



IAT/ML: a metamodel and modelling approach for discourse analysis

Cesar Gonzalez-Perez¹ · Martín Pereira-Fariña² · Beatriz Calderón-Cerrato¹ · Patricia Martín-Rodilla³

Received: 14 November 2023 / Revised: 8 July 2024 / Accepted: 5 August 2024 / Published online: 11 September 2024
© The Author(s) 2024

Abstract

Language technologies are gaining momentum as textual information saturates social networks and media outlets, compounded by the growing role of fake news and disinformation. In this context, approaches to represent and analyse public speeches, news releases, social media posts and other types of discourses are becoming crucial. Although there is a large body of literature on text-based machine learning, it tends to focus on lexical and syntactical issues rather than semantic or pragmatic. Being useful, these advances cannot tackle the nuanced and highly context-dependent problems of discourse evaluation that society demands. In this paper, we present IAT/ML, a metamodel and modelling approach to represent and analyse discourses. IAT/ML focuses on semantic and pragmatic issues, thus tackling a little researched area in language technologies. It does so by combining three different modelling approaches: ontological, which focuses on what the discourse is about; argumentation, which deals with how the text justifies what it says; and agency, which provides insights into the speakers' beliefs, desires and intentions. Together, these three modelling approaches make IAT/ML a comprehensive solution to represent and analyse complex discourses towards their understanding, evaluation and fact checking.

Keywords Natural language · Discourse · Argumentation · Ontologies · Metamodel · IAT/ML

1 Introduction

When we use language to communicate, we are doing different things at the same time. We are transmitting certain information to specific listeners, we are performing an action (e.g. warning someone that they might be in danger), and we are showing what we think to others (e.g. as a member of the health and safety team). Thus, the use of language is a practice, a “game” [68], which is guided by a set of implicit and explicit rules that determine who has acted appropriately and who has not, who is accepted in the group and who is not [20]. Discourse analysis focuses on the study of language in use, unpacking the discourse structures or strategies used in a specific context to communicate something and understand how it is done. This is useful in a number of ways: to gain a deeper understanding of complex discourses, to check

how well supported a claim is, or to develop a diagnosis and action plan for complex and polarised social issues such as migration or surrogacy motherhood.

However, discourse analysis is a hard task. It relies on an empirical analysis that goes beyond the automated application of quantitative measurements (such as word frequencies) and requires a theoretically grounded qualitative analysis to unpack the subtle strategies and patterns that are used to deliver a specific message to the audience and produce an impact on them. Therefore, discourse analysis requires the combination of quantitative methods to collect a representative and reliable selection of fragments of text and qualitative ones to manually analyse the selected texts under the umbrella of a specific theory. Thus, we propose to adopt a domain-specific approach to support the process of manual discourse analysis. To achieve this goal, this paper presents a formalised approach for discourse analysis that combines quantitative and qualitative techniques and is built upon three analysis perspectives: ontological, argumentation and agency.

The goal of ontological analysis is to capture the entities in the world that are referred to by the discourse and describe them via a conceptual model. Note that, here, we

Communicated by Dominik Bork and Henderik Proper.

✉ Cesar Gonzalez-Perez
cesar.gonzalez-perez@incipit.csic.es

¹ Incipit CSIC, Santiago de Compostela, Spain

² Department of Philosophy, University of Santiago de Compostela, Santiago de Compostela, Spain

³ IEGPS CSIC, Santiago de Compostela, Spain

use the terms “ontology” and “conceptual model” as equivalent, following [21]. The model developed as a result of ontological analysis must contain enough information as to allow users to reason about it and then apply conclusions of this reasoning back to the “real world” [22]. The result of ontological analysis is an ontology or conceptual model. Usually, one ontological model is enough to describe the part of the world that is described by a set of texts under a common theme. For example, when analysing a collection of texts about mass migrations, a single ontology should suffice. Only in cases where texts deal with clearly different subjects, multiple ontologies would be advisable.

The goal of argumentation analysis, in turn, is to unpack the structure of the argumentative discourse to determine the speakers’ main claims and describe how they are being justified, supported or attacked by other claims. The result of argumentation analysis is presented as a network of elements loosely based on Argument Interchange Format (AIF) [10] and is called an argumentation model. As opposed to the case of ontologies, one argumentation model is usually necessary for each individual text, because argumentation models describe the specific argumentative structures employed by each speaker in specific situations.

Finally, the goal of agency analysis is to gain insights into the beliefs, desires and intentions of the speakers in order to capture the social and political implications of the discourse and, potentially, intervene in order to mitigate potential injustices or biased controversies [20]. Results of agency analysis are usually expressed in structured natural language, guided by a set of questions that are to be answered from the text being analysed. A single agency model can be developed for a set of texts if these texts involve the same, or similar, agents, and discuss the same things.

Each of these three analysis approaches can be applied independently to a specific piece of text, obtaining a valuable result by itself. However, it is by combining the three together that we obtain the maximum potential of the proposed approach in terms of reproducibility and reliability, as many tasks of one type of analysis become easier and more reliable if we use the results of other types of analysis as input.

Finally, the three analysis approaches are connected together by a fourth sub-domain, context. Context analysis entails identifying and describing the overall situation where a discourse takes place, in terms of the issues being addressed, the themes being discussed, the positions being defended or attacked, and the agents doing all this. As further sections will explain, the integration of ontological, argumentation and agency analysis via context modelling is the major strength of IAT/ML, and something that distinguishes it from any other existing approach to discourse analysis.

When considering the pros and cons of discourse analysis as practised today, we grew interested in knowing how

well existing approaches cover the needs of practitioners (and what gaps exist), how feasible it would be to integrate the three modelling perspectives described above under a single method, how easy would be to develop an accompanying software tools, and how the overall approach would be kept modular and simple enough as to be adapted to varying needs. These concerns crystallised in the form of research sub-questions as described in Sect. 3.

The remainder of this article is organised as follows. Section 2 presents previous work and contextualises our proposal. Section 3 describes the requirements of the new approach as well as the methodology that was employed to develop it, focusing on design science. Section 4 presents the proposed approach in terms of its metamodel. Section 5 presents a case study that illustrates the proposed approach in practice. Section 6 provides details on further validation efforts, including application to different research contexts, teaching and tool implementation. Finally, Section 7 presents some conclusions and future lines of work.

2 Previous work

Numerous frameworks for ontological, argumentation and agency analysis exist in the literature. In every case, these frameworks have been developed without considering possible connections between them. The Integrated Argumentation Model (IAM) [17] is the only approach, to the best of our knowledge, that combines ontological and argumentation analysis to some degree.

Firstly, and in the realm of ontological analysis, a vast body of literature exists on ontology engineering and conceptual modelling. Although these two strands of work come from different historical backgrounds and traditions, more recent works [21, 35] have shown that ontologies and conceptual models are very similar kinds of artefacts and, to most purposes, completely equivalent. Consequently, we will not make big differences between these two research traditions and jointly refer to them as “ontological analysis”. Having said this, we must emphasise that “ontological analysis” in this paper refers to the development of human-oriented conceptual models rather than machine-oriented computational models. Approaches such as OntoUML [58] or ConML [38] are much closer to what IAT/ML needed than, for example, OWL [69] or RDF [70].

Complementary to these approaches, we should mention Named Entity Recognition (NER) from the field of Natural Language Processing (NLP). NER’s main aim is to automatically identify and classify the entities that are mentioned in a piece of text [2]. In principle, this would be valuable to an analyst who is aiming to identify entities by hand. However, most of the current NER techniques are only capable of recognising entities of a very limited range of pre-defined types, such

as places or people's names [44]. However, ontological analysis as part of discourse analysis requires a broader and richer range of entities. Going beyond this limitation in NER is a highly expensive and time-consuming process [47]. Moreover, doing this would involve corpus-dependent training of the NER algorithm, which makes it even less attractive for IAT/ML.

Secondly, argumentation analysis is also a field with a long research tradition. For this proposal, we focus on approaches that emphasise its communicative dimension and have some computational development, such as those of Perelman & Olbrechts-Tyteca [48] or Toulmin [59]. The Interchange Argument Format (AIF) [10] constitutes a milestone in this regard. AIF defines an abstract language for the representation and exchange of argumentation data and aims to be a standard in the argumentation community. Its core ontology defines three main categories of concepts: (i) arguments and argument networks; (ii) communication; and (iii) context. Arguments are represented as directed graphs, where the nodes stand for information contents (such as an utterance or a proposition) or the application of an argumentation pattern or scheme. Communications capture how the production of utterances and dialogue evolves, representing them in terms of protocols and sequence of utterances. Lastly, contexts can capture the non-strictly linguistic elements that play a role in the elaboration of arguments, such as speakers backgrounds or personal commitments.

Several different contemporary theories of argumentation have adopted AIF as their underlying ontology. One of them, Inference Anchoring Theory (IAT) [51], aims to capture how propositional reasoning involved in argumentation is anchored in discourse. IAT has been used by us to define the backbone of argumentation analysis in IAT/ML, and the name "IAT/ML" is a sign of this. Separately, the Periodic Table of Arguments (PTA) [64], which focuses on natural discourse as well, defines a categorisation and a procedure for a systematic analysis and evaluation of arguments based on identifying some internal characteristics in every argument. The Comprehensive Assessment Procedure for Natural Argumentation (CAPNA) [37], another systematic method for argument analysis and evaluation, analyses the semantic content of arguments to offer a reliable evaluation, as well as the argument structure. All these approaches focus exclusively on argumentation analysis.

In the NLP field, argument mining is the research topic that addresses the same goal of argument analysis: the identification, extraction and structural analysis of arguments expressed in natural language [43]. As described in [43], various approaches can be found in the literature, ranging from the most basic, where the main goal is to identify which text fragments are argumentative and what their components are by using argument indicators [56] to the most advanced that try to identify what type of argumentation scheme is being

applied in each case [71]. For the goals of IAT/ML, argument mining techniques are of limited application, as they are not accurate enough to be relied upon with no human supervision.

Thirdly, agency analysis, as defined in this paper, has little representation in the literature, as it constitutes an original development of the authors. However, it is strongly inspired by critical discourse analysis (often abbreviated as CDA) and, more recently, critical discourse studies (CDS) [73], which are indeed well-known approaches in linguistics and for which significant literature exists. CDA can be characterised as an interdisciplinary field, usually involving semiotics, anthropology, psychology, communication studies, and related fields. Gee [20] proposes a framework based on critical questions that address seven aspects: significance, practices, identities, relationships, connections, sign systems and knowledge. Johnstone [41] follows a more linguistic approach, based on Speech Act Theory [3, 4] and Grice's theory of meaning [31]. CDA's main assumption is that meanings strongly depend on speaker's intentions, which are captured by the illocutionary forces and the conversational implicatures (i.e. information that is implied rather than explicitly said), so these become the focus of critical discourse analysis. Agency analysis in IAT/ML takes some aspects of CDA, such as the process of "asking the text" a set of predefined questions, but avoids much of the political and ideological positioning and activism that are usually associated with the latter.

There are relevant works in the literature that do not correspond clearly to any of the three perspectives that we have identified (ontological, argumentation, agency). One example is the gIBIS approach [12], which aims to capture and represent the deliberative process as it occurs, and is oriented to intervene and help during a live process of decision making. IAT/ML, however, is more oriented to a forensic analysis, that is, to describe and analyse the discourse once it has happened rather than intervening while it occurs. Still, gIBIS shares some interesting ideas with IAT/ML, such as the notions of position (a statement or assertion which resolves the issue) and argument (as a support for a position). As a relevant difference, gIBIS focuses on argumentation analysis and does not address ontological or agency analysis, whereas IAT/ML contemplates all of them.

Finally, it is worth mentioning that, despite the conceptual similarities between IAT/ML, especially its agency analysis components, and multi-agent systems (MAS), we have not found much relevant literature in the domain of MAS that could be applied to IAT/ML. This is probably due to the fact that MAS are not centrally discursive in their nature, that is, the interactions between (artificial) agents are not linguistic but formal, i.e. they exchange information without the use of natural human language. Contrarily, IAT/ML is oriented towards human discourse, which involves the nuances,

intricacies and ambiguity that are characteristic of human language but mostly absent in MAS.

3 Development process

In this section, we describe why we decided that a new approach to discourse analysis was necessary, what requirements it had, and what process we followed to create it.

3.1 Motivation and requirements

The current theory and practice of discourse analysis, as described in the previous section, showed that several significant issues remained unsolved:

- Discourse analysis is systematically carried out from a single perspective, with very little or no cross-checking with other perspectives. For example, argumentation analysis is often performed with little or no attention to what the text refers to (ontology) [23], and critical discourse analysis is done with little or no regard to argumentation. This fragmentation of views hinders the robustness of the results and makes them more error prone.
- Critical approaches to discourse analysis are highly subjective, and yield results that are often criticised as not being traceable or replicable because they depend as much on the analyst as on the discourse being analysed [7].
- Most of the critical discourse analysis techniques that are being used lack proper formalisation, documentation and methodological guidance. This means that they are difficult to understand, adopt, integrate and implement in software tools.

These issues motivated us to tackle the development of a new discourse analysis approach with the following purpose: the new approach must allow a trained analyst to obtain a deep and nuanced understanding of a corpus in terms of issues, themes, positions and agents, and by using ontological, argumentation and agency analysis. The new approach is especially oriented towards polarised social issues, although it is applicable to any other domain of discourse.

The major requirements of this new approach were defined as follows:

- A. The new approach must integrate at least three modelling perspectives under a common and inter-connected metamodel: ontological (what the text is about), argumentation (how the speakers justify what they say) and agency (what are the beliefs, desires and intentions of the speakers). This will benefit robustness of results and help cross-validate them from different perspectives.

- B. The new approach must produce results that are traceable to intermediate products (such as ontological, argumentation and agency models) and eventually to the text itself, so that anyone can understand how they have been constructed and why. This will help with replication and explainability, especially across analysts.
- C. The new approach must be documented by a series of conceptual, procedural and technical specifications that can serve as guidelines to people willing to adopt the approach or implement it as part of software tools. This will facilitate adoption and implementability.
- D. The new approach must be as compatible as possible with the major existing approaches, such as IAT for argumentation analysis and ConML and similar for ontology analysis. This will also facilitate adoption.
- E. The new approach must be flexible enough as to be customisable to different projects and settings, allowing users to add or remove components, or integrate it with other techniques. This will also help with implementability and adoption.

The next sub-section briefly describes the users and stakeholders that are targeted by IAT/ML.

3.2 Users and stakeholders

As part of the development process, we liaised with the Outreach Unit at Incipit CSIC (were two of the authors work) plus Oxentia Ltd., a private firm in the UK specialising in knowledge transfer, to develop a map of potential users and stakeholders of IAT/ML. We present a summary of this work here for information on who is expected to use IAT/ML, how, and for which purpose.

The first and obvious stakeholder and user of IAT/ML are **researchers** interested in discourse analysis. Research concerns are related to the empirical production of rigorous knowledge that involve disagreement and conflict. As part of the pursuit of this aim, new tools and methods must be learnt, as well as the development of solid arguments to persuade others of the thesis on which the research is based on. As part of academia ourselves, we understand the difficulty to convince others (academics or the general public) of our conclusions, as well as the time needed and propensity to make mistakes when producing new knowledge in absence of a clear and solid methodology, especially when it is not supported by software tools. The approach presented here helps relieve these issues, since it is based on a rigorous discourse analysis methodology that includes argumentation analysis and is supported by a software tool to develop the analysis. We have developed a training programme in the method and tool, as well as a consultancy programme to assist other academics with their research problems. Please see Sect. 6 for more details on this.

A second stakeholder is **content creators**, which includes journalists, reporters, or even activists involved in spreading knowledge related to complex social issues such as prostitution, racism, or dissonant cultural heritage. Understanding and presenting issues like these is extremely complex when they are highly polarised. In particular, it is complicated to select and integrate reliable sources, to find points of agreements, to reconcile positions and to be objective and avoid injecting a strong personal bias into the report or news piece. Stakeholders of this kind are not expected to become users of IAT/ML or LogosLink themselves, but to become customers of a consultancy service that can provide traceable and rigorous discourse analysis on demand and about the very specific topics that are required.

A third stakeholder is composed of **politicians**, who are hopefully interested in solving social issues. When developing public policies that affect many people it becomes necessary to gather, understand and integrate distinct perspectives from different agents, and manage them with solid motivations. Through the IAT/ML analysis would be possible to produce an action plan including options plus consequences based on topics, positions and agents involved. We also propose training on how to argue better. Like in the previous case, politicians are not expected to be users of IAT/ML or LogosLink. Rather, technical advisors and staff would become customers of a consultancy service that can provide the necessary know-how.

Finally, Oxentia proposed the **police and judiciary** as a fourth kind of stakeholder. There are three areas where our proposal is interesting for this community. Firstly, police and judges have a strong relationship with argumentation in terms of understanding and assessing of the strengths and weaknesses of witnesses, expert reports and other documents in order to arrive to a verdict or conclusion, as well as in finding weak spots in arguments made by opposing lawyers. Secondly, police and judges must develop solid justifications to support their verdicts, decisions and actions. And, thirdly, judges must often assess whether a particular discourse constitutes a crime for defamation, hate speech or incitement to violence. In this sense, our proposal provides help to formulate detailed characterisations of the argumentation used in each discourse and training on how to improve argumentative practices.

We have taken some steps to approximate IAT/ML to these four kinds of stakeholders. Academia and content creators, more precisely fact-checking journalism, are the fields where the methodology is starting to have an impact through workshops and the relations established through different projects where the authors participate; please see Sect. 6.2 for details. Approaching politicians and the police and judiciary has shown to be more difficult, and we are considering additional avenues such as forensic linguistics.

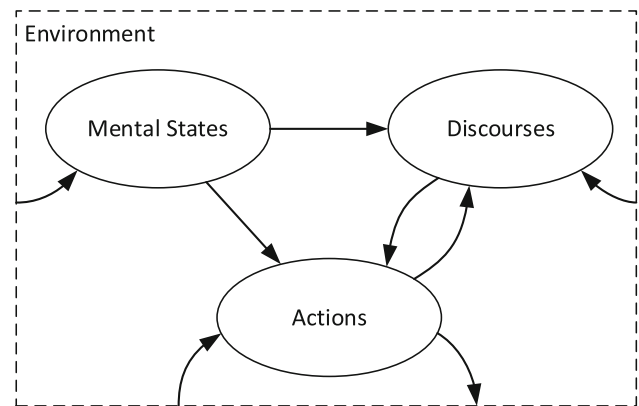


Fig. 1 Environment, mental states, discourses and actions (EMDA), plus the relations between them, constitute the focus of IAT/ML. Arrows represent influences

The next sub-section describes the theoretical background that was adopted for IAT/ML in the light of its requirements and users and stakeholders.

3.3 Theoretical underpinnings

IAT/ML is a methodology for discourse analysis. As such, it deals with discourse, understood as the practical and socially situated use of human language for communicative, pragmatic and symbolic purposes [20]. As a human activity, discourse implies the existence of agents, at least a speaker, who produces the discourse, and a receiver. Agents, in turn, produce at least two additional phenomena: mental states and actions. Mental states include beliefs, emotions, desires, and intentions [6]. Discourses produced by an agent are based on their mental states, that is, we often say what we think, want or plan to do. However, this does not mean that discourses faithfully follow our mental states all the time; in fact, people often speak words that do not match what is in their minds.

With regard to actions, the actions carried out by agents are often compatible with their mental states and discourses but, again, not always, as people sometimes do things that are not aligned with their thoughts or verbal commitments. Reverse connections are also relevant. For example, behaving in certain ways usually compels us to produce certain discourses, and even perhaps to think in certain ways. Mental states, discourses and actions occur within a given environment, composed of social and cultural elements that mediate what we think, say and do. Actions, in turn, modify this environment. In this manner, environment-situated mental states, discourses, and actions (EMDA for short) constitute the focus of IAT/ML, as shown in Fig. 1.

Consider the following example. Imagine a person who believes that everyone should have the same rights and be treated equally regardless of their ethnicity, sex or religion. This person is likely to state these beliefs when asked. However, they may prefer locals as opposed to immigrants when

looking for a carer for their children, perhaps due to mistrust and prejudice fuelled by mass media. This preference may be manifested as a systematic trend to hire only locals despite of the availability of immigrant carers. When confronted with this by peers, this person may rationalise their behaviour by using a discourse that justifies their choices.

In this scenario, mental states, discourses and actions occur in inter-related manners, sometimes aligned, sometimes not. IAT/ML is designed to look at how discourses are aligned or misaligned with mental states, and to what extent actions are aligned or misaligned with mental states and discourses. By “aligned” here we mean that a manifestation is complete and truthful. For example, a fully aligned discourse is one that exposes everything that the agent thinks about something, and nothing that the agent does not think. Similarly, a fully aligned action is one that exercises everything the agent thinks and has said about something, and nothing that the agent does not think or has not said. Of course, alignment is gradual and nuanced, rather than binary.

In this manner, IAT/ML focuses primarily on analysing the discourses, and from them reaches into the mental states and actions of agents. In particular, ontology analysis focuses on representing the mental states (and, especially, the beliefs) that are exposed by discourses. Argumentation analysis, in turn, tries to understand how agents justify their claims and what strategies they use to support or attack other discourses. And, finally, agency analysis aims to represent the relationships between agents’ mental states and their intended actions, together with enough information about the environment so that actions can be contextualised. In this manner, each element in the EMDA framework is addressed by one or many of the modelling approaches (ontology, argumentation and agency), albeit to different degrees.

This inclusive conception of discourse analysis, involving mental states and actions as well as discourse itself, is compatible with the usual descriptions of discourse analysis in terms of “saying, doing and being” [20], where “saying” refers to the discourses, “doing” to the actions, and “being” to the mental states of the agents.

Each of the three analysis approaches (ontology, argumentation and agency) is supported by some theoretical commitments, which are briefly described in the remainder of this section.

3.3.1 Ontologies

IAT/ML uses the word “ontology” as synonymous with “conceptual model”, following [21]. In this sense, an ontology is a model of a section of the world that focuses on the things that make it up, their properties and relationships. We use the word “thing” as it is usually employed in philosophy, i.e. to refer to discrete segments of the world that we can perceive and “pick out”. We adopt the classical view that some things

in the world are types and some are tokens [66]; types correspond to categories or classes, whereas tokens correspond to instances or individual objects that can be assigned one or more types. Types, in turn, can be organised in subtyping or subsumption hierarchies.

However, and as opposed to many mainstream modelling languages, we adopt in IAT/ML a multi-level modelling approach [11, 54] by which types are subtypes of tokens. In other words, a type in IAT/ML can be an instance of another type, and so on and so forth. This allows for more expressive and natural ontologies in situations where powertypes [28] are being dealt with, e.g. when describing an animal species such as Dog as both a token (an instance of AnimalSpecies) and a category (a subtype of Animal).

In addition, types can be described through features, which include properties (which are qualities or quantities) and connections (which point to other types). In parallel, tokens can be described through facets, which include values (instances of properties) and references (instances of connections). Please see [22] for a detailed description of this.

3.3.2 Argumentation

Argumentation in IAT/ML is strongly based on argumentation theory and, in particular, on the Argumentation Interchange Format (AIF) [10], which, despite its name, is an abstract model of argumentation as well as a data exchange format. For AIF, argumentation can be represented as a collection of propositions (i.e. statements by one or more speakers) connected by argumentation relationships. These relationships include inferences, which indicate that one or more propositions support another one; conflicts, which indicate that a proposition is incompatible with another one; and rephrases, which connect a proposition to another one that is being recast or reinterpreted in some way.

To this basic framework, IAT/ML adds the innovative notion of ontological proxies [23]. These are simplified representations of ontology elements inside an argumentation model, which work as anchors for denotations, that is, the semantic connections between a term in a locution and the concept represented by an element in the corresponding ontology. As we explain in Sect. 4.3, ontological proxies allow the formal and practical connection between the ontological and argumentation domains, which are often treated separately in the literature and in the practice of discourse analysis.

3.3.3 Agency

In IAT/ML, an agent is an entity that has mental states (at least beliefs, desires and intentions) plus the ability to act according to them. Consequently, IAT/ML adopts the beliefs/desires/intentions model popularised by Bratman [6]

and widely adopted in multi-agent systems (MAS) engineering [5].

In addition, agency analysis in IAT/ML borrows some aspects from critical discourse analysis (CDA) [15, 73], as indicated in the Introduction. However, and due to the very post-modern stance of much of the literature on CDA, it is extremely difficult to provide a formalised account of what constitutes the core concepts of this approach [7]. IAT/ML adopts the notion that a set of questions are asked against the text being analysed, and responses to these questions, elaborated by the analyst, constitute the results of the analysis.

The next sub-section describes how the requirements described above were used to develop IAT/ML based on these theoretical underpinnings and for the identified users and stakeholders.

3.4 Design science

Once the requirements and the theory were clear, our starting point was that a domain-specific approach was to be built to support the entire process of discourse analysis, including ontological, argumentation and agency modelling. In working through the methodology, the issue appeared of whether IAT/ML constituted a domain-specific modelling language (DSML) or not. This issue is not central to this paper but, from a methodological point of view, we considered that it would have been a mistake to not consider guidance on DSML development for the creation of IAT/ML.

We started off with the abstract research question of whether it is possible to develop a domain-specific approach to support the process of discourse analysis in an integral fashion, reconciling ontological, argumentation and some aspects of critical analysis (later renamed “agency analysis”). In order to address this, some research sub-questions (RSQ) were raised:

1. (a) What concepts and patterns are found in the discourse analysis process and its domain that are common to existing approaches? and (b) Which ones we do not find in existing approaches, but are necessary after our experience analysing discourses? (Requirement D)
2. Can we develop a domain-specific modelling language that fully describes and supports both the discourse analysis domain as well as the process for the three perspectives (ontological, argumentation and agency)? (Requirement A)
3. Is it viable to implement this language in a modelling tool? (Requirement C)
4. What degree of coverage and traceability does this language offer for the needs of discourse analysis? (Requirement B)

5. How viable, and at what cost, is to customise the language in order to cater for specific situations and project needs? (Requirement E)

Following Design Science [36], our main research question (and the subsequent sub-questions) implies the construction of an artefact (namely, a DSML, or something like it). Technical Action Research (TAR, [67]) has been used to answer research questions through artefact construction in other domains such as smart cities or health [42, 53], so we adopted it, as it would allow us to integrate the construction of the DSML into ongoing research projects. In this sense, the initial versions of IAT/ML constituted an artefact at the service of practitioners in ongoing discourse analysis projects for experimentation and improvement. These projects, due to their very interdisciplinary nature (cultural heritage, feminist identities, and communication of the COVID-19 pandemic) brought together professionals of different backgrounds, including linguists, cultural heritage specialists, philosophers of language and software engineers, which broadened and generalised the domain of application of the approach and allowed its validation by different stakeholders in later phases. By applying TAR, it was possible to integrate the process of developing IAT/ML into our own research process, thus being able to respond to the SRQs listed above.

There is no single or optimal methodology for DSML development, and approaches vary enormously, ranging from ad hoc proposals [13, 72], to those based on patterns [49], ontologies [32], or more oriented towards meeting the requirements of practitioners [19]. In our case, the empirical and incremental nature of the development of IAT/ML motivated the choice of an approach focussed on meeting the needs of practitioners, in our case team members who were performing discourse analysis tasks. Taking Frank’s method [19] as a reference, Fig. 2 illustrates the process followed for the design and development of IAT/ML.

Phase 1: clarification of scope and purpose. In this phase, we defined the scope of IAT/ML [29] by working with language and discourse experts working in the above-mentioned ongoing projects. We performed some gap analysis in relation to existing discourse analysis techniques, mostly IAT [8].

Phase 2: analysis of generic requirements. In this phase, we developed business-level requirements and a draft list of concepts that would be necessary to support discourse analysis from the described three perspectives.

Phase 3: analysis of specific requirements. In this phase, we developed a sketch metamodel for IAT/ML as well as a list of functional requirements that should be supported by IAT/ML, such as computation of argumentation statistics or argument structure analysis.

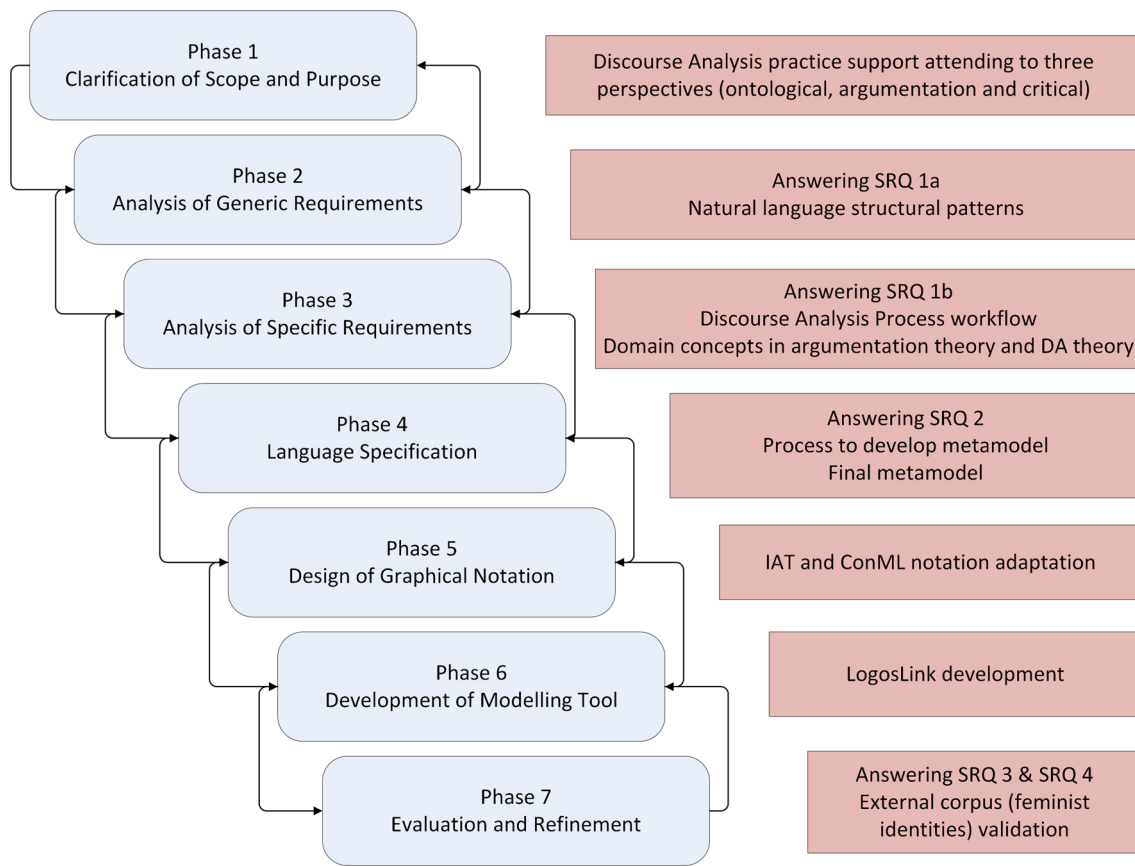


Fig. 2 IAT/ML design process, based on Frank's method. Grey shapes depict Frank's process, whereas red rectangles show IAT/ML development phases and specific results for each one

Phase 4: language specification. In this phase, we defined the IAT/ML metamodel as described in Sect. 4 of this paper. We also cross-validated the concepts, relationships and implications of the metamodel with project team members in terms of the required linguistic and discursive concepts. For some of the underpinning concepts, such as that of ontological proxy [23], we needed to go back to phase 3 and revisit existing requirements for their refinement and adjustment.

Phase 5: design of graphical notation. In this phase, we sketched some ideas for a graphical notation, mostly inspired by IAT [8] and ConML [38], and validated them with project team members for usability and acceptance.

Phase 6: development of modelling tool. In this phase, we initiated the development of LogosLink, a software toolset that implements most of IAT/ML. This required achieving a moderately stable metamodel so that software development could proceed on acceptably solid grounds. Section 6.1 briefly describes LogosLink.

Phase 7: evaluation and refinement. The evaluation of a DSML and a corresponding modelling tool recommends checking them against the requirements building on the use scenarios created for requirements analysis. As Frank's method specifies, each use case serves to analyse whether and

how corresponding requirements are satisfied by the DSML. In our case, the stable IAT/ML version was validated with respect to the initial requirements by the team members of the project for which IAT/ML was being experimentally used, as an initial way of verifying whether the general and specific requirements were met. Luckily, a project revolving around a new discourse analysis theme, that of feminist identities, was launched at that time, which allowed us to validate IAT/ML and LogosLink with a corpus and in relation to a topic that was radically different to those used during its initial development.

SRQ5 was addressed during the application of the language, as described in Sect. 6. The IAT/ML metamodel is presented in the next section.

4 Results

IAT/ML is inspired by the IAT argumentation analysis approach [8, 40] as well as the ConML conceptual modelling language [24, 38]. As explained before, IAT/ML covers

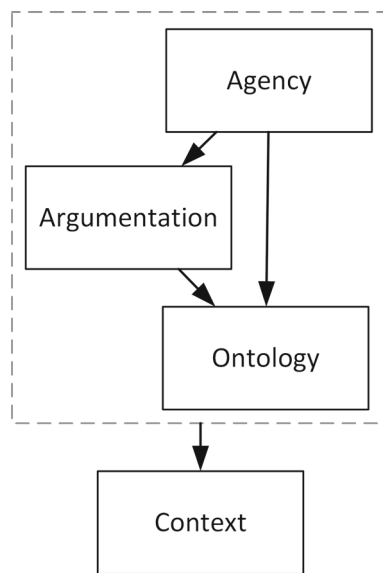


Fig. 3 Metamodel architecture of IAT/ML. Boxes represent components. Arrows represent dependencies

ontological, argumentation and agency analysis, plus context analysis as an inter-connecting infrastructure. Thus, the IAT/ML metamodel is composed of four components:

- *Context* which contains elements related to the overall context where the discourse takes place, including the themes it describes, the different positions being discussed, and the agents supporting or attacking each position.
- *Ontology* which contains elements related to the ontology being referred to by the discourse, including elements such as categories, properties, associations, atoms, values and links.
- *Argumentation* which contains elements related to the argumentative structure of the discourse, including the locutions and transitions uttered by speakers, the resulting propositions and argumentation relations such as inferences, conflicts and rephrases, and the connecting illocutionary forces. Argumentation elements are connected to ontology elements via denotations.
- *Agency* which contains elements related to the beliefs, desires and intentions of speakers, including entities that they mention, questions and their responses.

These components are organised in a metamodel as shown in Fig. 3.

This metamodel is used to provide a formalised view of the domain of study, namely human discourse. Furthermore, the metamodel is used to provide guidance to stakeholders on how to carry out discourse analysis, in the form of a methodology. And, thirdly, the metamodel is used as the structural backbone of the LogosLink software tool, described in further sections.

Using the metamodel to carry out discourse analysis involves both a manual (i.e. human-based) plus an automated phase. Instantiating the metamodel to construct context, ontology, argumentation and agency models is a time-consuming process that must be carried out by experienced human analysts. Once these models are available, then computerised tools can process them to obtain a wide range of analytical results (such as semantic collocations, argumentation structure or agent centrality) that would be impossible to obtain from the source texts alone with today's technologies.

The following sections provide additional details on each component of the metamodel. Diagrams are expressed in ConML [24, 38].

4.1 Context

This component contains metamodel elements related to the context in which the discourse takes place. Figure 4 depicts the major metamodel elements.

Contexts are described by using elements of four related types. An issue is a socially relevant problem or situation that is going to be addressed, such as “How can we guarantee safe food and water in developing countries?” or “What are the main drivers of present mass migrations?”. A Theme is a domain of discourse about which things can be said. Example themes are “International Politics” or “Feminism”. Themes can be nested within themes to cater for more general and more specific domains of discourse.

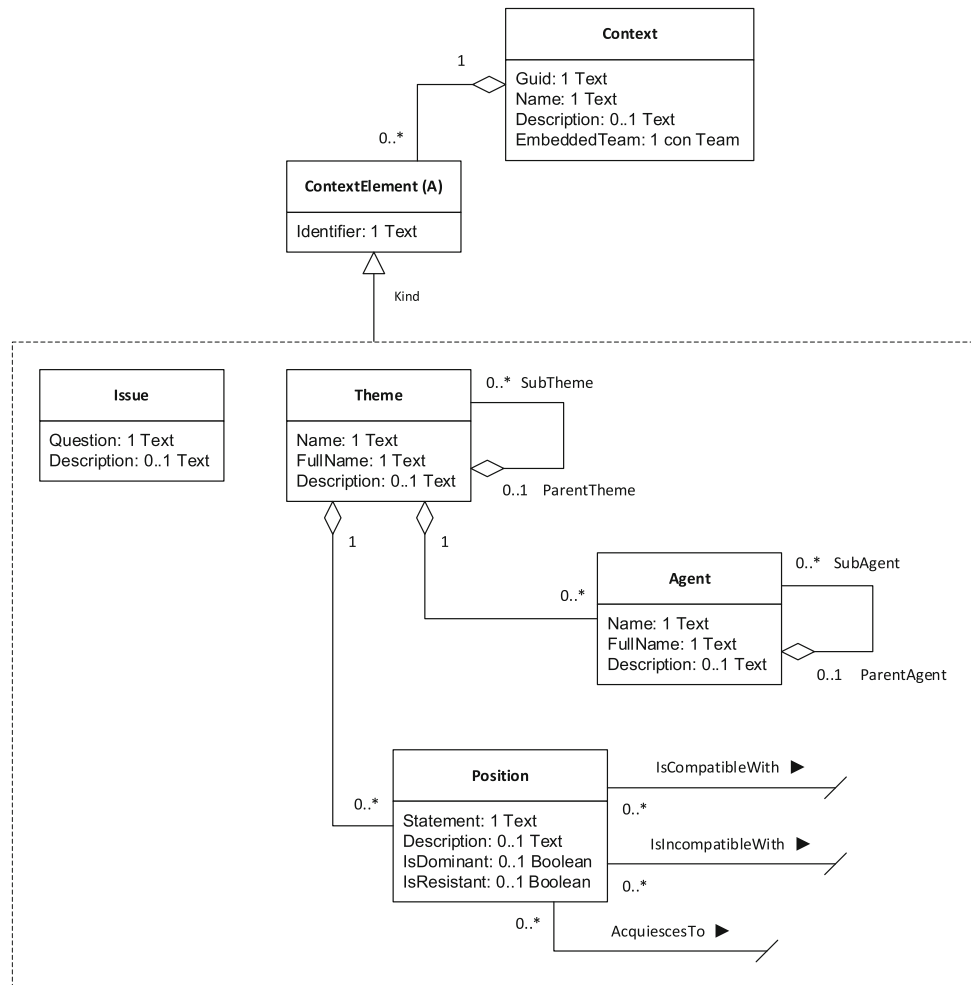
Each theme may involve positions and agents. A Position is a statement that expresses a belief that is defended by some and attacked by some others. For example, the “International Politics” theme may contain the position “The only solution for the Israeli–Palestinian conflict is a two-state scenario”. An Agent, in turn, is a person or group that defends, attacks or is otherwise involved in a position. Sample agents are “Immigrant women” or “Nelson Mandela”. Agents can also be nested within agents to express subgroups of people.

Together, issues, themes, positions and agents characterise the context where a discourse takes place. Elements in ontologies, argumentation models and agency models (see next sections) can be linked to elements in the context for grounding and cross-model connection.

4.2 Ontology

This component contains metamodel elements related to the ontology referred to by the speakers. Ontologies in IAT/ML are multi-level, in the sense that multiple levels of instantiation are possible [1, 11]. Also, ontologies in IAT/ML borrow from ConML [38], which was chosen for its explicit support for temporality and subjectivity modelling, which are crucial when representing discourse [25]. As opposed to other modelling languages such as UML, ConML can represent the

Fig. 4 Metamodel elements in the Context component



subjective perspectives held by each agent on each object or class and refer to them over time. Figure 5 depicts the major metamodel elements.

There are three kinds of ontology elements: entities, features and facets. An Entity is an ontology element that represents an identity-bearing thing in the world. Note that entities can be linked to themes, positions and agents in the underlying context, as entities in an ontology can represent either of these. There are two kinds of entities: atoms and categories. An Atom is an entity that represents a non-instantiable thing in the world. Atoms correspond to *urelements* in set theory, or individuals in philosophy. A Category, in turn, is an entity that represents a class of things in the world. Categories correspond to sets in set theory or universals in philosophy. Categories work as types in relation to entities, so that entities (either atoms or categories) can be instances of categories. Also, categories can be arranged in subtyping hierarchies, and multiple inheritance of features and facets is supported.

A Feature is an ontology element that represents a type of predication on entities of a given category, that is, some

shared property of all instances of a category. There are two kinds: properties and connections. A Property is a feature corresponding to quantities or qualities of the entities of the category, such as “Height” for the “Person” category. This is very similar to the concept of “attribute” in other modelling languages such as UML or ConML. A Connection, in turn, is a feature corresponding to relationships of entities of the category to entities of another category, such as “IsLocatedIn” for “City” towards “Country”. Connections are directional and are paired up to constitute bidirectional Associations.

Features work as types of facets. A Facet is an ontology element that represents a predication on an entity in the world, regardless of whether it is a category or an atom. There are two kinds of facets, corresponding to the two kinds of features: values and references. A Value is a facet corresponding to a quantity or quality of an entity, such as “Alice.Height = 171”. A Reference, in turn, is a facet corresponding to a relationship of an entity to another entity, such as “Rome.IsLocatedIn = Italy”. References, like connections, are unidirectional, so they are paired up to constitute bidirectional Links.

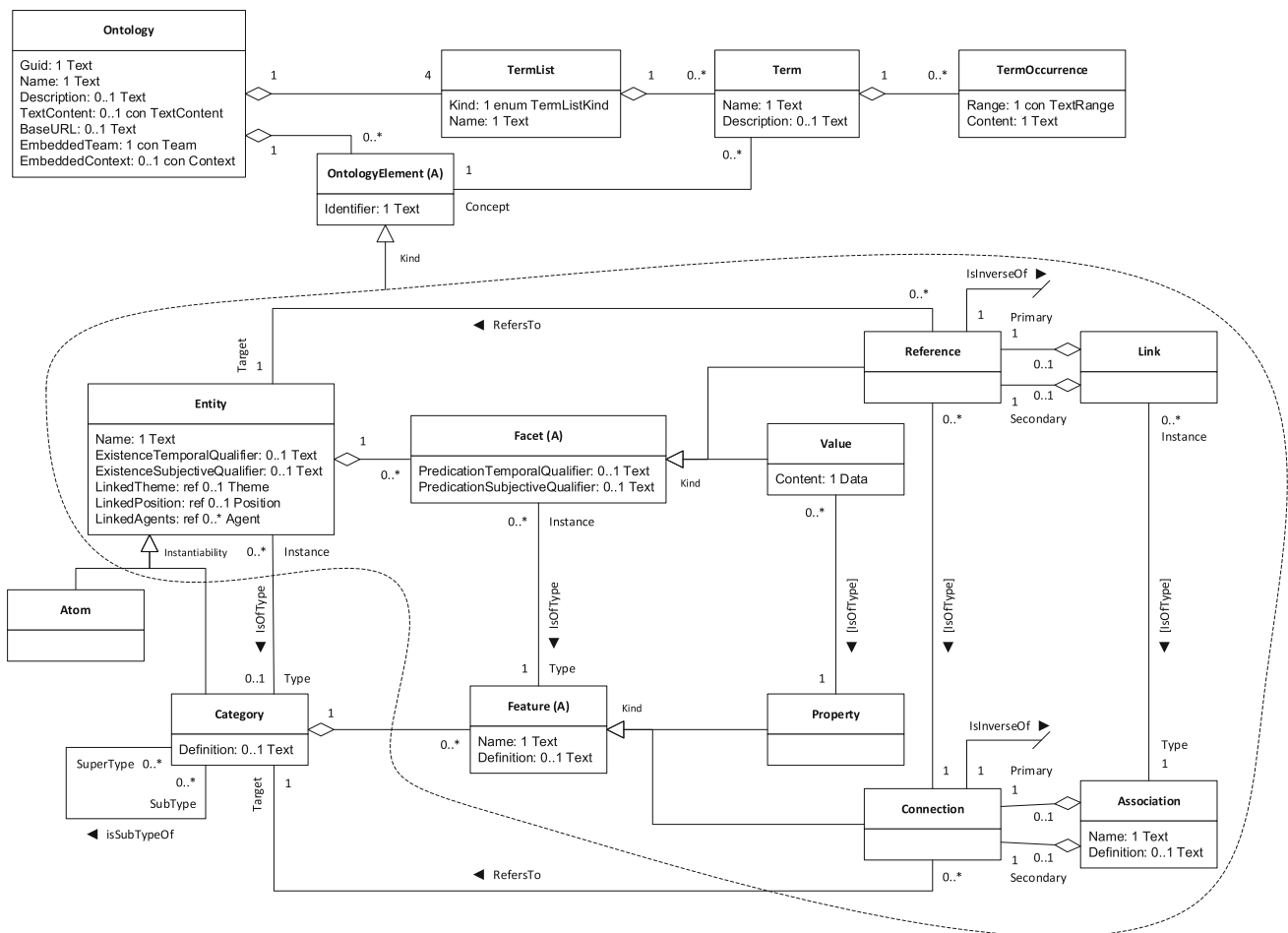


Fig. 5 Metamodel elements in the Ontology component

Drawing from ConML, IAT/ML ontology modelling provides support to describe existential and predication subjectivity and temporality [25], so that a modeller can record in a model when something is the case (when an entity exists, or when it has a particular property), or according to whom (according to whom it exists or has a particular property). The case study presented in the next section provides more details and illustrations of this.

4.3 Argumentation

This component contains metamodel elements related to the literal discourse as spoken by speakers, plus the argumentation elements and relationships employed by them. Many of these elements have been taken from IAT [8, 9], which has been thoroughly applied in practice and validated [52, 57]. Figure 6 depicts the major metamodel elements.

A Speaker in an argumentation model is an individual or group who participates in a discourse by speaking locutions and issuing propositions. Speakers can be linked to agents in the underlying context, to capture the fact that a speaker

can pertain to one (or even multiple) agents. For example, “Barak Obama” could be linked to agents “US Politicians” and also “African Americans”. Two kinds of elements may exist to describe what a speaker says: locutions and transitions. A Locution is an utterance made by a speaker in the discourse, whereas a Transition is a discursive relationship between locutions. Transitions show discursive dependencies and do not necessarily correspond to the chronological order of the discourse (which is given by timestamps of locutions), but must be compatible with it. Transitions provide the links that help the interpretation of a locution in relation to immediately related ones. For example, transitions of the Adding type indicate that a speaker adds something to what they said before; transitions of the TurnTaking type indicate that a speaker is talking after another speaker has finished. By combining locutions and transitions, we can represent a discourse as a sequence of utterances connected in a linear fashion, with the occasional branch for embeddings (such as in appositions, e.g. “My sister, who lives in France, will be arriving tomorrow”) or reportings (e.g. “Clinton said yesterday that she is not worried about the escalating tensions”).

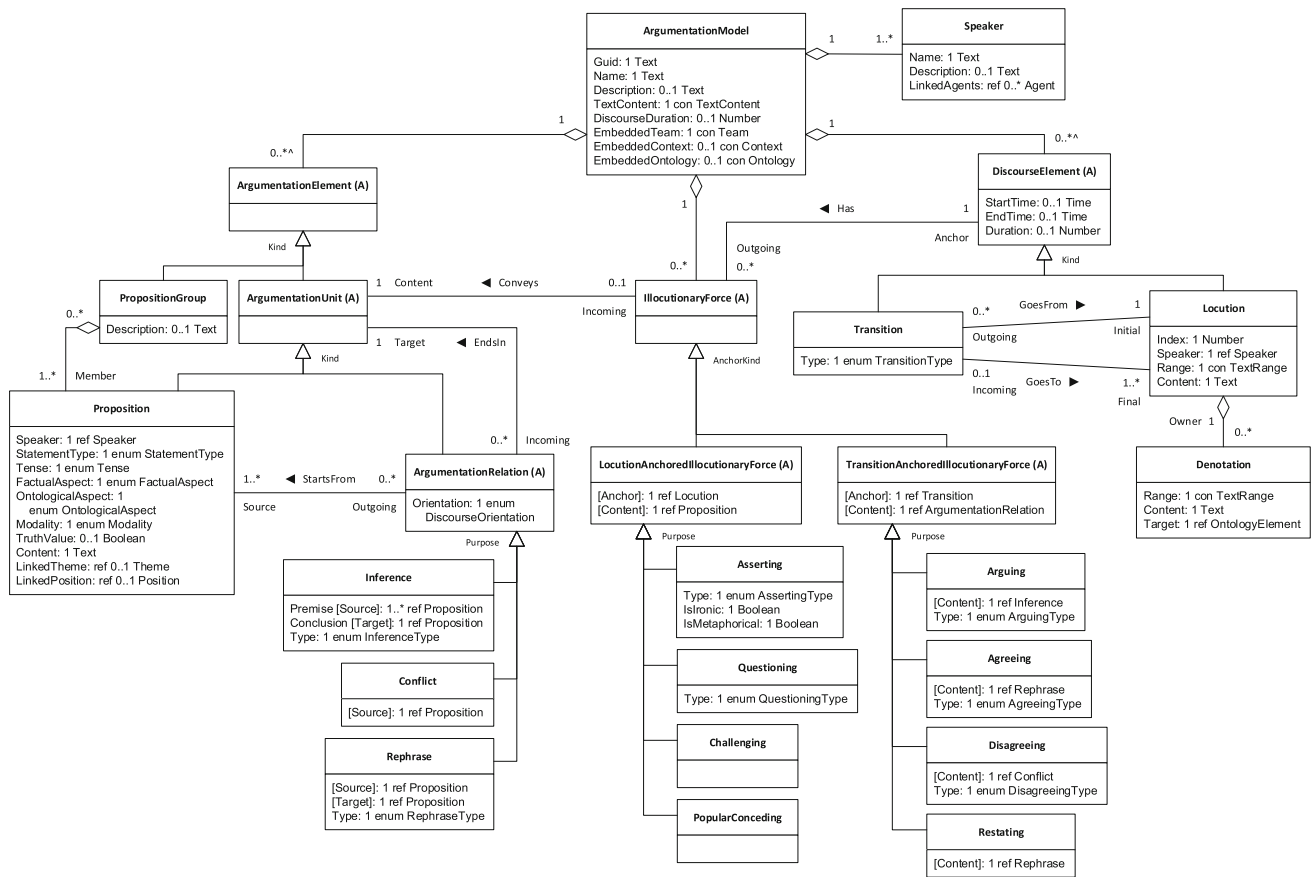


Fig. 6 Metamodel elements in the Argumentation component

In addition, two extra kinds of elements may exist in relation to the argumentation itself: propositions and argumentation relations. A Proposition is an argumentation unit corresponding to a state of affairs about the world. Propositions are self-contained and do not include unresolved references (such as anaphoric or deictic elements), so that their truth value is stable and as independent of the context as possible. Propositions can be characterised in a number of ways via attributes such as StatementType (fact or value), FactualAspect (existence, identity, predication, etc.), OntologicalAspect (logically necessary, physically possible, socially contingent, etc.), Modality (indicative, definitional, noetic, commissive, suggestive, etc.) and Tense (past, present, future or atemporal). Proposition characterisation via these properties is very relevant in situations where different stances on the world are kept by different speakers, allowing the analyst to represent each voice separately. Propositions can be also linked to themes and positions in the underlying context, to capture the fact that some propositions are about certain themes and support certain positions.

An Argumentation Relation, on the other hand, is an argumentation unit corresponding to a connection between two or

more argumentation units so that some of them are argumentally dependent on others. There are three kinds: inferences, conflicts and rephrasings. An Inference is an argumentation relation that indicates that one or more premise propositions are provided by a speaker to support a conclusion proposition. All the involved premise propositions are implicitly connected via conjunction. Inferences can be characterised through a Type attribute, which is based on the subtypes proposed by [61, 65]. A Conflict is an argumentation relation that indicates that a source proposition provided by a speaker is in any kind of conflict with a target proposition. Finally, a Rephrase is an argumentation relation that indicates that a source proposition is provided by a speaker as a reformulation of a target proposition. Rephrases can be of multiple types, such as Abstraction (i.e. the speaker repeats the target proposition but raising the level of abstraction), Agreement (i.e. the speaker expresses agreement with the target proposition), or Reinterpretation (the speaker reinterprets the target proposition by changing its contents without frontally contradicting it, including mechanisms such as analogies, adding emotional nuance, straw man fallacies, etc.).

To connect argumentation to discourse, argumentation models may contain illocutionary forces of different kinds.

An Illocutionary Force is a connection between a discourse element and an argumentation unit in terms of speaker intent. They are taken from the ample literature on speech acts such as [4, 55]. There are different kinds of illocutionary forces; some of them are always anchored on locutions, whereas others are anchored on transitions. Regarding locution-anchored illocutionary forces, an Asserting is an illocutionary force indicating that the speaker wants to communicate what they believe, as in e.g. “Today is a beautiful day”. A Questioning is an illocutionary force indicating that the speaker wants to obtain new information, as in e.g. “What’s your name?”. A Challenging is an illocutionary force by which a speaker requests another speaker to produce a new proposition that works as a premise for a base proposition, as in e.g. Alice: “Today is a beautiful day”; Bob: “How so?”; here, Bob is asking Alice to say something that justifies why she said that today is a beautiful day. Finally, a Popular Conceding is an illocutionary force indicating that the speaker wants to communicate that they believe a well-known and commonly accepted content proposition, as in e.g. “Everybody knows that the Earth is round”.

Regarding transition-anchored illocutionary forces, an Arguing is an illocutionary force indicating that the speaker produces an anchor transition to support a content inference, as in e.g. “Today is a beautiful day because it’s sunny”. An Agreeing is an illocutionary force indicating that the speaker produces an anchor transition to react affirmatively to a base proposition through a content rephrase, as in e.g. “Yes, of course”. Contrarily, a Disagreeing is an illocutionary force indicating that the speaker produces an anchor transition to react negatively to a base proposition through a content conflict, as in e.g. “No way!”. Finally, a Restating is an illocutionary force indicating that the speaker produces an anchor transition to recast a base proposition through a content rephrase, as in e.g. “Most large cities are heavily polluted. In particular, Beijing’s concentrations of nitrogen dioxide and PM10 concentrations are well above national standards”; where the second sentence is rephrasing the first by providing a particular example.

Locutions can be connected to ontology elements via denotations. A Denotation is a semiotic connection between a segment of a locution and a target ontology element. Denotations are based on the concept of ontological proxies [23, 26], which work to connect the argumentation and ontological aspects of discourse modelling in a single mesh of relationships so that semantics can be captured and managed.

4.4 Agency

This component contains metamodel elements related to the beliefs, desires and intentions (BDI) of the speakers in the discourse. The BDI framework is well-known in the literature on intelligent agents [6, 50] and was adopted for IAT/ML for

its strong support of agents’ mental states and plans. Figure 7 depicts the major metamodel elements for agency modelling.

Agency models are collections of responses to predefined questions. Questions, in turn, are organised into question sets. A Question Set is a collection of questions, optionally arranged in Question Groups. Questions may tackle the beliefs, desires and/or intentions of the speakers in the text and therefore are equipped with some specific guidance as to how each contributes to each of the BDI dimensions. Some example questions may be “What are the most repeating ideas in the text?”, “What role is played by each agent according to each speaker?” or “What strategies are used by each speaker to defend their main thesis?”. Some questions may refer to one or more Entity Lists, such as “What strategies are used by each speaker to defend their main thesis?”, which refers to the list of speakers in the text. There are different kinds of questions: short text, which are responded via a brief free text, option list, which are responded by selecting options from a predefined list, and itemised, which are responded by freely listing individual items.

Responses, in turn, may refer to speakers, and speakers, as in the case of argumentation, can be linked to agents in the underlying context. Responses may also refer to entities in the associated question’s entity lists, and must be of the same subtype (short-text, option list or itemised) as their associated question. Once responses have been developed for a speaker, the speaker’s beliefs, desires and intentions can be characterised from the gathered information.

5 Case study

Recent research shows that disinformation constitutes an extremely important problem in our society, revealing, for instance, that fake news spread more broadly, deeper, and faster than truth [63], and that social network posts containing disinformation are 70% more likely to be shared than truthful posts [18]. This situation constitutes a threat especially for European democracies, because disinformation often supports radicalism, opinion manipulation, and extremism against minorities and vulnerable populations.

The verification of data as a professional task is usually carried out by media professionals and journalists who must, very often manually, verify the different sources of information about a specific fact, as well as the dissemination that public figures do about it [62]. As a result, fact-checkers arrive at a certain conclusion about the degree of truth of the fact. This conclusion can be binary (true or false) or gradual. Several works have highlighted the heterogeneity of sources and processes and the time burden that fact-checking implies [14, 33], and have tried to assist this process through modelling and software techniques. Also, large companies

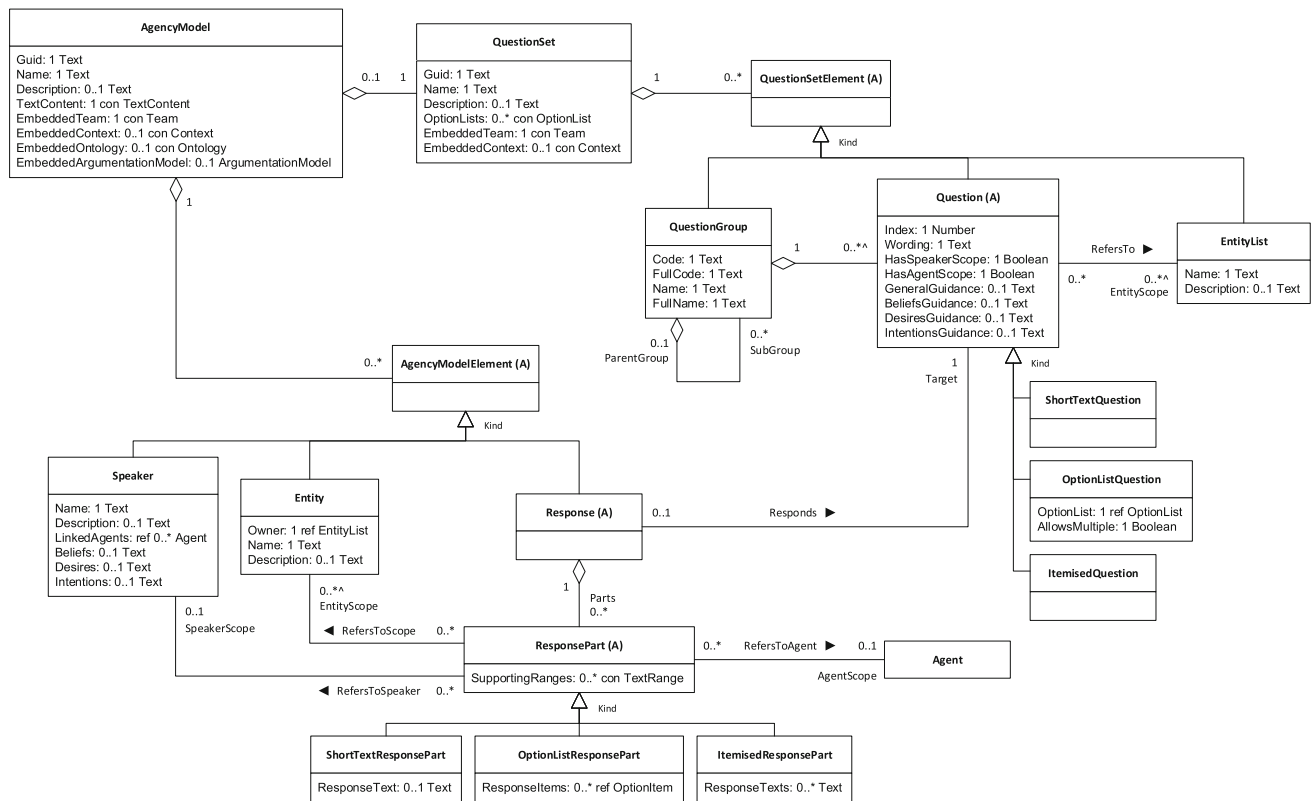


Fig. 7 Metamodel elements in the Agency component

like Google have technological suites [30] that allow fact-checkers to manage and tag their sources for more efficient work organisation.

Even with these tools, fact-checking always involves intense work on the different discourses about the target fact, which must be carefully analysed in order to obtain a result and, above all, to determine the reasons why the result is what it is [60]. It is this need to justify the verdict in any fact-checking process that makes fact-checking an excellent domain for a case study to validate IAT/ML.

Language, being so intuitive to humans, can be deceptive in what it can hide. For example, the truth or falsehood of some discourses may look easily decidable by any critical reader, and discourse analysis may look like a cumbersome unnecessary way to restate what a careful reading would already show. However, having a strong intuition about the facts presented in a discourse is one thing, but being able to demonstrate a solid backing for this intuition is a very different matter. Using IAT/ML (or any other discourse analysis approach) is not about affirming what we already believe, but being able to provide support when challenged, and show others why we believe what we believe. The following case study illustrates how to accomplish this.

5.1 Presentation

In September 2023, some news appeared describing the expense of over one million US dollars in a Cartier store by Olena Zelenska, wife of Ukraine's president Volodymyr Zelenskyy. According to the news release, the shopping episode occurred on 22 September 2023 during Volodymyr Zelenskyy and Olena Zelenska's visit to New York to participate in the UN General Assembly. A former Cartier employee, who was able to obtain a copy of the purchase receipt, was involved. According to the international press, Volodymyr Zelenskyy and his wife landed in Ottawa on 22 September 2023, so it was unlikely that the purchase occurred on that date. The news, which originated from a Nigerian newspaper and was echoed by the Russian media, was later denied by several fact-checking agencies, including Newtral (www.newtral.es) in Spain.

The selection of this case study responds to the fact that, as a verified piece of fake news, we were able to study the original source and how it was echoed and spread by mass media. Analysing discourses like this becomes especially important when the subject matter affects a highly polarised issue such as the war in Ukraine. At the same time, this is a convenient example to explore counterarguing, namely the work done by a fact-checking agency to refute the false claim.

5.2 Analysis

To address the claim described above, we composed a corpus of five documents: two pieces reporting the claim, two of the sources used as evidence to deny it, and a final verdict from a fact-checking agency, Newtral. The two pieces reporting the fake news include the original source from *The Nation* Newspaper in Nigeria as well as one from *Rossiyskaya Gazeta* in Russia, who echoed the false claims. Both pieces support their view on the basis of declarations made by a former Fifth Avenue Cartier store employee who, as reported, was fired after Olena Zelenska spent over one million dollars at the store. Regarding the sources used as denying evidence, they are from *The New York Times* in the USA and *CTV News* in Canada. The first is a piece of conventional news explaining Zelensky's visit to Canada and his address to the Canadian Parliament and provides some context about Canadian–Ukrainian relations. The second is a summary of the visit.

5.2.1 Context analysis

The issue being addressed by this case study could be worded as “How is misinformation being used to smear Ukrainian politicians?”. The theme is the conflict in Ukraine and, more specifically, the questioning of the morality and respectability of Ukraine's president and his wife. If successfully discredited, this would most likely affect Zelensky's ability to secure additional funding for the war from the USA and Canada, and increase polarisation and social division about the conflict due to the ability of media to echo and amplify the message [16]. Within this theme, there are two incompatible positions in this corpus: that Olena Zelenska made a millionaire purchase at Cartier in New York, and that she did not. Relevant agents include at least Olena Zelenska, Volodymyr Zelensky, and the Ukrainian and Russian governments.

5.2.2 Ontology analysis

Ontology analysis allows us to portray the main entities mentioned by the different texts in the corpus, together with their properties and relationships. In addition, we can add subjectivity and temporality markers to some of these entities. In the case of polarised discourses, properly managing subjectivity becomes a crucial aspect of modelling, because different model elements may exist or have specific properties only according to certain agents. Adding subjective markers to model elements, in turn, makes it possible to observe the conflicts in different discourses about the same things, what entities are affected by it, and in what sense. As described in previous sections, ontological modelling in IAT/ML is based on ConML, which provides explicit support for temporality, subjectivity and other “soft” issues as part of its metamodel.

To represent the major ideas in the theme being analysed, we developed an ontology containing some types (Fig. 8) and instances (Fig. 9).

Four categories were considered necessary: Place, since the opposing positions revolve around where Zelenska was on 22 September; Agent, since Zelenska and her husband are the focus of the debate; Event, to represent the purported millionaire purchase; and Perspective, to represent the different perspectives about the issue given by different media outlets. A capital “T” in parenthesis in the diagram indicates a temporal feature, whereas a capital “S” indicates a subjective one, i.e. one that may vary depending on who is speaking.

In this manner, instances were added to represent each of the relevant entities, as shown in Fig. 9: agents corresponding to Zelenska, her husband, and the Cartier's employee who revealed the news to the media, the places where they were supposed to be on 22 September according to different agents, the presumed expensive purchase event, and the different perspectives on it. The two perspectives in the ontology map to the two positions in the context model (see Sect. 5.2.1).

Note that time markers (indicated by an “@” sign) appear at several points in the ontology, because they are crucial, in this particular case, to determine whether Zelenska was in New York or not on that day. Subjective markers (indicated by a “\$” sign) also play an important role; for example, the expensive event exists only according to Cartier's employee, and it is only this person who places Zelenska at the 5th Avenue Cartier Store on 22 September. Similarly, it is various eyewitnesses who place both Zelenska and her husband in Canada on that date.

5.2.3 Argumentation analysis

An argumentation model was developed for each of the five documents in the corpus, adding up to 5,476 words and 123 individual propositions. Each proposition in the model is connected to the speaker who uttered the associated locution; this allows us to map speakers to agents in the context model. In addition, propositions themselves can be mapped to positions in the context model (see Sect. 5.2.1).

From an argumentation point of view, the analysis of the two texts containing what we now know are fake news shows a significant number of unsupported statements, which are then employed to build more elaborate conclusions on Zelenska's character (Fig. 10).

Here, proposition PR103 is mapped to position “Olena Zelenska made a millionaire purchase at Cartier in New York” in the context model. In addition, it is important to remark that statements such as “The wife of the Ukrainian president has spent \$1,100,000 on jewellery” or “Olena Zelenska had an aggressive behaviour” are unsupported, that is, they are stated with no backing arguments. Also, note

Fig. 8 Types in the case study ontology. The diagram is expressed in ConML

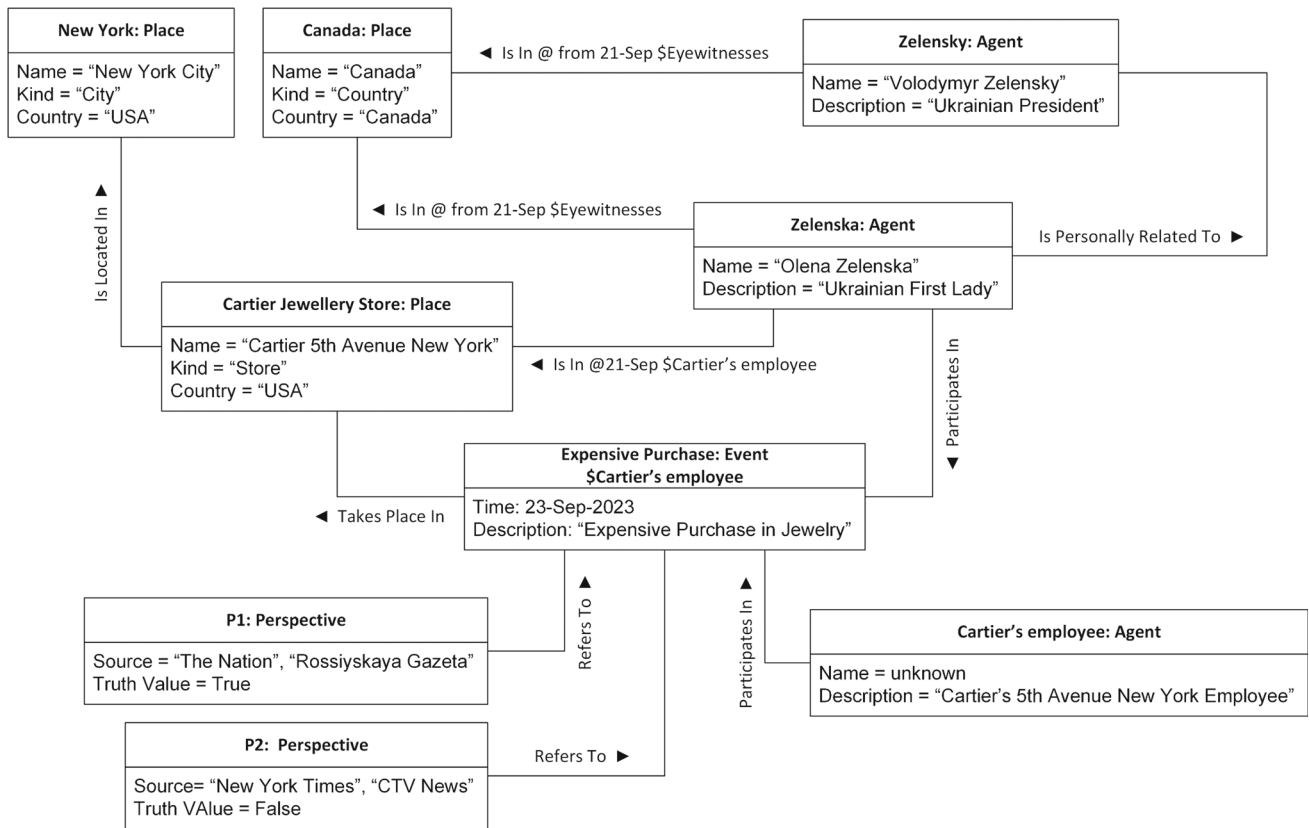
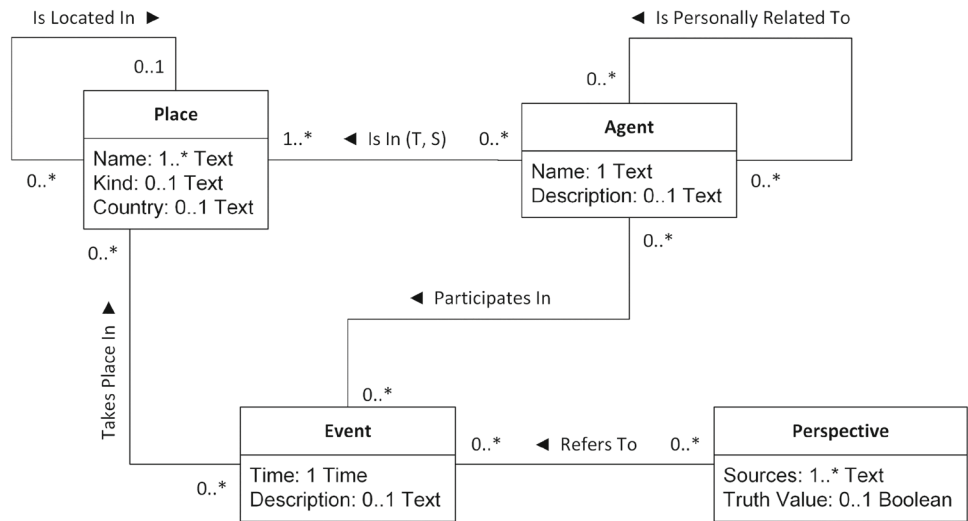


Fig. 9 Instances in the case study ontology. The diagram is expressed in ConML

that the unsupported proposition is, in fact, the main thesis being defended. By combining these unsupported propositions with additional claims, the speakers conclude that “It seems that her appetite has grown dramatically as the time passed” and “Olena Zelenska’s shopping habits went public again”. In addition, note the subjectivity being injected in appreciations like “It seems that her appetite has grown dramatically”.

In contrast, the fact-checkers’ report (Fig. 11) uses a combination of eyewitnesses’ reports and more objective information about the assumed receipt to argue against the position by using a convergent structure that supports the conclusion very strongly. Also, their major thesis appears strongly supported rather than unsupported as in the previous case.

Fig. 10 Fragment of an argumentation structure diagram for one of the fake news pieces. Large rectangles represent propositions, whereas small ones represent inferences

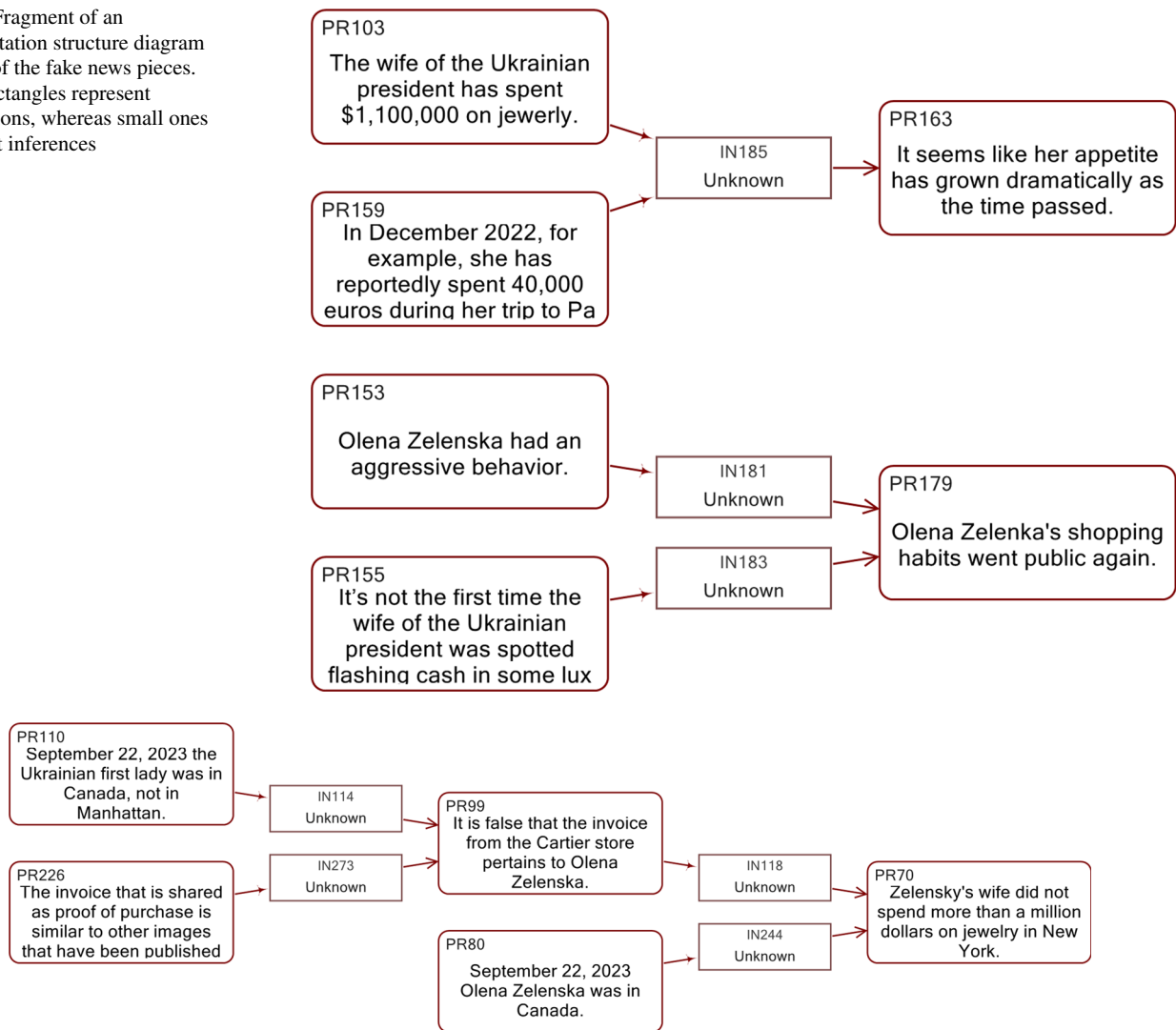


Fig. 11 Fragment of an argumentation structure diagram for the fact-checkers' report. Large rectangles represent propositions, whereas small ones represent inferences

Here, proposition PR70 is mapped to the “Olena Zelenska did not made a millionaire purchase at Cartier in New York” in the context model. This contrasts with proposition PR103 from the previous model (see Fig. 10). In this manner, we can identify a pair of propositions from different documents that support incompatible positions, thus allowing for an inter-textual analysis and traceability to the individual pieces of evidence on which it is based.

5.3 Discussion

In this section, we have shown a case study applying IAT/ML for a fact-checking process. In particular, both an ontology and five different argumentation models were constructed

and then compared to draw some conclusions. Agency analysis was not carried out in this case study, as it usually works better with a much larger corpus.

Ontological analysis, in particular, allowed us to identify the relevant entities in the opposed discourses. The fact that there is a single ontology for the complete corpus helps integrate the different perspectives, connect them via common or shared entities, and compare the potentially different foci of each position. For example, in this case study, it was found that Cartier's employee played a central role in defending the position that Zelenska had spent a million dollars in New York, whereas the opposed view did not rely at all on this employee. Imbalances like this highlight potential weak points for either of the parts.

Argumentation analysis, in turn, allowed us to visualise the depth, complexity and nature of the supports that each

speaker provides about their position. In the case study, defenders of the fake news piece boldly claimed their main thesis with no backing support, while opposed media only stated their major thesis once after a chain of inferences was in place. This allows us to compare and decide on which position is more likely true.

Although denotations were not used in this particular case study, they would become very useful in scenarios with a larger corpus. Denotations would allow the analyst to connect each proposition to the elements in the ontology being referred to. For example, “The wife of the Ukrainian president” in proposition PR103 in Fig. 10 clearly refers to the Zelenska entity in Fig. 9. By recording denotations in the models, an intertextual analysis becomes possible that can shed additional light on connections across documents in larger and more complex scenarios.

It is also interesting to highlight how the context model (composed of one theme, two positions and various agents) acts as a supporting infrastructure to which other models may refer. For example, the entities in the ontology representing perspectives (P1 and P2 in Fig. 9) are mapped to positions in the context, and each speaker in the various argumentation models is mapped to an agent in the context. Furthermore, and as described above, propositions from different documents are also mapped to opposing positions. This allows for intertextual analysis and powerful cross-model analytics that take these connections into account. Although this case study is small and any analyst can keep all the involved information in mind at the same time, inter-textual scenarios involving hundreds or thousands of documents, and tens of thousands of propositions, which would be unmanageable for a human analyst, can be as easily handled by LogosLink and IAT/ML.

6 Further validation

IAT/ML has been further validated through different mechanisms, including the implementation into the LogosLink toolset, the modelling of discourses in various projects, and the teaching of several courses.

Given that IAT/ML was developed under a practice-driven design science approach, validation focussed on two aspects:

- A. On the one hand, we aimed to stress-testing the **expressive power** of the metamodel by confronting it with as many different discourses from as many different sources and agents as possible. This is connected to satisfying requirements A, B, D and E in Sect. 3.1.
- B. On the other hand, we wanted to verify the **usefulness** of the metamodel to support a methodology-oriented software tool, which, in turn, would assess its understandability and usability. This is connected to satisfying requirement C in Sect. 3.1.

6.1 Implementation in LogosLink

The implementation of IAT/ML in the form of a software application was guided by Requirement C as described in Sect. 3.1 and aimed to satisfy validation goal B above. IAT/ML was implemented in the form of the LogosLink toolset, available from www.iatml.org/logoslink. LogosLink is a collection of libraries and user interface applications developed in C# on top of the Microsoft.NET Framework. It consists of over 245.000 lines of code organised in a modular structure so that it can be used as an interactive stand-alone application or integrated as part of other projects. The implementer was one of the authors (Gonzalez-Perez).

The Argumentation component of IAT/ML has been fully implemented as part of the ArgumentationEngine library, which offers a complete object model for discourse and argumentation modelling together with the functionality to save and load models, obtain statistics, and other related functions. The Ontology component, in turn, has been implemented as a separate library, OntologyEngine, which offers analogous features. A third library, Analytics, works on top of the previous two to carry out complex analytical techniques such as argument structure analysis or denotation analysis. Finally, the Desktop executable offers a Microsoft Windows-based desktop user interface capable of diagramming argumentation models and offering full features for argumentation and ontological modelling of discourses. Documentation for LogosLink (both user’s and developer’s) can be found at www.iatml.org/logoslink.

Ontological and argumentation models are stored as JSON files by LogosLink. These models can be formally validated against the metamodel by the libraries and used to perform an array of analytical procedures on them. For example, LogosLink can carry out centrality or argumentation structure analytics to show what elements in an argumentation model are most central, which are the major theses being defended, which are the key foundations on which each speaker bases their discourse, etc. Some other analytics work at the corpus level, operating on multiple ontologies and argumentation models at once. Some examples include different collocation flavours (lexical, semantic and lexical/semantic), denotation (which examines the terms used by each speaker to refer to each concept in the ontology), or intertextuality (which assigns a score to each pair of texts depending on how related they are via common denotations or explicit references).

The implementation of a theory or metamodel, such as the one presented in this paper, in the form of a software tool, constitutes a good validation mechanism, especially in relation to quality factors such as usefulness, integrity, performance and understandability. A metamodel constructed with a modelling tool can be formally validated, but when it is exposed

to actual users in a variety of environments and incarnations, issues arise that could never be detected by a modelling tool alone. For example, in relation to integrity, implementing IAT/ML in LogosLink helped us detect redundant and lacking associations between classes in the metamodel; once code was written, it was easy to see what references were being used in run-time and which ones were unused. In relation to performance, implementation helped us experiment with different type structures and hierarchies. In particular, the metamodel class structure for argumentation described under Sect. 4.3 is the third iteration after two attempts that did not map too well to users' expectations of how information should be organised in an argumentation model. Finally, and in relation to understandability, it is obvious that the user interface of an interactive application somehow maps to the metamodel being implemented. Although the mapping is not always one-to-one, the "shape" and structure of the metamodel determine, to a large extent, how well the information is understood by users. In our case, for example, it was not clear whether denotations, which mediate between ontology and argumentation models, should be shown to users as part of ontology or argumentation analysis, or both. In the end, we decided to include them in argumentation analysis only, and this had a significant impact on the metamodel itself.

Finally, we must say that context and agency analysis are being in the process of being implemented, but are not part of the publicly available version of LogosLink yet. This step-wise implementation approach is a natural consequence of the incremental and practice-oriented development approach that was taken to the construction of IAT/ML, as described in Sect. 3.

6.2 Research projects

Aiming to satisfy validation goal A above, IAT/ML was used to model discourses in a number of projects, using LogosLink for ontology and argumentation modelling, and a word processor for agency modelling. One of these projects was "COVID19 en español: investigación interdisciplinaria sobre terminología, temáticas y comunicación de la ciencia" [COVID-19 in Spanish: Interdisciplinary Research on Terminology, Themes and Science Communication], funded by the Spanish National Research Council between 2020 and 2022. This project gathered a corpus of 877 COVID-related popular science articles published by The Conversation Spain [74] throughout the critical year of 2020, amounting to 962,886 words, and developed ontological and argumentation models to find out the main strategies and mechanisms used to disseminate information about the pandemic within the Spanish-speaking world. Figure 12 shows a screenshot of an argumentation model developed during this project.

Four analysts worked on this project, including one of the authors (Gonzalez-Perez) plus additional technical staff from

three different organisations. This meant that the issue of traceability and reproducibility (Requirement B in Sect. 3.1) of analysis results was present from the beginning. Having clear and comprehensive documentation, as well as a software tool that assisted them during analysis, proved to be crucial.

Another project where IAT/ML and LogosLink have been used is "Heritage 3.0: Argumentation and Conceptual Modelling for Enhanced Cultural Heritage Participation and Management Policies" (<https://www.incipit.csic.es/en/project/acme>), which aims to analyse discourses amounting to 323.174 words across 517 texts, which include transcribed interviews, historical documentation from the 1950s onwards, student essays and social media posts about five different case studies. The project is led by one of the authors (Gonzalez-Perez) and involves a team of 18 people, including the other three authors, all of which have used IAT/ML and LogosLink extensively. Each case study involved a particular heritage element (such as a monument or a cultural landscape) and was coded as a topic in the corpus. The resulting corpus was quite large for a full manual analysis, so IAT/ML had to be adapted to the time and resources that were available, thus testing the approach against Requirement E as described in Sect. 3.1. Adaptations included focussing on topic-level ontology analysis, selecting a sample of texts from each case study for argumentation and agency analysis, and using the analysis results for action planning that would be useful to local governments in the management of cultural heritage. At the time of writing, analysis is mid-way through, with a clear indication that customisation capabilities are good. Also, early tests with some selected heritage managers from two of the five case studies have shown that traceability (Requirement B) has shown to be extremely valuable to demonstrate the soundness of the analysis results with non-technical stakeholders. For example, the traceability capabilities of IAT/ML and LogosLink allow us to explain why a certain conclusion is obtained, on which discourses it is based, by whom, and in which context.

A third project where IAT/ML and LogosLink have been used is the ongoing doctoral work of one of the authors (Calderón-Cerrato), supervised by Gonzalez-Perez and Pereira-Fariña. So far, this project has gathered a corpus of 61 texts, amounting to 53,543 words, related to identity and polarisation in cultural heritage and feminism, from sources such as legislation, press articles, transcribed interviews and social media posts. Although this corpus is smaller than the previous ones, this project is carrying out agency analysis as well as ontological and argumentation analysis, and producing integrated diagnostics of dissonant heritage situations supported by the context common sub-domain, thus allowing the integration of conclusions from ontologies, argumentation models and agency models in terms of themes, positions and agents. In this regard, Requirement A as described in

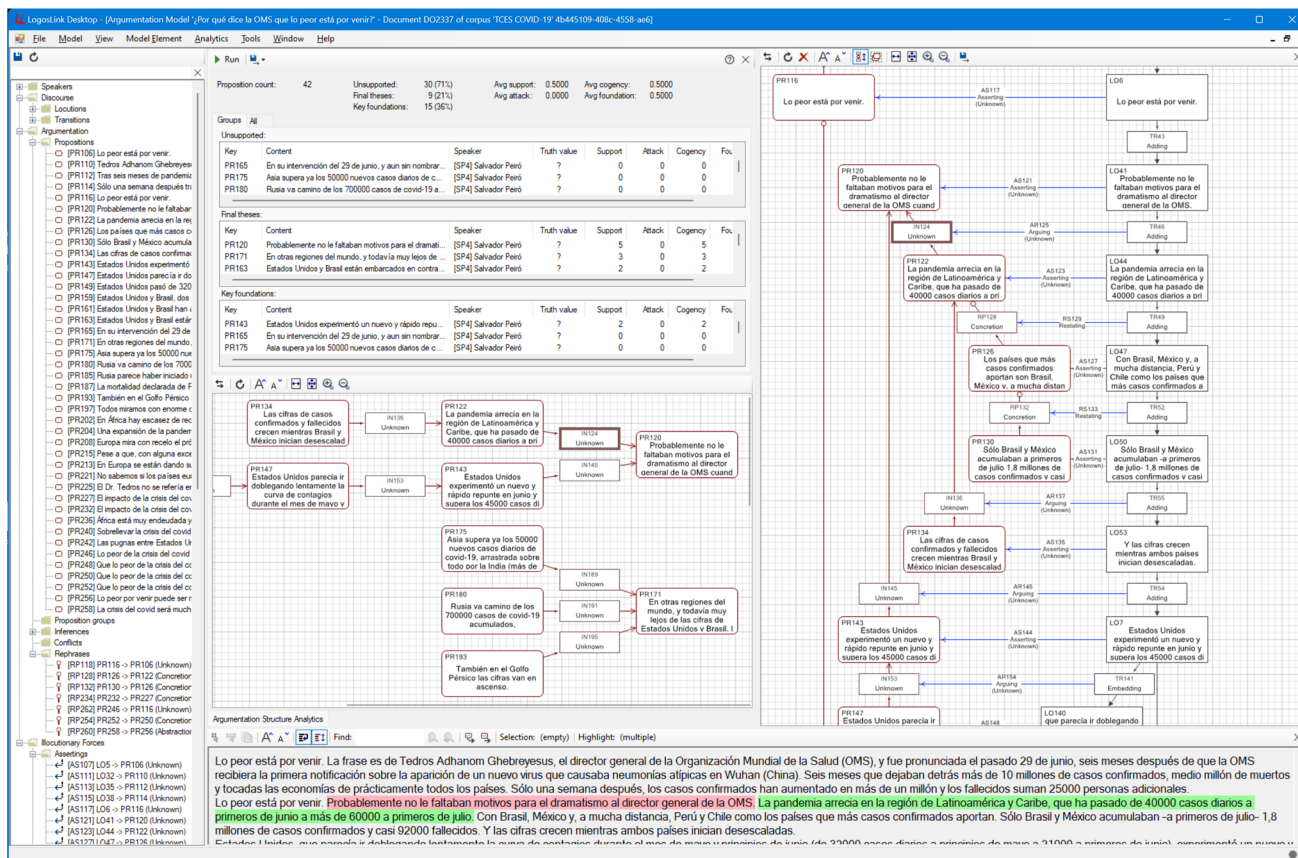


Fig. 12 Screenshot of LogosLink Desktop showing an argumentation model for an article on COVID-19. Black boxes represent locutions and transitions. Red boxes represent propositions and argument relations.

Blue arrows represent illocutionary forces. The central panel shows the results of an argumentation structure analytics

Sect. 3.1 has been intensely put to test by this project and satisfactorily validated.

In addition, the incorporation of feminist identities as an extra theme at a later stage of this project allowed us to further validate IAT/ML in relation to its usefulness for a corpus and topic other than those employed during development. This is one of the suggested approaches to validation offered by [34].

Finally, argumentation models developed during this project were also employed during discussions of Calderón-Cerrato and Pereira-Fariña with the staff of ArgTech, the original creators of IAT [8, 40], which were experts in IAT but did not have previous exposure to IAT/ML. IAT/ML's notational, to some extent, conceptual, compatibility with IAT, as dictated by Requirement D, worked very well to facilitate the communication.

6.3 Teaching

Following Requirement C in Sect. 3.1, and aiming to satisfy validation goals A and B above, comprehensive documentation has been developed for IAT/ML. In addition to a process-oriented document that provides recipe-style guidance on how to apply the methodology, pattern-oriented documents have been also developed to provide details on how to model individual situations of many kinds. These documents are available from www.iatml.org.

This documentation, plus additional materials, were used to teach a 21-h postgraduate course in Santiago de Compostela, Spain, in March 2023. Participants included PhD students, professionals and professors in archaeology, sociology, geography, architecture and law, each of them providing a different usage scenario to which IAT/ML was applied. All the participants used IAT/ML and LogosLink on their particular fields of study. Teaching not only served to flesh out complex details of the methodology and explain them to others; it also worked as a test bed for the documentation, which was evaluated as of “outstanding value” by course students.

Teaching also provided a wealth of feedback from students about conceptualisations, process and usability aspects, most of which has been incorporated into the approach.

Follow-up teaching workshops, as well as a consultancy service on discourse analysis with IAT/ML, have been developed in collaboration with the Outreach Unit at Incipit CSIC, where two of the authors work. At the time of writing, three additional workshops have been carried out in Spain and Portugal, as well as two consultancy projects, involving the use of IAT/ML and LogosLink by an additional 20 + people from multiple organisations. The fact that IAT/ML and LogosLink have been used in different settings and by different users with no previous involvement in their development showed, once again, the usefulness and expressive power of the approach.

6.4 Lessons learned

All in all, we can confidently state that, by using IAT/ML, an analyst (or a team of analysts) can obtain a deep and nuanced understanding of a corpus of texts in terms of issues, themes, positions and agents by using ontological, argumentation and agency analysis, as determined by the original purpose as described in Sect. 3.1. Furthermore, teaching experiences have shown that becoming a novice but competent analyst is as easy as attending a 3-day, 21-h workshop.

Three major lessons have been learned from using IAT/ML and LogosLink over the last four years. Firstly, it is evident that IAT/ML and LogosLink are very useful resources to understand complex polarised situations through the associated discourses. This has been shown in a variety of situations and projects as well as in teaching.

Secondly, we have seen that any methodology, process or tool that aims to be useful to a wide array of users must be modular and flexible. Although this article focuses on the metamodel aspects of IAT/ML, IAT/ML is indeed a full methodology, as it recommends a series of steps and intermediate products. We have worked in situational method engineering in the past (see, e.g. [27, 45]), so many of the methodological principles of modularity, composability and separation of concerns (especially between the process and product realms) that are part of IAT/ML have been taken from ISO/IEC 24744 [39] and related works. Still, it is difficult to produce one methodology that fits all, and expert guidance and consultancy is often required to successfully apply IAT/ML in complex scenarios. We are working to make IAT/ML easier to use by incorporating many of the suggestions that we receive from users and clients.

Thirdly, we have learned that developing and maintaining a non-trivial metamodel-based software tool like LogosLink is a double-edged sword. On the one hand, implementing a theory or metamodel as a software tool provides excellent validation opportunities, as discussed in Sect. 6.1. On the

other hand, doing this is very expensive in terms of time and effort, especially when working from academia. We have invested over 2290 person-hours in software development alone over 3 years, and the foreseen maintenance cycle for LogosLink version 1 will span another 3. While we maintain this, we are already working on LogosLink version 2, which will be (partially) multiplatform, much more modern, and more functional. Software technologies evolve fast, and despite the long-term support provided by Microsoft for the .NET platform, it is still difficult to keep up the pace and evolve LogosLink fast enough as to keep obsolescence under control.

7 Conclusions

In this paper, we have presented IAT/ML, a domain-specific approach for the modelling and representation of discourses based on the combination of three modelling perspectives: ontological, argumentation and agency. Ontological and argumentation analyses have been fully incorporated into the LogosLink supporting tool. In this regard, IAT/ML and LogosLink are the first of its kind, as no other approaches, as far as we know, integrate different perspectives under a common and inter-connected modelling approach.

IAT/ML has been constructed empirically from clear requirements and by following a known method for DSML development while practitioners were using it. The same practitioners worked to validate and enrich the approach, both as part of the initial projects as well as under new discourse themes that were added later. We have also shown a case study developed with IAT/ML and the LogosLink toolset, focusing on fact checking. Additional validation was described as part of various projects and teaching efforts.

Regarding the initial research sub-questions (see Sect. 3.4), we can now briefly answer them:

1. (a) What concepts and patterns are found in the discourse analysis process and its domain that are common to existing approaches? (b) Which ones we do not find in existing approaches, but are necessary after our experience analysing discourses?

We took the overall ontological modelling approach from ConML and adjusted them to make them multi-level modelling compliant. We took most of the concepts in IAT and extended them for comprehensive argumentation modelling. We developed an agency modelling conceptualisation from scratch. Overall, about 75% of the metamodel elements in IAT/ML are original, and 25% have been taken more or less straight from previous developments.

2. Can we develop a domain-specific modelling language that fully describes and supports both the discourse analysis domain as well as the process for the three perspectives (ontological, argumentation and agency)?

Yes. This was feasible and practical. Furthermore, the three perspectives were successfully integrated via the fourth domain of analysis Context.

3. Is it viable to implement this language in a modelling tool?

Indeed, we developed LogosLink, which has been extensively used over the last four years in a number of projects, both inside and outside of the research team that developed it.

4. What degree of coverage and traceability does this language offer for the discourse analysis process, given a corpus to be analysed that is different from those used during development?

Experiences show that coverage is good, in the sense that IAT/ML allows you to describe the discourse under analysis from three different integrated perspectives. In addition, traceability is excellent, especially when compared to the previous state of the art, which lacked much in terms of documentation, guidelines and reproducibility.

5. How viable, and at what cost, is to customise the language in order to cater for specific situations and project needs?

This depends on the degree of customisation. According to our experience, very little time and effort are needed for basic and moderate customisation and integration. More time and effort are expected to be needed in more complex customisation settings, but these have not been found so far, although they are anticipated for use cases related to some of the stakeholders described in Section 3.2.

Additional future work includes adding automatic assistance for the analyst to segment the text and reconstruct propositions by using Large Language Models (LLMs), and detecting ontology elements through Named-Entity Recognition (NER). Additional analytics that operate on top of ontology, argumentation and agency models to produce quantitative and visual results are also being developed and tested. Metamodeling-wise, extensions are being planned to cater for diachronic argumentation analysis (as, for example, in changing one's mind over time) and stronger support for intertextual connections. An additional line of work for the future is that of a large-scale validation effort of the full approach. This has been partially tackled already, especially in the ontological realm, by works such as [45, 46]. These works have shown that an ontological analysis of discourse is relevant to various communities. Still, it would be very beneficial to carry out a comprehensive exercise that also includes argumentation and agency analysis, and involves the different types of users and stakeholders that are described in Sect.

- 3.2. The ongoing HYBRIDS MSCA DN project (<https://hybridsproject.eu>), which focuses on disinformation and hate speech, is a perfect setting for this, and we plan to conduct such a study over the next 2 years. Finally, LogosLink is being further developed to cover agency analysis.

IAT/ML is documented online on www.iatml.org, and LogosLink can be downloaded for free. We hope that this domain-specific approach, together with its tooling support, will contribute to better and more reliable discourse analysis projects and easier and more powerful discourse understanding, evaluation and fact-checking.

Acknowledgements The work presented in this paper has been partially funded by the AEI (Spanish National Research Agency) through grants PID2020-114758RB-I00, MCIN/AEI/<https://doi.org/10.13039/501100011033> and PID2020-115482GB-I00, MCIN/AEI/<https://doi.org/10.13039/501100011033>.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Almeida J. P. A., Frank U., and Kühne T.: Multi-Level Modelling Dagstuhl Seminar 17492, Wadern, Germany (2018). <https://doi.org/10.4230/DagRep.7.12.18>
- Alqaaidi S. and Bozorgi E.: A survey on recent named entity recognition and relation classification methods with focus on few-shot learning approaches (2023)
- Austin J. L.: How to do things with words: The William James lectures delivered at Harvard University in 1955, 2nd ed. Oxford [etc.]: University Press (1989)
- Austin, J.L.: How to Do Things with Words. Martino Fine Books, Reprint (2018)
- Beydoun, G., et al.: FAML: a generic metamodel for MAS development. IEEE Trans. Software Eng. **35**(6), 841–863 (2009). <https://doi.org/10.1109/TSE.2009.34>
- Bratman, M.E.: Intention, plans, and practical reason. CSLI Publications (1999)
- Breeze, R.: Critical discourse analysis and its critics. Pragmat. Q. Publ. Int. Pragmat. Assoc. (IPrA) (2022). <https://doi.org/10.1075/prag.21.4.01bre>
- Centre for Argument Technology: A quick start guide to inference anchoring theory (IAT) (2017)

9. Centre for Argument Technology: Annotation guidelines for inference anchoring theory (IAT) with support for conventional implicatures (CIs) (2018) [Online]. Available: <https://typo.uni-konstanz.de/add-up/wp-content/uploads/2018/04/IAT-CI-Guidelines.pdf>
10. Chesñevar, C., et al.: Towards an argument interchange format. *Knowl. Eng. Rev.* **21**(4), 293–316 (2006). <https://doi.org/10.1017/S0269888906001044>
11. Clark T., Gonzalez-Perez C., and 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), vol. 1286, C. Atkinson, G. Grossmann, T. Kühne, and J. de Lara, Eds. Regensburg, Germany: CEUR-WS.org, pp. 43–52 (2014)
12. Conklin, J., Begeman, M.L.: gIBIS: a hypertext tool for exploratory policy discussion. *ACM Trans. Inf. Syst.* **6**(4), 303–331 (1988). <https://doi.org/10.1145/58566.59297>
13. Crăciunean, D.-C., Volovici, D.: Conceptualization of modelling methods in the context of categorical mechanisms, pp. 543–565. Springer (2022)
14. Daniel A., Flew T. and Spurgeon C.: The promise of computational journalism. In: Proceedings of the Australian and New Zealand Communication Association (ANZCA) Conference 2010: Media, Democracy and Change, pp. 1–19 (2010)
15. Van Dijk, T.A.: Critical Discourse Analysis. In: Tannen, D., Schiffrin, D., Hamilton, H. (eds.) *Handbook of Discourse Analysis*. Blackwell, Oxford (2001)
16. van Dijk T.: *Elite Discourse and Racism*. 2455 Teller Road, Thousand Oaks California 91320 United States : SAGE Publications, Inc., (1993)
17. Doerr, M., Kritsotaki, A., Boutsika, K.: Factual argumentation—a core model for assertions making. *J. Comput. Cultural Herit.* **3**(3), 1–34 (2011). <https://doi.org/10.1145/1921614.1921615>
18. European Commission and European Political Strategy Centre, *10 trends shaping democracy in a volatile world*. Publications Office, (2019)
19. Frank U.: Some guidelines for the conception of domain-specific modelling languages, Enterprise modelling and information systems architectures (EMISA 2011), (2011)
20. Gee, J.P.: *An Introduction to Discourse Analysis: Theory and Method*. Routledge, London (2014)
21. Gonzalez-Perez C.: How Ontologies Can Help in Software Engineering. In: Grand Timely Topics in Software Engineering, vol. 10223 LNCS, no. 10223, J. Cunha, J. P. Fernandes, R. Lämmel, J. Saraiva, and V. Zaytsev, Eds. Springer, pp. 26–44 (2017)
22. Gonzalez-Perez, C.: *Information Modelling for Archaeology and Anthropology*. Springer, Berlin (2018)
23. Gonzalez-Perez, C.: Connecting discourse and domain models in discourse analysis through ontological proxies. *Electronics (Basel)* **9**(11), 1955 (2020). <https://doi.org/10.3390/electronics9111955>
24. Gonzalez-Perez C.: Conceptual Modelling Language for the Humanities and Social Sciences. In: Sixth International Conference on Research Challenges in Information Science (RCIS), 2012, C. Rolland, J. Castro, and O. Pastor, Eds. IEEE Computer Society, pp. 396–401 (2012)
25. Gonzalez-Perez C.: Modelling Temporality and Subjectivity in ConML. In: 7th IEEE International Conference on Research Challenges in Information Science (RCIS 2013), R. Wieringa and S. Nurcan, Eds. Paris (France): IEEE Computer Society, pp. 1–6 (2013)
26. Gonzalez-Perez, C.: Ontological Proxies to Augment the Expressiveness of Discourse Analysis. In: Gamallo, P., García, M., Martín-Rodilla, P., Pereira-Fariña, M. (eds.) *Hybrid Intelligence for Natural Language Processing Tasks 2020*, vol. 2693, pp. 1–3. CEUR-WS.org (2020)
27. Gonzalez-Perez, C.: Supporting Situational Method Engineering with ISO/IEC 24744 and the Work Product Pool Approach. In: Ralyté, J., Brinkkemper, S., Henderson-Sellers, B. (eds.) *Situational Method Engineering: Fundamentals and Experiences*, pp. 7–18. Springer, Boston, MA (2007). https://doi.org/10.1007/978-0-387-73947-2_3
28. Gonzalez-Perez, C., Henderson-Sellers, B.: A powertype-based metamodeling framework. *Softw. Syst. Modell.* **5**(1), 72–90 (2006). <https://doi.org/10.1007/s10270-005-0099-9>
29. Gonzalez-Perez C., Pereira-Fariña M., and Calderón-Cerrato B.: IAT/ML, <http://www.iatml.org/> (2021)
30. Google: Google Fact-Check Tools, <https://newsinitiative.withgoogle.com/en-gb/resources/trainings/verification/google-fact-check-tools> (accessed Nov. 13, 2023) (2023)
31. Grice H. P.: Logic and Conversation. In: *The Logic of Grammar*, D. Davidson and G. Harman, Eds., pp. 64–75 (1975)
32. Guizzardi G., Ferreira Pires L., and van Sinderen M.: An ontology-based approach for evaluating the domain appropriateness and comprehensibility appropriateness of modeling languages, pp. 691–705 (2005)
33. Guo, Z., Schlichtkrull, M., Vlachos, A.: A survey on automated fact-checking. *Trans. Assoc. Comput. Linguist.* **10**, 178–206 (2022). https://doi.org/10.1162/tacl_a_00454
34. Hamdaqa M., Metz L. A. P., and Qasse I.: icontractml: A domain-specific language for modeling and deploying smart contracts onto multiple blockchain platforms, pp. 34–43 (2020)
35. Henderson-Sellers, B.: Bridging metamodels and ontologies in software engineering. *J. Syst. Softw.* **84**(2), 301–313 (2011). <https://doi.org/10.1016/j.jss.2010.10.025>
36. Iivari, J.: Twelve Theses on Design Science Research in Information Systems. In: Hevner, A., Chatterjee, S. (eds.) *Design research in information systems: Theory and practice*, pp. 43–62. Springer, Boston (2010). https://doi.org/10.1007/978-1-4419-5653-8_5
37. Hinton, M., Wagemans, J.H.M.: Evaluating reasoning in natural arguments: a procedural approach. *Argumentation* **36**(1), 61–84 (2022). <https://doi.org/10.1007/s10503-021-09555-1>
38. Incipit CSIC, ‘ConML Technical Specification’, Incipit CSIC, 2020. [Online]. Available: <http://www.conml.org/Resources/TechSpec.aspx>
39. ISO/IEC, ‘Software Engineering - Metamodel for Development Methodologies’. ISO/IEC, Geneva, [Online]. Available: <https://www.iso.org/standard/62644.html> (2014)
40. Janier M., Aakhus M., Budzynska K., and Reed C.: Modeling argumentative activity with Inference Anchoring Theory. In: *Argumentation and Reasoned Action. Volume I Proceedings of the 1st European Conference on Argumentation*, vol. 1, no. 62, D. Mohhamed and M. Lewinski, Eds. College Publications, (2016)
41. Johnstone, B.: *Discourse Analysis*. Wiley, New Jersey (2018)
42. Krämer M.: Controlling the processing of smart city data in the cloud with domain-specific languages’, pp. 824–829 (2014)
43. Lawrence, J., Reed, C.: Argument mining: a survey. *Comput. Linguist.* **45**(4), 765–818 (2020). https://doi.org/10.1162/coli_a_00364
44. Mao, R., et al.: A survey on semantic processing techniques. *Inf. Fusion* **101**, 101988 (2024). <https://doi.org/10.1016/j.inffus.2023.101988>
45. Martín-Rodilla P. and Gonzalez-Perez C.: An ISO/IEC 24744-Derived Modelling Language for Discourse Analysis. In: *Research Challenges in Information Science (RCIS), 2014 IEEE Eighth International Conference on*, M. Bajec, M. Collard, and R. Deneckère, Eds. IEEE Computer Society, (2014)
46. Martín-Rodilla, P., Gonzalez-Perez, C.: Same text, same discourse? Empirical validation of a discourse analysis methodology for cultural heritage. *Digital Scholarsh. Humanities* **38**(1), 224–239 (2023). <https://doi.org/10.1093/llc/fqac038>

47. Moscato, V., Postiglione, M., Sperli, G.: Few-shot named entity recognition: definition, taxonomy and research directions. *ACM Trans. Intell. Syst. Technol.* **14**(5), 1–46 (2023). <https://doi.org/10.1145/3609483>
48. Perelman C. and Olbrechts-Tyteca L.: *Traité de l'argumentation: La nouvelle rhétorique*. Presses Universitaires de France, (1958)
49. Pescador A., Garmendia A., Guerra E., Cuadrado J. S., and de Lara J.: Pattern-based development of domain-specific modelling languages', pp. 166–175 (2015)
50. Rao A. S. and Georgeff M. P.: Modeling Rational Agents within a BDI-Architecture. In: Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning, pp. 473–484 (1991)
51. Reed C. and Budzynska K.: How Dialogues Create Arguments', in *ISSA Proceedings 2010*, [Online]. Available: <http://rozenbergquarterly.com/issa-proceedings-2010-how-dialogues-create-arguments/> (2010)
52. Reed, C., et al.: The argument web: an online ecosystem of tools, systems and services for argumentation. *Philos. Technol.* **30**(2), 137–160 (2017). <https://doi.org/10.1007/s13347-017-0260-8>
53. Reyes Román, J.F., León Palacio, A., García Simón, A., Beyrouiti, R.C., Pastor, O.: Integration of clinical and genomic data to enhance precision medicine: a case of study applied to the retina-macula. *Softw. Syst. Model.* **22**(1), 159–174 (2023)
54. Rolland, C., Prakash, N., Benjamin, A.: A multi-model view of process modelling. *Requir. Eng. J.* **4**(4), 169–187 (1999)
55. Searle, J.R., Vanderveken, D.: *Foundations of Illocutionary Logic*. Cambridge University Press, Cambridge (1985)
56. Sidorova, E.A., Akhmadeeva, I.R., Kononenko, I.S., Chagina, P.M.: Argument extraction based on the indicator approach. *Pattern Recognit Image Anal.* **33**(3), 498–505 (2023). <https://doi.org/10.1134/S1054661823030410>
57. Snaith, M.: An argument-based framework for selecting dialogue move types and content. *Comput. Models Argument* **326**, 355–362 (2020)
58. Suchánek M.: *OntoUML Specification*, <https://ontouml.readthedocs.io/> (accessed Oct. 09, 2020) (2018)
59. Toulmin, S.E.: *The Uses of Argument*. Cambridge University Press, Cambridge (2003)
60. Visser, J., Lawrence, J., Reed, C.: Reason-checking fake news. *Commun. ACM* **63**(11), 38–40 (2020). <https://doi.org/10.1145/3397189>
61. Visser, J., Lawrence, J., Reed, C., Wagemans, J., Walton, D.: Annotating argument schemes. *Argumentation* **35**(1), 101–139 (2021). <https://doi.org/10.1007/s10503-020-09519-x>
62. Vlachos A. and Riedel S.: Fact Checking: Task definition and dataset construction', in Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science, pp. 18–22. <https://doi.org/10.3115/v1/W14-2508>. (2014)
63. Vosoughi, S., Roy, D., Aral, S.: The spread of true and false news online. *Science* **359**(6380), 1146–1151 (2018). <https://doi.org/10.1126/science.aap9559>
64. Wagemans J.: *Period Table of Arguments*, <https://periodic-table-of-arguments.org/> (accessed Oct. 16, 2020). (2020)
65. Walton, D., Reed, C., Macagno, F.: *Argumentation Schemes*. Cambridge University Press, Cambridge (2008)
66. Wetzell L.: Types and Tokens: The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, [Online]. Available: <https://plato.stanford.edu/archives/fall2018/entries/types-tokens/>. (2018)
67. Wieringa and Moral A.: Technical action research as a validation method in information systems design science, pp. 220–238 (2012)
68. Wittgenstein L.: *Philosophical investigations*, 3. ed., re. Oxford: Blackwell, (1989)
69. World Wide Web Consortium, 'OWL 2 Web Ontology Language'. World Wide Web Consortium, [Online]. Available: <http://www.w3.org/TR/2012/REC-owl2-overview-20121211/>. (2012)
70. World Wide Web Consortium, 'RDF/XML Syntax Specification (Revised)'. World Wide Web Consortium, [Online]. Available: <http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/>. (2004)
71. Zasytkin A. S., Pimenov I. S., and Salomatina N. V.: The Combined Approach to Identifying Argumentation Structures in Short Scientific Papers', in 2023 IEEE 24th International Conference of Young Professionals in Electron Devices and Materials (*EDM*), Jun. 2023, pp. 1800–1805, <https://doi.org/10.1109/EDM58354.2023.10225223>
72. Zhou, S., Wang, N., Wang, L., Liu, H., Zhang, R.: CancerBERT: a cancer domain-specific language model for extracting breast cancer phenotypes from electronic health records. *J. Am. Med. Inform. Assoc.* **29**, 1208–1216 (2022)
73. Flowerdew, J., Richardson, J.E. (eds.): *The Routledge Handbook of Critical Discourse Studies*. Routledge (2017). <https://doi.org/10.4324/9781315739342>
74. *The Conversation*, Spanish Edition', <https://theconversation.com/es> (accessed Oct. 16, 2020). (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Cesar Gonzalez-Perez Cesar Gonzalez-Perez is a Staff Scientist at the Institute for Heritage Sciences (Incipit) at the Spanish National Research Council (CSIC), where he leads a co-research line in software engineering, language and cultural heritage. Previously, Cesar has worked in various academic and industry organisations in Spain and Australia, in the areas of conceptual modelling, meta-modelling, situational method engineering and software systems development, with a special orientation towards heritage issues. Cesar is active in the international standardisation community, has been promoter of 3 technology-based companies, and has authored over 100 publications.



Martín Pereira-Fariña Martín Pereira-Fariña is an associate professor at the Department of Philosophy and Anthropology at the University of Santiago de Compostela, Spain. His main research interests include argumentation theory, the philosophy of language, and computational models for argumentation. Martín has been a postdoctoral researcher at the Centre for Argument Technology (University of Dundee, Scotland) and Incipit CSIC (Santiago de Compostela, Spain). He

has authored over 30 publications.



Beatriz Calderón-Cerrato Beatriz Calderón-Cerrato is a PhD researcher at the Institute for Heritage Sciences (Incipit) at the Spanish National Research Council (CSIC). Previously, she graduated in Spanish Language and Literature from the University of the Balearic Islands, and studied a master's degree in applied linguistics at the University of Santiago de Compostela. Her current research focuses on discourse analysis applied to dissonant heritage and feminism

using argumentation, ontologies, and agency analysis.



Patricia Martín-Rodilla Patricia Martín-Rodilla is a tenured scientist at the Spanish National Research Council (IEGPS-CSIC), leading a research line on Digital Humanities. Previously, she was an assistant professor at the Information Retrieval Laboratory of the University of A Coruña. Patricia has participated in a multitude of regional, national and European projects. Her main interests are the intersection between language, software engineering and the humanities. This

includes information modelling and ontologies, information retrieval in the humanities, computational linguistics, discourse analysis and corpus and social networks in the humanities.