Connecting Discourse and Domain Models in Discourse Analysis through Ontological Proxies

Cesar Gonzalez-Perez

Institute of Heritage Sciencies (Incipit), CSIC, Avda. de Vigo, s/n, 15705 Santiago de Compostela, Spain; cesar.gonzalez-perez@incipit.csic.es

check for updates

Abstract: Argumentation-oriented discourse analysis usually focuses on what is being said and how, following the text under analysis quite literally, and paying little attention to the things in the world to which the text refers. However, to perform argumentation-oriented discourse analysis, one must assume certain conceptualisations by the speaker in order to interpret and reconstruct propositions and argumentation structures. These conceptualisations are rarely captured as a product of the analysis process. In this paper, 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 and its argumentation structures, facilitates intertextual analysis, and improves understandability of the analysis products. 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.

Keywords: ontological proxy; argumentation-oriented discourse analysis; conceptual modelling; ontologies; model expressiveness

1. Introduction

Discourse analysis helps us understand the structure, content and objectives of texts, contributing to better insights into how people say what they say, how they justify their claims and overall, how we construct knowledge. Usually, discourse analysis focuses on "saying, doing and being" [1], where saying refers to what is said, doing to the practice of speaking by the speaker, and being to his or her the social roles. Different discourse analysis techniques such as RST (Rhetorical Structure Theory) [2] or IAT (Inference Anchoring Theory) [3,4] focus on different purposes, being one of them the identification and study of argumentation structures. Argument-oriented discourse analysis usually proceeds by breaking down a text into meaningful chunks, such as 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 [5]. The final products of argument-oriented discourse analysis, in this manner, are diagrams and accompanying texts that describe what argumentation devices such as inferences, conflicts or rephrasings are being employed by the speaker.

Naturally, argument-oriented discourse analysis focuses on what is being said and follows the source text as literally as possible. This is a desirable property, as being faithful to the text minimises unwanted biases and spurious information that the analyst might otherwise inject. However, this also has the consequence that little or no attention is paid to the actual things in the world to which the text refers. But 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. In particular, proposition reconstruction (i.e., rewording the literal locutions in the text so that standalone propositions can be obtained) often plays a central part in argument-oriented analysis discourse, as illustrated by the IAT Guidelines [5]. And reconstructing propositions requires the analyst to guess or unveil what was in the mind of the speaker or authors, that their words make

sense. This mental model of the discourse domain that the analyst constructs is rarely mentioned in the discourse analysis literature, despite its apparent centrality. Consequently, it is rarely captured as a product of the analysis process, and usually lost forever. Readers or users of the analysis products must re-create this mental model in their heads again, possibly diverging from the interpretation adopted by the analyst, and thus hindering the communication and utility of the analysis products.

When the text being analysed involves a dialogical situation in which two or more agents exchange arguments, this issue becomes even more important. Analysis of dialogical texts requires the analyst to discover the common ontology shared by the speakers and interpret 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 rarely refer to it. In this manner, the reconstructed propositions and argumentation relations are only anchored on the text but not on the world external to it, 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.

In this paper we argue that the mental model that the analyst develops in relation to the discourse being analysed should be captured during argument-oriented discourse analysis, and documented as a proper analysis product, so that users of the diagrams or other artefacts that result from the analysis can refer to it as necessary. To do this, we propose the use of conceptual models to represent the relevant parts of the world that the text refers to. In addition, we argue that detailed connections should be made between the conventional products of argument-oriented discourse analysis, usually diagrams, and these conceptual models, so that tracing between discourse and world becomes feasible. These connections are mediated by conceptual artefacts named ontological proxies.

2. Materials and Methods

The approach followed in this paper is based on conceptual modelling. This means that we consider that the product of an argument-oriented discourse analysis effort is a *conceptual model*, i.e., a formalised representation of a part of the world in terms of concepts as dictated by a given formalism or modelling language. Conceptual models are powerful because they represent a part of the world through controlled simplification so that we can reason on them and apply the results of our reasoning back to the part of the world being represented [6]. For example, we can represent the geography of a place through a digital map in a Geographical Information System, reason on the digital map (for example, by measuring the distance between two villages), and then apply the conclusions of our reasoning back to the physical world (we expect these villages to be at the measured distance). Conceptual models are composed of modelling elements, which are formalised concepts that adhere to a given formalism or modelling language. This modelling language is usually described through a metamodel, which defines what kinds of modelling elements, or primitives, there may be. For example, many modelling languages such as ConML [7] or UML [8] establish that the world is to be described in terms of primitives such as *Type* and *Instance*, or equivalent ones.

In our case, the part of the world being modelled is the discourse under analysis, and the modelling language is a more or less explicit collection of primitives from which the analysis products are constructed. In our work we use an extended version of IAT [3,4], which defines basic modelling primitives such as *Locution*, *Proposition*, *Inference* and *Illocutionary Force*, as well as specific relationships between them. Even though IAT has not been described through an explicit metamodel, its major "building blocks" (locutions, propositions, inferences, etc.) can be readily characterised from the literature. In this manner, performing a discourse analysis with IAT entails re-expressing what the text says in terms of IAT's primitives, i.e., what locutions there are, how they are reconstructed into propositions, how illocutionary forces anchor each proposition onto a locution, how inferences connect propositions to drive the argumentation from premises to conclusions, and so on. In this manner, the final product of an argumentation-oriented discourse analysis effort is a conceptual model of the discourse, which describes the discourse in terms of the above-mentioned modelling primitives. We will call this model a *discourse model*.

In addition, the central thesis in this paper is the need for every discourse model to be accompanied by a conceptual model of the discourse domain, or part of the world to which the text refers. We will call this model a *domain model*.

At this point, we must make a clarification. Within information technologies, the representation of the world has been approached from two different disciplinary traditions and has thus generated two different sets of terms and assumptions. In the world of software engineering, the term "conceptual model" is often used, whereas in the tradition of artificial intelligence and computer systems, the term "ontology" is more common. The commonalities between conceptual models and ontologies are far more numerous than their differences [9–11], so we will use "conceptual model" in this paper despite the fact that "ontology" should work equally well.

In this manner, the fact that both the discourse model and the domain model are both conceptual models allows us for a homogeneous treatment as well as their interconnection, as we explain in further sections. Figure 1 summarises our approach.
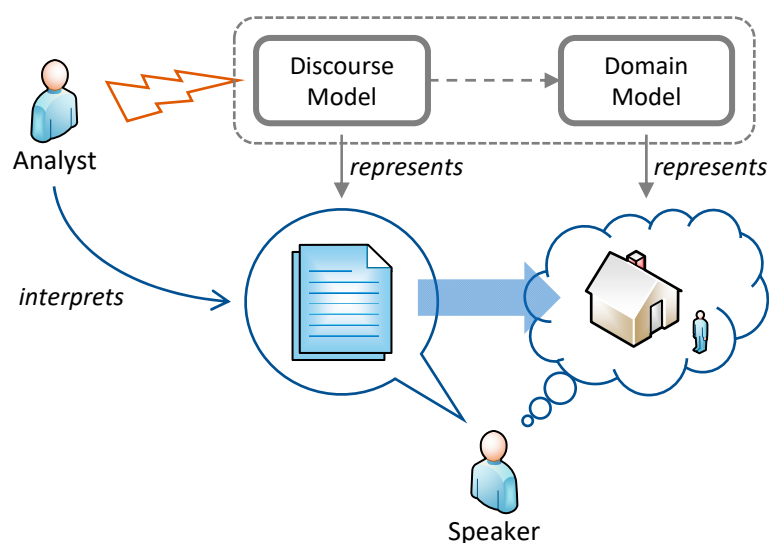
Figure 1. A speaker produces a discourse (centre) referring to a part of the world in their mind (right-hand side). By looking only at the discourse, an analyst creates a discourse model to represent the discourse, plus a domain model to represent the associated domain. Since the discourse refers to the domain (thick arrow, bottom), the discourse model must somehow refer to the domain model (dashed arrow, top).

There is an extensive body of literature on conceptual modelling (as well as ontologies), and conceptual modelling is practised today through the use of many techniques, languages and tools, such as ConML [6,7], OntoUML [12] or OWL [13]. To express discourse models, as introduced above, we employ a slightly modified version of IAT [3,4], supplemented with details from the Periodic Table of Arguments [14,15], which we tentatively call IAT+. The details of IAT+ are out of the scope of this paper, but they should not matter for the current discussion, as the approach that we propose is independent of the particular formalisms chosen for modelling; this is elaborated further in Section 4.

On the other hand, we chose ConML to express domain models, as it is especially suited to the representation of soft issues such as vagueness, temporality and subjectivity [16], which are often important in discourse analysis. A full description of ConML is out of the scope of this paper, but we can offer a brief description. ConML is a general-purpose conceptual modelling language especially oriented towards the humanities and social sciences. It is based on the object-oriented paradigm, so its metamodel defines modelling primitives such as *Class*, *Attribute*, *Association*, *Object* and *Link* [6,7]. This means that ConML models represent parts of the world in terms of what categories of things (classes) there are, what properties they have (attributes), how they relate to each other (associations), what particular entities exist (objects), and how they are connected one another (links).

Even though the discourse and domain models are both conceptual models, they are expressed in terms of different languages (IAT+ and ConML, respectively), and thus they must be considered two separate models rather than one. Keeping these models separate also makes sense for modularity reasons. For example, an intertextuality study addressing commonalities and differences between related texts may want to use a common domain model for the whole collection of texts, but obviously one discourse model for each of them. In this manner, the relationship between discourse models and domain models (top of Figure 1) is many-to-one.

An example my help here. Consider the following excerpt of an interview with Spanish architect and cartoonist Peridis [17]:

People tend to go down South, where there is wealth and work. And they expel the Muslim population.

Here, the speaker is describing the fact that in the past people migrated from the North of Spain to the South, which was wealthier, and in doing so they expelled the Muslim population living there. Note that the North is not mentioned in this fragment, but it is in a sentence right before the text in the excerpt. Similarly, the fact that the speaker is talking about Spain is not stated in the text, but we infer it from the context. Using IAT+, we would model this fragment as depicted in Figure 2.
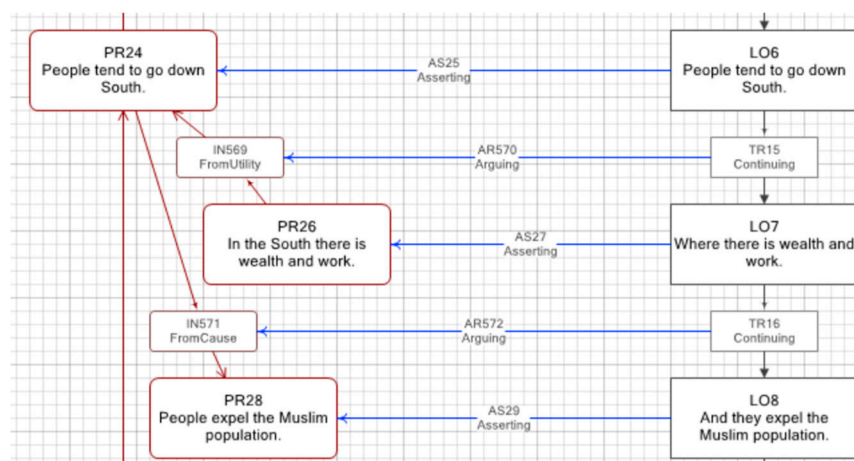
Figure 2. An IAT+ diagram showing the text fragment mentioned above. Locutions are shown as large boxes on the right-hand side, whereas propositions are shown as large boxes on the left. Note that two inferences, labelled IN569 and IN571, indicate how propositions are argumentatively related. The diagram was prepared with LogosLink, a software tool developed by the author.

The diagram in Figure 2 constitutes a small part of a larger discourse model. To construct this model, the analyst had to interpret what the speaker meant. Expressions such as "the North" or "the South" are especially difficult, as the text bears no reference to what geographical area is being discussed. Similarly, the text does not state that the mentioned Muslim population was living in the South; this must be derived through interpretation. In the absence of an explicit domain model, the discourse model depicted above fails to convey the necessary information to the reader, who must interpret the diagram themselves to, luckily, arrive at the same mental model as the analyst who created it.

A domain model of this text fragment would look like the one depicted in Figure 3.
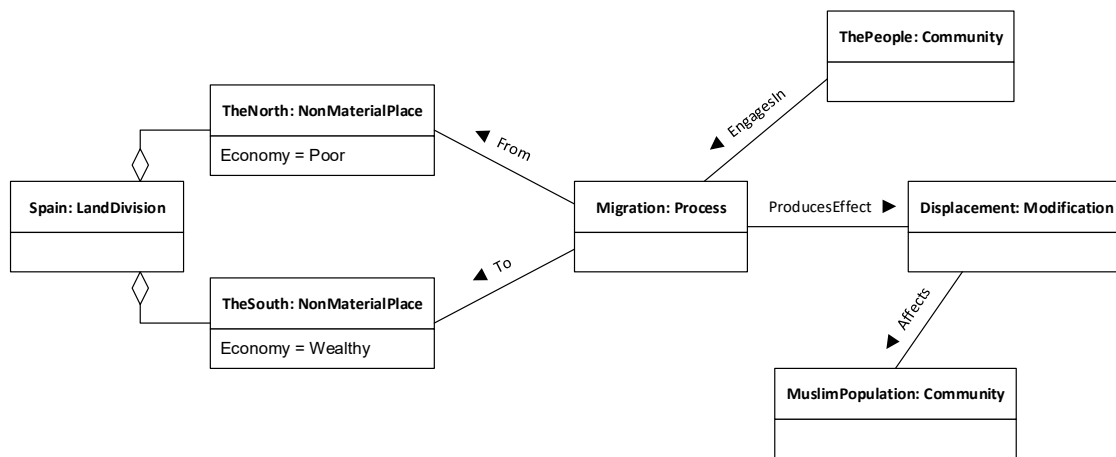
Figure 3. A ConML diagram showing a domain model for the text fragment mentioned above. Boxes represent entities in the world. For each one, an identifier and a category are given, separated by a colon. For some entities, values are stated, such as in the case of *Economy = Wealthy* for *TheSouth*. Lines connecting boxes stand for links between entities and are labelled accordingly.

This domain model represents the major things that are explicitly mentioned by the speaker, such as the South or the Muslim population. It also represents other things that do not appear in the text but we know are there, such as the North (which weas mentioned by the speaker in previous locutions), Spain (which the analyst inferred from the context), or the migration process (which is implied by the speaker). All in all, this domain model captures the interpretation that the analyst made of the discourse and can be used as a reference to better understand the discourse model.

At this point, the question remains as to how elements in the discourse model should be connected to elements in the domain model, as depicted by Figure 4.
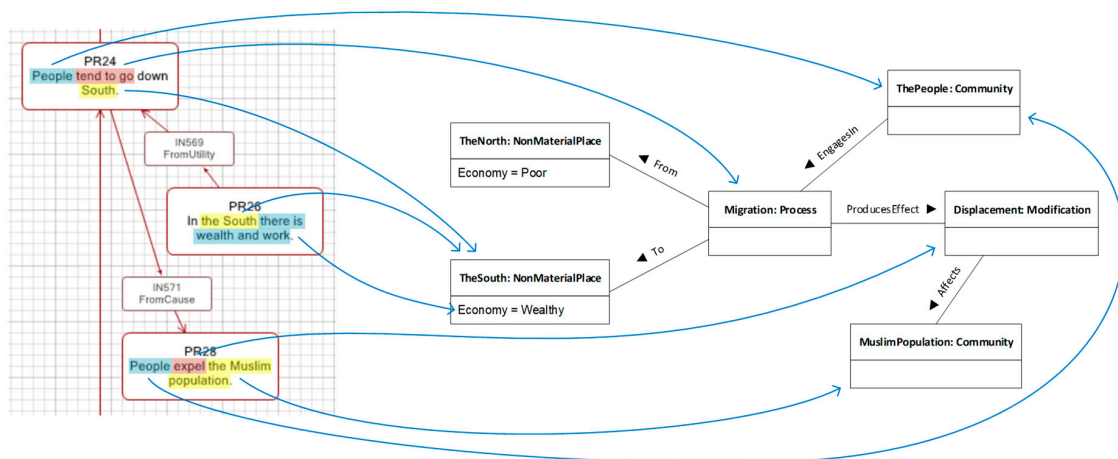
Figure 4. Diagram fragments for the discourse and domain models are displayed here. Blue arrows connecting them stand for the expected connections between elements in the discourse and elements in the domain. Discourse fragments have been highlighted in different shades for clarity. For example, the words "tend to go" in proposition PR24 (top left) must be connected to the *Migration: Process* entity (centre right).

The discourse and domain models are different models, each using a different language, so there is no common formalism that may establish the rules for the necessary connection. In other words, neither the metamodel of IAT+ or ConML can represent both propositions and entities in the world. In addition, IAT offers no modelling primitive to represent fragments of a proposition, such as "tend

to go" or "there is wealth and work" in Figure 4. To address these issues, we propose the notion of *ontological proxy*, as well as the related notion of *denotation*.

3. Results

An ontological proxy is an element in a discourse model that stands for another element in the associated domain model, and which may be referenced by multiple propositions. Let us unpack this definition and explore its consequences.

- Ontological proxies are model elements. This means that, like any other model elements, they are formalised concepts in the mind of the analyst [6] and are usually communicated via depictions in diagrams or other media.
- Ontological proxies are elements in the discourse model. This means that the IAT+ metamodel must contain suitable modelling primitives to accommodate them. In other words, the IAT+ metamodel must define primitives for ontological proxies as well as locutions, propositions and inferences.
- Every discourse model must have an associated domain model. As we introduced above, a common domain model may be shared by multiple discourse models, but every discourse model must have one and only one domain model.
- Each ontological proxy stands for one element in the associated domain model. By "stand for" here we mean that they can work as simpler replacements of the referred to domain elements, since both an ontological proxy and the associated domain element represent the same thing in the world. It is for this reason that they are called "proxies".
- Ontological proxies must be simpler than the associated domain elements; otherwise, there would be no point in using them. Also, and for the sake of modularity, ontological proxies must be as independent as possible from the modelling language employed to express the domain model. For these two reasons, ontological proxies must be lightweight and minimal.
- Each ontological proxy may be referenced by multiple propositions. Actually, it is fragments of propositions what refer to ontological proxies, as highlighted in Figure 4. Each proposition fragment that refers to an ontological proxy is called a denotation.

These consequences have been used as design criteria to extend the IAT+ metamodel and incorporate the necessary constructs to support ontological proxies. The following subsections describe these criteria and the associated implementation in greater detail.

3.1. IAT+ Metamodel

As described above, the IAT+ metamodel must provide modelling primitives to express ontological proxies and denotations. Figure 5 shows the relevant part of the metamodel.
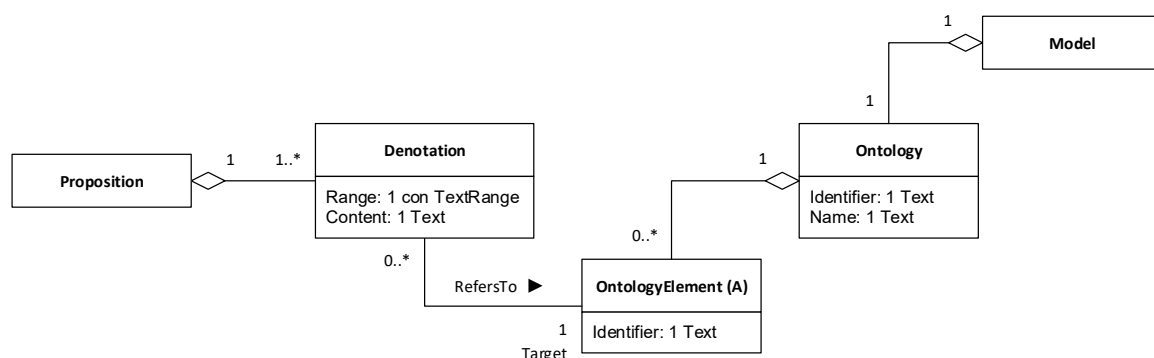
Figure 5. Diagram depicting a section of the IAT+ metamodel. *Model* on the top right refers to discourse models. *Ontology* refers to domain models (but see text below for details).

According to the metamodel, every discourse model (simply called *Model* in Figure 5) has an associated domain model (called *Ontology* in the figure). We said in previous sections that multiple discourse models can share a common domain model. However, the Ontology class in Figure 5 does not represent domain models themselves, but the proxy image of a domain model that is kept by a discourse model. In other words, and from the perspective of a discourse model (*Model* in Figure 5), *Ontology* represents a private and simplified copy of the associated ontology. Consequently, this relationship has been modelled as a one-to-one whole/part association.

Furthermore, every private and simplified ontology contains a number of ontological proxies, called ontology elements in the metamodel. *OntologyElement* is an abstract class, as indicated by the "(A)" marker in Figure 5. This means that it has a number of subtypes representing different kinds of ontology proxies, which we discuss below.

Reading now from the left-hand side of the diagram, every proposition has a number of denotations. A denotation is a fragment of a proposition that refers to an ontology element. The concept of denotation allows us to pick specific words or phrases in a proposition that clearly refer to an element in the ontology, such as "tend to go" in PR24 or "the South" in PR26 in Figure 4.

Figure 6 depicts a sample instance model conforming to the metamodel in Figure 5.
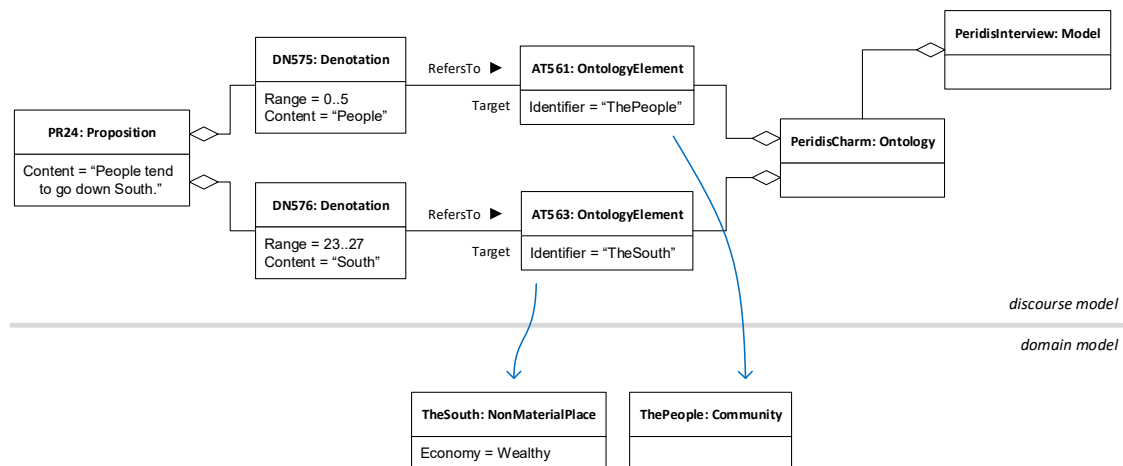

Figure 6. Diagram depicting how ontological proxies work. Above the line, an instance model conforming to the metamodel in Figure 5 is shown, stating that proposition PR24 has two denotations for "People" and "South". Each denotation refers to a particular ontological element of the discourse model's associated domain model (ontology). Below the line, a fragment of the associated domain model from Figure 3 is shown. Blue arrows across the line depict the fact that ontological elements work as proxies to elements in the domain model, as shown by the matching identifiers "*ThePeople*" and "*TheSouth*".

In the figure, the ontological proxies are the objects of type *OntologyElement*. These objects have an *Identifier* value whose contents match the identifiers of elements in the domain model. This matching relationship is what makes ontological proxies to work as, precisely, proxies. Note that, in the diagram, proxy relationships are shown as blue arrows between the associated elements, but they do not exist as formal relationships as such, since, as we explained above, the discourse and domain models are expressed using different languages. In any case, both human users of the models as well as computers processing them can easily find these matches and thus follow the proxy relationships.

As we said above, and as depicted in Figure 5, *OntologyElement* is an abstract class and has a number of subtypes, corresponding to the different kinds of ontology elements that are common in domain models. Of course, there are many languages that one could use to express a domain model, so the IAT+ metamodel must be generic enough as to cater for as many as possible. For this purpose, we decided to implement a small but varied range of subtypes of *OntologyElement*, which the design goal that at least languages such as ConML, OntoUML and OWL should be supported. Most conceptual

modelling languages adopt an object-oriented approach and hence include primitives such as *Class*, *Attribute*, *Object* and *Link*. However, terminology varies between languages, and the specific semantics of the major primitives are also slightly different. Most languages, however, share the fact that they distinguish clearly between types and instances (or categories and entities, depending on the terminology used) as a major architectural principle around which their metamodels are organised. This means that ontological elements could also be organised along these lines. However, we felt that adopting a multilevel modelling approach [18,19] would entail little extra complexity but provide a much richer and more expressive ontological infrastructure. Multilevel modelling allows chains of type/instance relationships of arbitrary length, thus enabling the homogeneous treatment of types and instances for many common purposes and supports higher-order types with a rather simple structure. For these reasons, we adopted the multilevel modelling principles sketched in [20] and designed the *OntologyElement* subtype hierarchy shown in Figure 7.
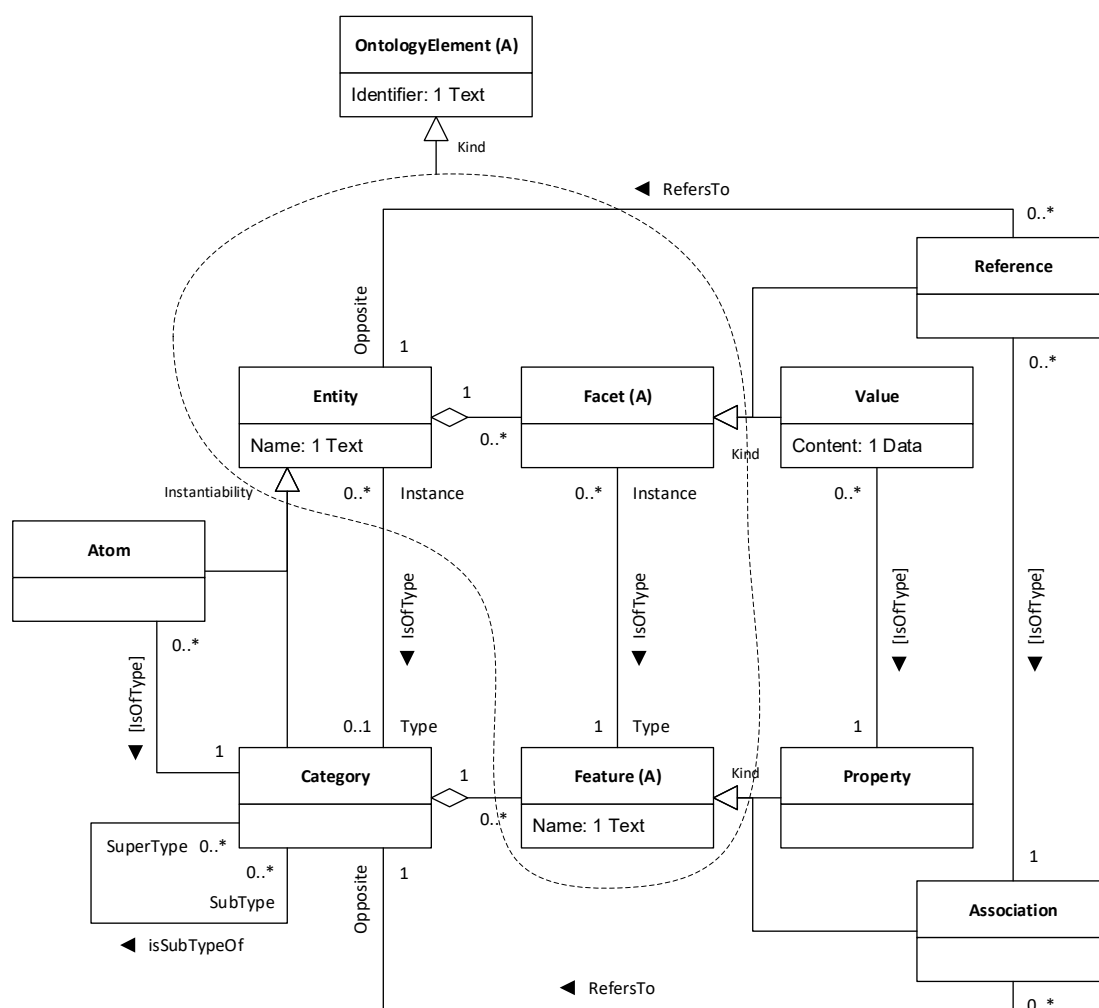
Figure 7. Part of the IAT+ metamodel showing the class hierarchy under *OntologyElement*. Please see the text below for a detailed description of each model element.

The first subtype of *OntologyElement* is *Entity*, which represents things in the world such as the computer I am using, my house, the Second World War, or the 5/2016 Act on Cultural Heritage, for example. Anything in the world may be an entity. Entities are characterised through facets of two kinds: values and references. Values represent atomic qualities or quantities of entities, such as the fact that I am 53 years old or that the Second World War began in 1939. References, in turn, represent

connections between entities, such as the fact that I (an entity) work at Incipit CSIC (another entity), or that the 5/2016 Act on Cultural Heritage (an entity) applies in Galicia, Spain (another entity).

Entities come in two kinds, depending on whether or not they can be instantiated, as described in the multilevel modelling literature [18,20]. Some entities are not instantiable, that is, they cannot work as templates for other entities. These are called "particulars" (and sometimes "atoms") in philosophy, "ur-elements" in mathematics, or "objects" in the object-oriented approach in software engineering. We call them atoms. Some examples of atoms include myself, the Second World War, or the 5/2016 Act on Cultural Heritage.

Some other entities, as opposed to the previous, can be instantiable into other entities, working as templates for them, and usually corresponding to generic concepts or ideas. For example, the notion of *Tree* can be instantiated into individual trees, such as each of the trees I can see through the window as I type this sentence. Similarly, the notion of *Person* is instantiated into each individual person. These instantiable entities are called "universals" in philosophy or "classes" in object-oriented software engineering. We call them categories. In general, we can say that every entity has a category as type, since, in the words of George Lakoff, "There is nothing more basic than categorization to our thought, perception, action, and speech" [21]. For example, I am of the *Person* category, the Second World War is of the *ArmedConflict* category, and the 5/2016 Act on Cultural Heritage is of the *Law* category. In practice, and especially when constructing ontologies with some degree of uncertainty, we do not know or are not interested in the category of some entities, so specifying them is not mandatory.

Now, since categories are also entities, they can have values and references. In addition, they can be characterised through two extra kinds of features: properties and associations. Properties define possible values of the entities of the category. For example, since every person has a value for their age, then we can capture this fact by stating that the *Person* category has an *Age* property. Similarly, associations define possible references of the entities of the category. For example, since every person has been born in a particular place, then we can capture this fact by stating that the *Person* category has a *WasBornIn* association towards the *Place* category.

In this manner, the IAT+ metamodel supports ontological proxies of six concrete kinds: atoms, values, references, categories, properties, and associations. Although some types of modelling primitives are not covered (such as OntoUML non sortals, for example), these six kinds map nicely to the major modelling primitives of almost any conceptual modelling language, as exemplified by Table 1.

Table 1. Mappings between IAT+ ontology element subtypes and modelling primitives of common conceptual modelling languages.

IAT+	ConML	OntoUML	OWL
Atom	Object	*(not supported)*	Individual
Value	Value	*(not supported)*	DataProperty
Reference	Reference	*(not supported)*	ObjectProperty
Category	Class	RigidSortal	Class
Property	Attribute	Property	*(handled through axioms)*
Association	Semi-Association	Relation	*(handled through axioms)*

We must also remark that the notation used in Figure 6 is convenient to visualise the details of the data structures implementing the models. However, we suggest a different notation for most practical purposes, which is shown in Figure 8.
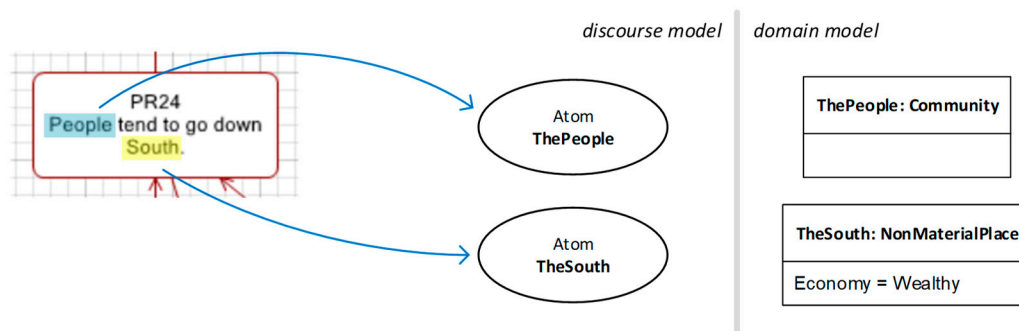
Figure 8. This depicts the same situation that was shown in Figure 6, but using the IAT+ notation introduced earlier plus some additional lines and symbols. Ellipses represent ontological proxies, that is, instances of *OntologyElement* in Figure 7. Matching elements in the domain model are shown to the right.

The following sections provide guidance on how to find ontological proxies as well as some examples to illustrate how they can be used in practice.

3.2. How to Construct Ontological Proxies

As we described in previous sections, ontological proxies are model elements. This means that they are mental constructs that adhere to a well-known formalism or modelling language. In this section we tackle the issue of how ontological proxies, as model elements, are constructed.

As explained above, ontological proxies are referred to by fragments of propositions. In Figure 8, for example, the fragments "People" and "South" are highlighted to indicate that they correspond to denotations, each of them referring to an ontological proxy. So, in order to determine what ontological proxies must be constructed for a given proposition, we must take into account the following guidelines.

First, it is important to acknowledge that conceptual modelling is always done for a purpose, i.e., it is a situated activity driven by a goal. Two models of the same part of the world but pursuing different goals are likely to be very different. In addition, conceptual modelling, as a concept-creation process, is clearly dependent on the subjective traits of the analyst such as academic and cultural background or personal preferences. Consequently, it is impossible to provide clear-cut rules as to how construct ontological proxies; only approximate guides can be offered.

Having said this, it is safe to say that the process to construct ontological proxies is often driven by an examination of the lexicon and grammar employed by the proposition at hand, with the goal of answering the question "what is this sentence talking about?". For example, in "People tend to go down South" in Figure 8, we can observe the following:

- The subject "People" refers to an uncertain group of persons.
- The verb "tend to go down" indicates a movement of said person group.
- The complement "South" refers to the destination place of this movement.

This means that the proposition contains three denotations, which in turn hint at three potential entities: a group of people, a movement process, and a destination place. It also hints at some connections between them: "People" points at the thing that is moving, and "South" points at the destination of such a movement. The source place of movement is unsaid, at least by this proposition. We can represent this by the domain model depicted in Figure 9.
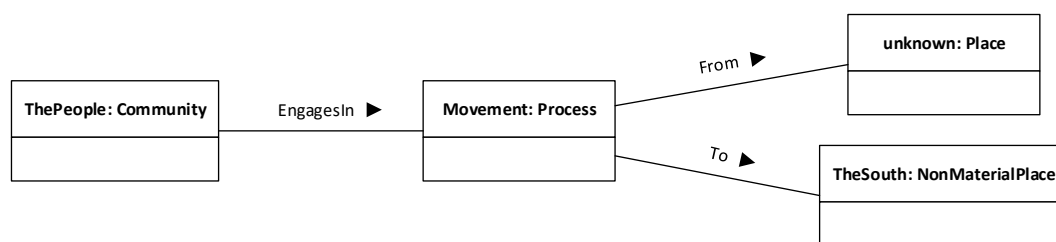
Figure 9. Domain model representing the observations made from the analysis of proposition PR24 in Figure 8.

Note that, in the domain model, the source place of the movement is unknown. We show it in the diagram for completeness sake, and because it is likely that the analysis of another proposition in this discourse does refer to it, which would allow us to refine this domain model. The domain model, as is, contains four entities, named by the identifiers *ThePeople*, *Movement* and *TheSouth* plus the keyword *unknown*. Note also that we have chosen particular categories for these entities: *ThePeople* is a community, *Movement* is a *Process*, *TheSouth* is a *NonMaterialPlace*, and the unknown source of the movement is a *Place*. Other options may be also valid. For example, stating that the people referred to by "People" in the proposition make up a community may be too bold, as we have no guarantee, from the text being analysed, that they in fact do; these people may actually be a scattered collection of groups and families with little or no relation to each other, so we are not justified in categorising them collectively as a community. In this case, we should rather employ a different category such as the non-committal *GroupOfPeople*. Choosing the right category is not always easy, as often there is not much information in the text about what "right" means in this context. Using a domain-specific reference model or ontology can be useful, as it would offer a catalogue of common concepts in the domain to choose from. For our examples we have used CHARM, the Cultural Heritage Abstract Reference Model [22–24], which lists over 200 concepts related to cultural heritage and associated topics plus their properties and relationships.

In this example, all the denotations refer to entities in the world. Other propositions may refer to other kinds of ontological elements, such as values or references. For example, PR26 in Figure 4 states that "In the South there is wealth and work". Here, the fragment "there is wealth and work" can be interpreted as denoting a value for the entity denoted by "the South", namely the predication that the economy of the South is good (or wealthy, as depicted in the figure).

In general, proper nouns or qualified noun phrases, such as "the South" or "the Muslim population" usually denote material or immaterial entities. Verbal phrases headed by dynamic verbs such as "tend to go down" or "expel" usually denote processes or activities. Both can be modelled through *Entity* ontological proxies. Verbal phrases with stative verbs, such as "there is" or "have" often denote predications of values or references on the subject entity, which can be modelled through *Value* and *Reference* ontological proxies. Adjectival clauses such as "wealthy" or "long and difficult" usually denote the content of values or references. A special mention should be made of phrases with the verb "to be", as this verb may carry different meanings in many languages. In English, for example, "to be" may indicate either existence ("there is a person"), which would be modelled through an *Entity*; identity ("she is my mother"), which can be also modelled as an *Entity* plus a *Reference*; predication ("she is tall"), which is best modelled as a *Value* or a *Reference*; classification ("this is a house"), which can be modelled through an *Entity* and a *Category*; or subsumption ("a house is a structure"), which should be modelled through two related instances of *Category* [6]. Sentences containing "to be" must be carefully analysed.

Not that this lexical and grammatical analysis of propositions allows us to construct a domain model, rather than ontological proxies themselves. Ontological proxies, by definition, are lightweight replacements for elements in the domain model, so once this model is clear, an ontological proxy can be constructed for each model element. Coming back to the example in Figure 9, we would construct three ontological proxies, all of them of the *Atom* kind: one for *ThePeople*, one for *Movement*, and one for *TheSouth*.

As we proceed to analyse more propositions in the same discourse, we would be adding to the domain model, or altering it to accommodate new elements. For example, it is likely that another proposition tells us something relevant to identify the source place of the movement in Figure 9, or add extra detail to any of the associated entities. Conceptual modelling is usually an iterative and incremental task, which eventually converges to a stable resolution.

3.3. Usage Examples

Let us look at some examples of ontological proxies in practice. Firstly, let us focus on the issue of how ontological proxies may help us to document particular interpretations of the discourse. Consider the following fragment:

Alice: The 5/2016 law says that you cannot build close to a protected site.

Bob: But the law also says that I have the right to buy and possess any land.

A first approach to analysing this fragment may interpret the exchange as a conflict, since "the law" in Bob's line refers to the same thing as "The 5/2016 law" in Alice's. In fact, the "But" lexical marker heading Bob's retort is a usual indicator of conflict. This interpretation is captured by the models depicted in Figure 10.
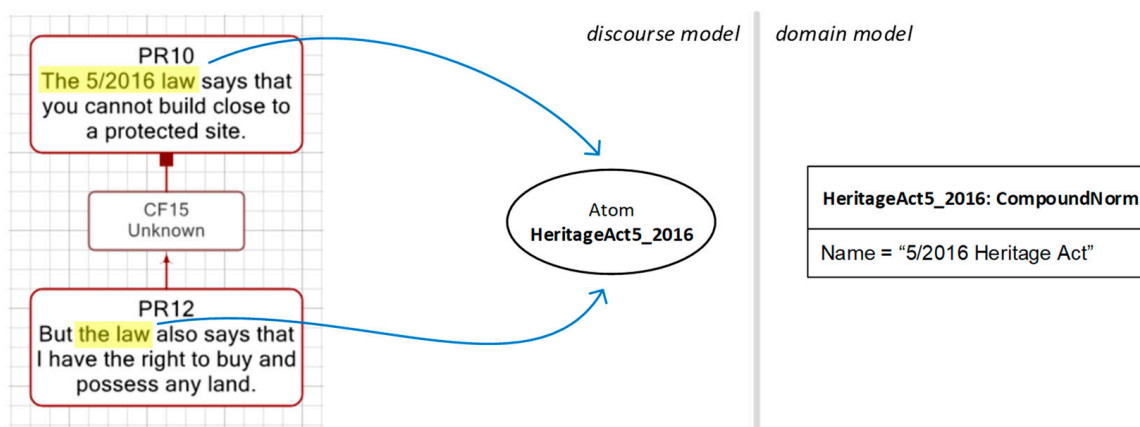
Figure 10. Discourse and domain models for the interpretation that "The 5/2016 law" and "the law" refer to the same thing.

However, an alternative interpretation is possible. The denotation "the law" in Bob's line may refer to the general laws and regulations that apply, rather than the 5/2016 Heritage Act in particular. If this is the case, then Bob is saying that regulations, in general, allow you to buy and possess any land, which may not be a conflict with Alice's proposition after all, as the 5/2016 Heritage Act could be making an exception to the general right to buy and possess land. This alternative interpretation is captured in Figure 11.
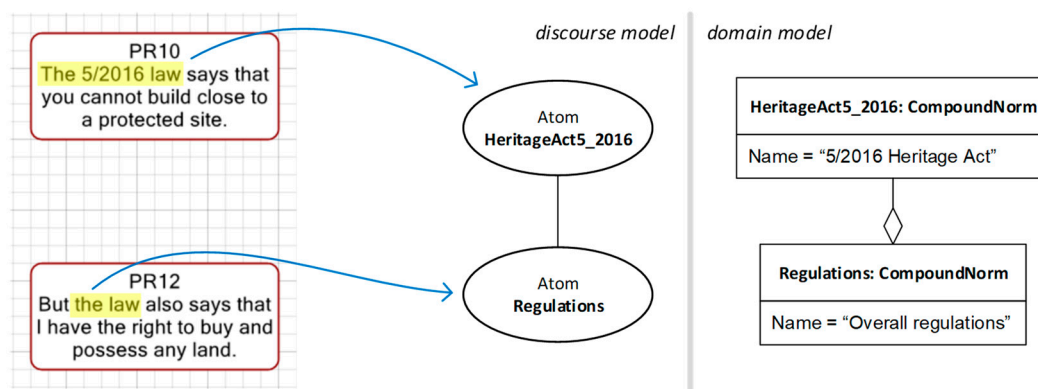
Figure 11. Discourse and domain models for the interpretation that "The 5/2016 law" and "the law" refer to different but related things.

Here, two ontological proxies exist, capturing the facts that the 5/2016 Heritage Act is part of a larger set of overall regulations. Once this interpretation has been established, it is clear that there is no necessary conflict between propositions PR10 and PR12, as shown. Note that, in the absence of ontological proxies, the two discourse diagrams (corresponding to the boxes displayed on a grid) from Figures 10 and 11 would show different options but with no associated explanation. A reader of these models would find no information as why a conflict was or was not described between the propositions. Once we incorporate the ontological proxies, however, and even in the absence of the domain model, the interpretation of the discourse becomes clear.

Let us now move to a different example and focus on how ontological proxies can work to assist in lexical/semantic studies. Consider the following text [17]:

> *People tend to go down South, where there is wealth and work. And they expel the Muslim population. The North was hard, and they got rumours about Al Andalus being like an Eden.*

Here, two terms, "the South" and "Al Andalus", are being used to refer to the same thing. This interpretation is shown in Figure 12.

First, note that propositions PR24 and PR26 use "South" or "the South" to refer to the southern region of Spain, whereas PR43 uses "Al Andalus" to refer to the same place. This is interpretation is clearly documented by the single ontological proxy labelled *TheSouth*. Once this has been established, it is easy to see why PR43 works as a premise (together with PR30) for inference IN573 and leading to the conclusion PR24: living in the North was hard, and since people got rumours that Al Andalus was like an Eden, they moved there. This argument only makes sense if we assume that Al Andalus and the South are the same thing. Again, this assumption is clearly documented through ontological proxies and thus works as grounding to support inference IN573.

Finally, let us consider how ontological proxies may be useful to intertextual studies. Consider the following fragments, taken from different tweets in March 2020:

> *Speaker 12: She was a simple woman who became a heroine.*

> *Speaker 35: Her father was a victim, and so is she.*

Here, speakers 12 and 35 are not engaged in a dialog, and probably they do not even know about each other. But both are discussing the late Ascensión Mendieta, a Spanish activist for Historic Memory who struggled to restore the memory of her father, killed by Franco's dictatorship in 1939. We know this because both tweets were inserted in threads where Mendieta was named. Figure 13 depicts the models for both fragments.
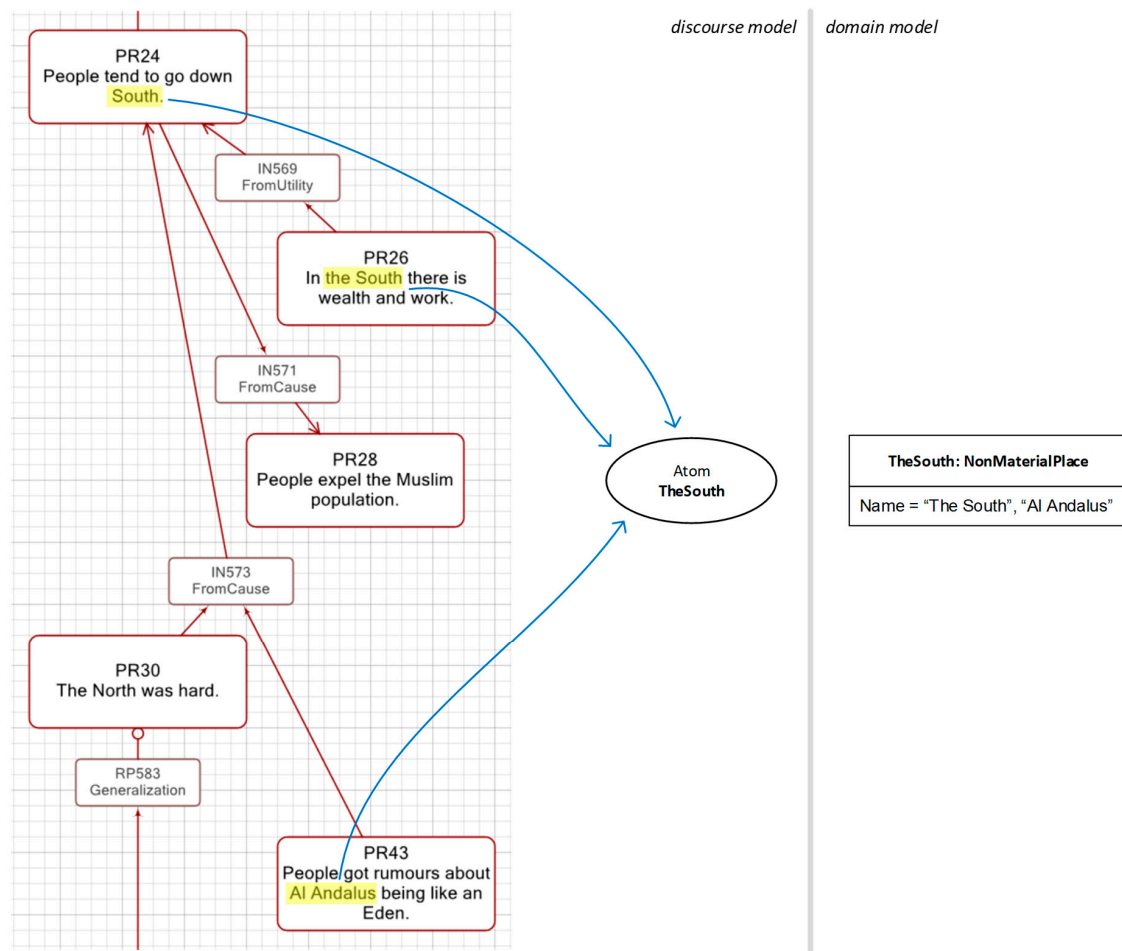
Figure 12. Discourse and domain models for the interpretation that "the South" and "Al Andalus" refer to the same thing.

In this example, the denotations "she" in both discourse models point to an atom labelled *Mendieta*. Both models contain a denotation pointing to a *Role* value for this atom, but with different contents: model 12 states that Mendieta is a heroine, whereas model 25 states that she is a victim. The domain model is shared between the two discourse models. In it, we can see a single object *Mendieta* with subjectively marked values for the *Role* attribute, corresponding to each of the *Role* values in the discourse models. The modelling of subjectivity is out of the scope of this paper, but a brief introduction can be found in [16]. Essentially, each of the lines starting with "Role" in the Mendieta box in the domain model stands for a value given to this object by a different agent, namely, our speakers 12 and 35. In this manner, two discourse models that were in principle disconnected and structurally unrelated are linked together through a common domain model that documents the associated speaker perspectives. This captures the fact that both discourses are referring to a common set of concepts in the world. This example only involves two discourse models, but this approach can be applied with any number of discourse models as long as all of them refer to a common set of things in the world.
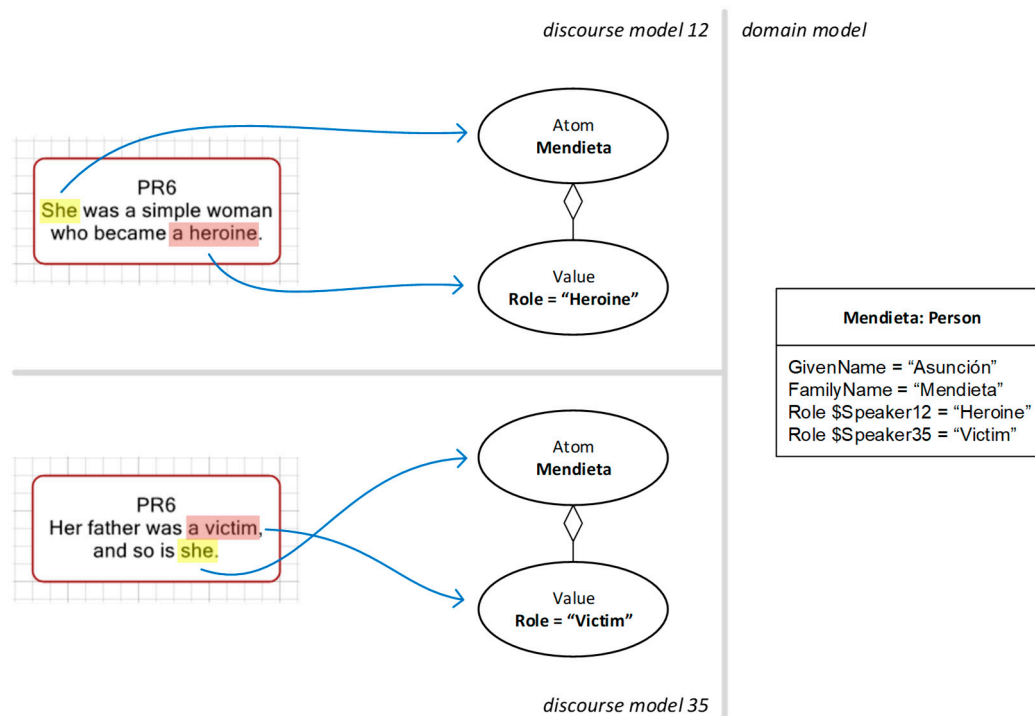
Figure 13. Discourse and domain models for the fragments above. Note that the two discourse models share a common domain model.

4. Discussion and Conclusions

The previous sections have presented the notion of ontological proxy and described how ontological proxies can be used to better express domain facts that are relevant to the discourse being analysed.

Various aspects must be highlighted. Firstly, ontological proxies are independent of the specific languages or approaches that one employs for discourse or domain modelling. We have chosen IAT+ and ConML, but ontological proxies do not rely on these choices. Rather, they are an abstract device that mediates between a discourse model and a domain model, whatever formalisms are used to express them. As we stated in Section 3.1, the six concrete kinds of ontological proxies (atoms, values, references, categories, properties, and associations) map nicely to the major modelling primitives of almost any conceptual modelling language.

Secondly, ontological proxies are part of the discourse model. This means that the discourse model is autonomous and does not need an accompanying domain model to stay expressive. In fact, we could remove the right-hand side in every figure in Section 3.3, and the diagrams would still be understandable. Of course, ontological proxies are proxies, and therefore lightweight, so they do not contain every detail that the full domain model can offer. This is especially clear, for example, in Figure 13, where the fact that there are multiple perspectives on Mendieta's historical role cannot be seen but in the domain model. Still, ontological proxies provide a good balance between expressiveness and conciseness, which arguably would minimise the need to retrieve and examine the domain model in most situations. In addition, the fact that the connections between discourse and domain models are established via lightweight elements acknowledges the principle of modularity that has been crucial in software engineering since at least the 1980s [25]. According to this principle, discourse and domain models are kept separate (they are different "modules") but connected through few and weak links, namely, the mappings between ontological proxies and elements in the domain model. This allows each of these two artefacts to live separately, using whichever formalism is required for each one, but still be connected when needed.

Another relevant issue is the fact of limited expressiveness. Since ontological proxies are simpler replacements for domain model elements, they are limited by how expressive the chosen modelling

language is. In this paper we have used ConML, which is capable, for example, of representing different subjective views on the same things, or temporal change, with minimum burden, as it provides specific mechanisms to do it. Not all modelling languages do this. If the domain modelling language of choice did not offer a similar mechanism to represent subjective views, for example, propositions such as "As opposed to the local government, tourists often think that the cathedral urgently needs repairs" would difficult to analyse and express, as the opposed subjective views described by it could not be satisfactorily represented by any primitive in the language. In this regard, and despite the fact that ConML is highly expressive [16], it still lacks support for irrealis moods such as conditionals or imperatives, so ontological proxies for denotations using these modalities are difficult or impossible to represent properly.

The theoretical proposal introduced in this paper has been implemented in the LogosLink software tool, as mentioned in Section 2, and has been applied to the analysis of texts from different sources, including live debates, tweets, press news and popular science articles. Currently, it is being applied to the analysis of a corpus of over 620 articles about covid-19 from the Spanish edition of The Conversation [26].

Future research directions include the following. First, the ConML language will be extended to support inequality predication, so that facts such as "I am older than 40" can be captured. Also, ConML will be extended to support various linguistic modalities such as deontic or hypothetical structures. This will allow domain models to become much richer and expressive, as described above. The subclasses of *OntologyElement* in IAT+ will be extended likewise so that propositions containing constructs like these can be adequately linked to domain elements. Additional extensions will be made to allow denotations to refer not only to specific ontological proxies, but also to the changes associated to them. This will allow, for example, to cater for statements expressing persuasion or change of mind, such as "I was convinced that the cathedral was fine, but now I see that it needs some repairs".

Finally, a comprehensive specification of IAT+, including a proper graphical notation, will be prepared and published. From the point of view of tool implementation, LogosLink will be updated with the new additions to IAT+, and support will be added for multi-model projects so that additional analytical options become possible, especially in relation to intertextual analysis.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Gee, J.P. *An Introduction to Discourse Analysis: Theory and Method*; Routledge: London, UK, 2014.
2. Mann, W.C.; Thompson, S.A. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text Interdiscip. J. Study Discourse* **1988**, *8*. [CrossRef]
3. Reed, C.; Budzynska, K. How Dialogues Create Arguments. In Proceedings of the ISSA Proceedings 2010. Available online: http://rozenbergquarterly.com/issa-proceedings-2010-how-dialogues-create-arguments/ (accessed on 20 October 2020).
4. Janier, M.; Aakhus, M.; Budzynska, K.; Reed, C. Modeling argumentative activity with Inference Anchoring Theory. Argumentation and Reasoned Action. Volume I. In Proceedings of the 1st European Conference on Argumentation, Lisabon, Portugal, 9–12 June 2015.
5. Centre for Argument Technology. Annotation Guidelines for Inference Anchoring Theory (IAT) with support for Conventional Implicatures (CIs). 2018. Available online: https://typo.uni-konstanz.de/add-up/wp-content/uploads/2018/04/IAT-CI-Guidelines.pdf (accessed on 20 October 2020).
6. Gonzalez-Perez, C. *Information Modelling for Archaeology and Anthropology*; Springer: Berlin/Heidelberg, Germany, 2018.
7. Incipit. ConML Technical Specification. Incipit CSIC. 2020. Available online: http://www.conml.org/Resources/TechSpec.aspx (accessed on 20 October 2020).

8.　OMG. Unified Modeling Language 2.5.1. 2017. Available online: https://www.omg.org/spec/UML/ (accessed on 20 October 2020).

9.　Henderson-Sellers, B. Bridging Metamodels and Ontologies in Software Engineering. *J. Syst. Softw.* **2011**, *84*, 301–313. [CrossRef]

10.　Atkinson, C.; Gutheil, M.; Kiko, K. On the Relationship of Ontologies and Models. In Proceedings of the 2nd International Workshop on Meta-Modelling (WoMM), Karlsruhe, Germany, 12–13 October 2006; Volume 96, pp. 47–60.

11.　Gonzalez-Perez, C. How Ontologies Can Help in Software Engineering. In *Grand Timely Topics in Software Engineering*; no. 10223; Cunha, J., Fernandes, J.P., Lämmel, R., Saraiva, J., Zaytsev, V., Eds.; Springer: Berlin/Heidelberg, Germany, 2017.

12.　Suchánek, M. OntoUML Specification. 2018. Available online: https://ontouml.readthedocs.io/ (accessed on 9 October 2020).

13.　World Wide Web Consortium. *OWL 2 Web Ontology Language*; World Wide Web Consortium: Cambridge, MA, USA, 2012.

14.　Wagemans, J. Four Basic Argument Forms. *Res. Lang.* **2019**, *17*, 57–69. [CrossRef]

15.　Wagemans, J. Period Table of Arguments. 2020. Available online: https://periodic-table-of-arguments.org/ (accessed on 16 October 2020).

16.　Gonzalez-Perez, C. Modelling Temporality and Subjectivity in ConML. In Proceedings of the 7th IEEE International Conference on Research Challenges in Information Science (RCIS 2013), Paris, France, 29–31 May 2013.

17.　Mantilla, J.R. *'Peridis': En Comarcas De La Montaña Palentina Nacen Ya Más Osos Que Niños*; El País: Madrid, Spain, 2020.

18.　Clark, T.; Gonzalez-Perez, C.; Henderson-Sellers, B. A Foundation for Multi-Level Modelling. In Proceedings of the Workshop on Multi-Level Modelling co-Located with ACM/IEEE 17th International Conference on Model Driven Engineering Languages & Systems (MoDELS 2014), Valencia, Spain, 28 September 2014; Atkinson, C., Grossmann, G., Kühne, T., de Lara, J., Eds.; CEUR-WS.org: Regensburg, Germany, 2014; Volume 1286, pp. 43–52.

19.　Atkinson, C.; Kühne, T. The Essence of Multilevel Metamodelling. In *«UML» 2001: Modeling Languages, Concepts and Tools*; Springer: Berlin/Heidelberg, Germany, 2001; Volume 2185, pp. 19–33.

20.　Almeida JP, A.; Frank, U.; Kühne, T. Multi-Level Modelling (Dagstuhl Seminar 17492). In *Dagstuhl Reports*; Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik: Wadern, Germany, 2018. [CrossRef]

21.　Lakoff, G. *Women, Fire, and Dangerous Things*; University of Chicago Press: Chicago, IL, USA, 1990.

22.　Gonzalez-Perez, C.; Martín-Rodilla, P.; Pereira-Fariña, M. Computer-Assisted Analysis of Combined Argumentation and Ontology in Archaeological Discourse. In Proceedings of the 46th Computer Applications and Quantitative Methods in Archaeology (CAA 2018), Tübingen, Germany, 19–23 March 2018.

23.　Gonzalez-Perez, C.; Parcero Oubiña, C. A Conceptual Model for Cultural Heritage Definition and Motivation. In Proceedings of the Revive the Past: 39th Conference on Computer Applications and Quantitative Methods in Archaeology, Beijing, China, 12–16 April 2011.

24.　Incipit. CHARM White Paper. Incipit, CSIC. 2016. Available online: http://www.charminfo.org/Resources/Technical.aspx (accessed on 20 October 2020).

25.　Meyer, B. *Object-Oriented Software Construction*, 2nd ed.; Prentice-Hall: Upper Saddle River, NJ, USA, 1997.

26.　The Conversation, Spanish Edition. 2020. Available online: https://theconversation.com/es (accessed on 16 October 2020).