
- **Reference**, which represent connections between particulars, such as `me.WorksAt = Incipit`.
- **Category**, which represent universals such as `Computer`, `Person` or `Law`.
- **Property**, which represent values of categories, such as `Person.Age` or `Law.PublicationDate`.
- **Association**, which represent references between categories, such as `Person.WorksAt` or `Law.AppliesIn`.

Once these primitives are in place, we can assume that any proposition in the analysis is able to *denote* instances of them. For example, the proposition “my house was built in 1994” denotes an entity through the fragment “my house” plus an associated value through the fragment “was built in 1994”.

3 EXAMPLE

Let’s consider the following exchange:

Speaker 1: The 5/2016 law on cultural heritage says that you cannot build close to a protected site.

Speaker 2: But the law also says that I have the right to buy and possess any land, so...

Here, speaker 1 is clearly referring to the 5/2016 Act on Cultural Heritage, which is a very specific legal text in Galicia, and using it as grounds to back up their claim. Speaker 2 replies that “the law” apparently contradicts this. However, it is not clear from the text whether “the law” refers to the very same 5/2016 Act on Cultural Heritage or, to the contrary, to ‘the law’ as the abstract entity of the regulations that apply here and now. When analysing this text, we must develop a mental model that allows us to interpret speaker 2’s utterance and connect it to speaker 1’s claim. We may be right or wrong in our interpretation, but we should capture what it is, document it, and connect it to the discourse analysis. In this case, I imagine that speaker 2 refers to the law in general, rather than the 5/2016 Act on Cultural Heritage. This means that:

- The fragment “The 5/2016 law on cultural heritage” in the locution spoken by speaker 1 denotes an entity, which I will call *ach* (for “Act on Cultural Heritage”).
- The fragment “the law” in the locution spoken by speaker 2 denotes a different entity, which I will call *glr* (for “general laws and regulations”).

Just by doing this, we have documented our interpretation that the speakers are referring to different things, and thus there may be no conflict after all. Any reader of our discourse analysis will be able to see this explicitly. But we can go further. For example, we can construct a conceptual model that describes the common ontology shared by speakers 1 and 2, by using the evidence provided by the text and its context. For example, I may construct a ConML model that includes objects for the *ach* and *glr* ontological proxies, plus a reference *IsPartOf* from the former to the latter. In this manner, I am capturing the fact that the 5/2016 Act on Cultural Heritage is a part of a bigger whole. Note that nowhere in the text is said that the 5/2016 Act on Cultural Heritage is part of a larger legal body to which speaker 2 is referring; this is something that only exists in the associated conceptual model. Furthermore, by capturing this fact in the conceptual model, we are documenting our position that it is reasonable to think that both speakers share the belief that the 5/2016 Act on Cultural Heritage does not exist in isolation but exists as part of a larger legal body.

Now, we can reason about the claims made by the speakers in their exchange using the extended information provided by the conceptual model. For example, the connotation of the “but” marker in speaker 2’s locution can be reassessed in the light of the above mentioned whole/part relationship. What initially may have looked like a conflict, now it is more nuanced, as the speakers seem to be dealing with a case of a legal exception or particular case. This provides additional information that allows us to understand their apparent disagreement, and even act to reconcile their points of view if that is what we wanted.

4 CONCLUSION

Ontological proxies are simple and lightweight elements in a discourse model that allow us to connect what is being said in the discourse to the associated things in the world, via a conceptual model. Doing this provides additional information to reason about the argumentation structure and contents.

In my talk I will provide additional details on how ontological proxies are defined in the context of a discourse modelling language, how denotations work, and how conceptual models can be used to connect different discourses in order to perform intertextual analyses. This is especially interesting when analysing texts that refer to a common set of things in the world, such as political discourses from different candidates addressing a hot topic, or tweets by different users about the same thing.

REFERENCES

- Atkinson, Colin, Matthias Gutheil, and Kilian Kiko. 2006. "On the Relationship of Ontologies and Models." In *Proceedings of the 2nd International Workshop on Meta-Modelling (WoMM)*, edited by Saartje Brockmans, Jürgen Jung, and York Sure, LNI 96:47–60. Karlsruhe, Germany.
- Gonzalez-Perez, Cesar. 2018. *Information Modelling for Archaeology and Anthropology*. Springer. <https://doi.org/10.1007/978-3-319-72652-6>.
- Henderson-Sellers, Brian. 2011. "Bridging Metamodels and Ontologies in Software Engineering." *Journal of Systems and Software* 84 (2): 301–13. <https://doi.org/doi:10.1016/j.jss.2010.10.025>.
- Mann, William C., and Sandra A. Thompson. 1988. "Rhetorical Structure Theory: Toward a Functional Theory of Text Organization." *Text - Interdisciplinary Journal for the Study of Discourse* 8 (3). <https://doi.org/10.1515/text.1.1988.8.3.243>.
- OMG. 2017. "Unified Modeling Language." Object Management Group.
- Reed, Chris, and Katarzyna Budzynska. 2010. "How Dialogues Create Arguments." In *ISSA Proceedings 2010*, edited by Frans van Eemeren, Bart Garssen, David Godden, and Gordon Mitchell. <http://rozenbergquarterly.com/issa-proceedings-2010-how-dialogues-create-arguments/>.
- World Wide Web Consortium. 2012. "OWL 2 Web Ontology Language." World Wide Web Consortium. <http://www.w3.org/TR/2012/REC-owl2-overview-20121211/>.