



White Paper

version 1.0.2 · 8 May 2024



IAT/ML by Institute of Heritage Sciences (Incipit), Spanish National Research Council (CSIC) & University of Santiago de Compostela (USC)



is licensed under a Creative Commons Attribution 4.0 International License.

This document and its contents were created by Cesar Gonzalez-Perez, Martín Pereira-Fariña and Beatriz Calderón-Cerrato.

Partial funding was provided by Incipit · CSIC, USC, and project ACME (grant number PID2020-114758RB-I00 funded by MCIN/AEI/10.13039/501100011033, Retos de la Sociedad, Plan Estatal 2017-2020, Spain).

8 May 2024 14:36 · revision 844

Table of Contents

Introduction	3
Background and Motivation.....	4
IAT/ML Overview.....	5
Major Properties	5
Overall Process.....	5
Main Analysis Perspectives	5
Usage Scenarios	10
Benefits and Limitations	10
Practical Advice for Adoption.....	11
Alternatives	12
IAT	12
ConML	12
Critical Discourse Analysis.....	12
RST	12
Artificial Intelligence	13
No Particular Approach.....	13
Conclusion	14
Acknowledgements.....	15
References.....	16

Introduction

IAT/ML is a method for the joint analysis of discourse that combines ontological, argumentation and agency perspectives. It is based on Inference Anchoring Theory (IAT) and the ConML conceptual modelling language.

This document presents a brief overview of IAT/ML, its background and limitations, some usage scenarios, alternative approaches, and some practical advice to put it into practice.

This document is intentionally brief and simple. If you are interested in a complete description of the IAT/ML process, please see the *IAT/ML Analysis Process Guidelines*. If you are interested in a technical specification of IAT/ML, please see the *IAT/ML Technical Specification* document.

For more information on IAT/ML and additional documentation, please visit www.iatml.org.

Background and Motivation

Understanding what people say is a crucial part of all aspects of social life. Understanding why people say what they say is equally important, but often more difficult, as the reasons why someone expresses an idea or holds a particular belief are not always evident and explicit. This understanding is even more important, and more difficult, when we move from the personal realm to that of public discourses such as politics or mass media.

Many scientific approaches have been proposed to discourse analysis. However, most of the efforts in this area focus too much on what is being said and pays little attention to the actual things in the world referred to by the speakers, the reasons behind their discourse, or their beliefs, desires and intentions. At the same time, many approaches have been suggested to represent things in the world in a manner that can be studied and reasoned about with ease, such as those related to ontology modelling. However, most of the efforts of this kind focus exclusively on the things themselves and pay little attention to how these things are described through language, and how statements about them are made by speakers. Finally, there is extensive work on critical discourse analysis, which relates to the political and social positioning of the analyst in relation to the text. However, critical discourse analysis is often criticised for being too subjective and lacking proper grounding.

Neither of these approaches, by itself, is sufficient to gain a comprehensive understanding of complex social phenomena such as cultural heritage, public policy, or gender identity. For example, IAT [4], [6] offers good support to analysing the argumentation inherent to a discourse, but does not consider the world that the discourse refers to or the overall social or political stances of speakers. Similarly, technologies such as the ConML conceptual modelling language [2], [3] can help with the representation and analysis of the things in the world, but is unable to help with the associated linguistic and discursive aspects. In order to obtain a comprehensive understanding of what people say, and why, a triple joint analysis is needed, one that combines ontological, argumentation and agency analysis in an inter-connected manner.

In this document we provide an overview of IAT/ML, a method for joint ontological, argumentation and agency discourse analysis that has been developed to address this gap.

IAT/ML Overview

IAT/ML stands for *Inference Anchoring Theory / Modelling Language*. This obeys to the fact that IAT/ML is strongly based on Inference Anchoring Theory (IAT), a proven approach for argumentation analysis [4], [6], plus the ConML conceptual modelling language, a simple and powerful modelling language especially oriented towards the humanities and social sciences [2], [3]. In addition, IAT/ML takes some techniques from critical discourse analysis for agency analysis.

The following sections present a description of how IAT/ML is organised, how it can be used in different scenarios, what its benefits and limitations are, and some practical advice for its adoption.

Major Properties

IAT/ML is characterised by the following major properties:

- It focuses on **discourse analysis**. Although IAT/ML can help you shed light on the mental states and actions of people, it is always done via their discourses.
- It is **systematic and rigorous**. With IAT/ML, you work through a sequence of steps, each of them anchored on the previous. In this manner, you or anyone else can always trace a conclusion back to its sources and assess how well founded it is.
- It is **trans-disciplinary**. IAT/ML is based on principles from linguistics, philosophy of language, argumentation theory, conceptual modelling, sociology and ontology engineering. This provides a wider and richer conceptualisation of discourse and a better-informed array of analytical techniques.

Overall Process

IAT/ML works by guiding you through the following phases:

1. **Initiation**, during which you establish the situation or problem being addressed, define the context of your project in terms of themes, positions and agents, and build a corpus with the relevant texts to analyse.
2. **Analysis**, during which you manually analyse the gathered texts from different perspectives: ontological, argumentation and agency (see next section).
3. **Analytics**, during which you apply automated algorithms to process the results of the analysis and obtain additional data and visualisations.
4. **Results**, during which you compose one or more results documents based on the previous findings, and which will be used to understand or act on the situation or problem being addressed.

Please see the *IAT/ML Analysis Process Guidelines* document for more details.

Main Analysis Perspectives

IAT/ML is organised around three major analysis perspectives, as shown in Figure 1:

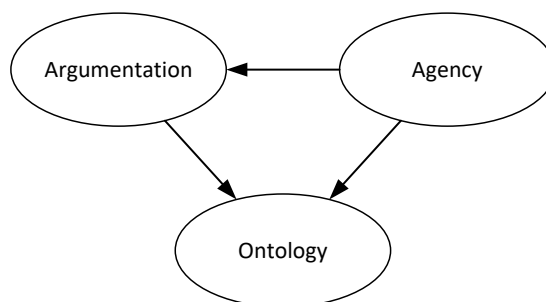


Figure 1. Main analysis perspectives in IAT/ML. Arrows represent dependencies.

Ontology analysis is concerned with the things in the world being referred to by the discourse, expressed in terms of entities plus their properties and relationships.

Argumentation analysis deals with the inferential structure of the discourse, including its propositions and argumentation relations such as inferences, conflicts and rephrasings. Argumentation relies on the ontology as people argue about things in the world.

Finally, agency analysis is connected with the beliefs and intentions of the speakers in relation to their discourse. It relies on the previous two kinds of analysis.

It is not mandatory to carry out the three kinds of analysis when working with IAT/ML. Although the most information is obtained when doing so, you may choose to carry out analysis of only one or two kinds depending on your research objectives.

The following sections provide additional details.

Ontology

Ontological analysis in IAT/ML proceeds by identifying the entities that the discourse refers to, either individual things (such as the Eiffel Tower, Nelson Mandela or the city of Amsterdam) or categories of things (such as Monument, Person or City). It also looks at the properties of these entities (such as Nelson Mandela being born in 1928 or the fact that monuments may be protected by law) and the relationships between them (such as the Eiffel Tower being located in Paris or the fact that different people may visit different monuments over time).

The ontological analysis of a text produces an ontological model, also called an ontology. An ontology provides a simplified and rigorous representation of the segment of the world that a text or a collection of texts talk about. In this regard, an ontology is very useful to express what the text is about, and to detect things that are referred to by multiple texts.

For example, consider the following text, taken from [1]:

Brennan: The New Yorker reported that when her name was first floated to you by- by Senator Daniel Patrick Moynihan, you had some misgivings.

Clinton: I don't remember that. I, first of all, there were- everybody that I know is taking credit now for- for 27 years for nominating her, but I didn't have misgivings.

In this fragment, there are clear references to some individual entities, such as The New Yorker, Senator Daniel Patrick Moynihan, and an unknown woman referred to by “she” and “her”. In addition, the text shows that Clinton, according to Brennan, had misgivings about this woman, indicating that people may have opinions about other people. All this information, together with some additional details, is captured in the ontology shown in Figure 2.

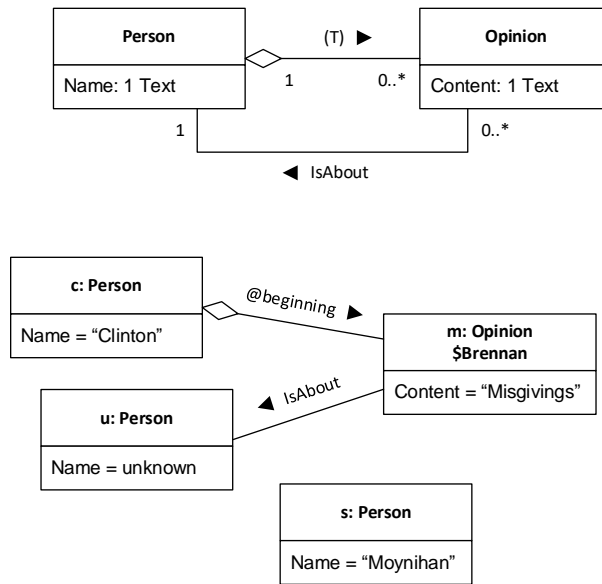


Figure 2. Ontological model for the sample fragment. Boxes represent entities. Text inside the lower sections of boxes represent properties. Connecting lines represent relationships.

Argumentation

Argumentation analysis in IAT/ML, in turn, starts by segmenting the discourse into discrete utterances made by the speakers, which are named locutions, as well as the connections between them, called transitions. The example above can be segmented as follows.

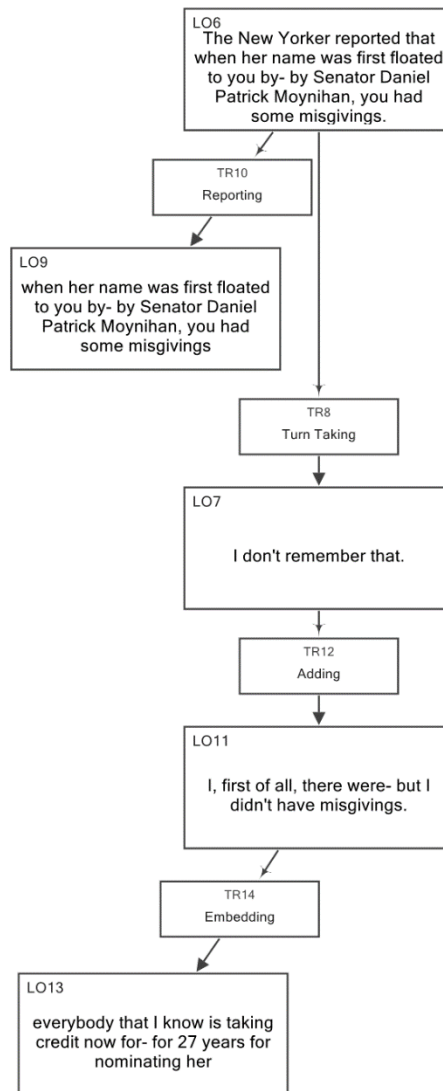


Figure 3. Sample segmentation of a discourse fragment. Large boxes represent locutions. Smaller boxes represent transitions.

Once the discourse has been segmented, each locution is reconstructed as a proposition, that is, a self-contained statement that represents faithfully the intended message. Finally, propositions are connected via argumentation relationships, which may be inferences (which connect premises and conclusions), conflicts (which connect incompatible statements) and rephrases (which connect propositions that are being recast). Locutions and propositions are connected through illocutionary forces, which describe the intention of the speaker when uttering each locution.

The final product of argumentation analysis is an argumentation model, such as that in Figure 4.

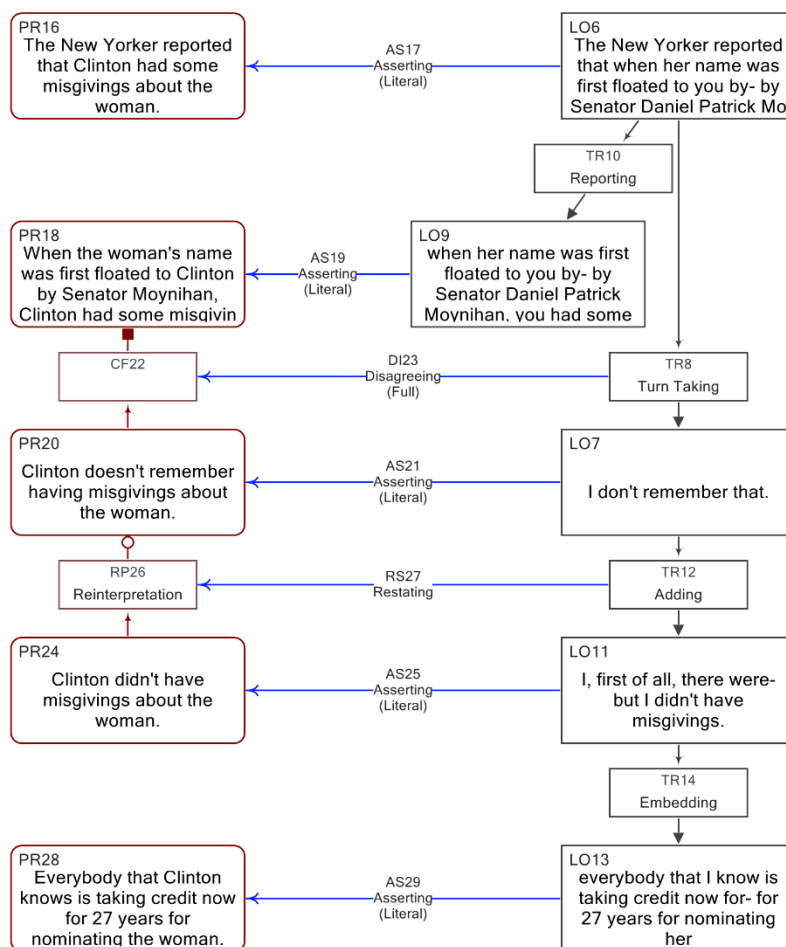


Figure 4. Argumentation model of the sample discourse fragment. In addition to locutions and transitions on the right-hand side, the left-hand side shows propositions (as large boxes) and different kinds of argumentation relationships (as smaller boxes). Illocutionary forces are depicted by horizontal arrows connecting both sides of the diagram.

Finally, you can add denotations to the argumentation model. Denotations are connections between parts of propositions and elements in the ontology, so that both models get related. For example, you may add a denotation in the previous example to show that the phrase “some misgivings” in locution LO6 refers to the ontology element “m: Opinion”.

Agency

Agency analysis in IAT/ML proceeds by asking a number of questions to the text being analysed. These questions address issues such as the themes, agents, situations and rhetoric that appear in the discourse, and are sorted from the more objective and direct to the more subjective and interpretive. Answering these questions often requires information that can be obtained from ontological and argumentation analyses.

By answering questions one by one, you can build an agency model of the discourse that provides insights into the speakers’ beliefs, desires and intentions.

For example, one typical question is “What agents or agent types are referred to by the text?”. The answer to this question is expected to produce a list of agents. A second question is “Who are the friends of each agent?”. Answering this question produces a network of “friendship” or positive relationships among agents. Although a standard question set is provided as part of IAT/ML, questions can be added, reworded or even removed depending on research objectives.

Usage Scenarios

You can use IAT/ML in a number of scenarios. IAT/ML is oriented towards discourse-oriented research, so any activity that pursues gaining a better understanding of discourse may benefit from using IAT/ML. This includes:

- Personal research projects such as doctoral or master theses
- Group research projects on discourse analysis
- Fact-checking efforts in journalism or information studies
- Studying reports, testimonies or speeches in political science, law, sociology or anthropology
- Any other need that requires a deep and contextual comprehension of what people say and why

Benefits and Limitations

IAT/ML has many benefits and some limitations, which are described in the following sections.

Benefits

The most obvious benefit of using IAT/ML is the fact that you can carry out discourse analysis from three simultaneous perspectives (ontological, argumentation and agency) in a combined manner. This is far superior to using only one of them, and produces information that would be very difficult to generate by using a conventional approach. Combining approaches not only bridges the gaps between them, but also works as methodological triangulation so that findings obtained by each perspective are validated and contrasted by the other perspectives.

A second benefit of using IAT/ML is its ability to study collections on texts as a unit. In fact, IAT/ML has been designed for the analysis of corpus, and provides a number of features for this. For example, if you are analysing a set of documents with a shared theme, you can develop a common ontological model for all of them and then one argumentation analysis for each document that relies on the shared ontology. In this manner, you can analyse the intertextual connections that may exist between documents via their references to common ontology elements.

A third benefit of using IAT/ML is the fact that LogosLink, a freely available software tool, has been developed specifically for the methodology. By using LogosLink, you can manage your corpus, carry out ontological and argumentation analysis, and produce automated analytics to support agency analysis. LogosLink can be downloaded from www.iatml.org/LogosLink. In addition, you can integrate LogosLink into your own software tool developing efforts if you wish.

Limitations

The major limitation of IAT/ML is common to all discourse analysis approaches, and is related to the fact that analysing discourse is a time-consuming manual task. In this regard, IAT/ML is not worse than other approaches. Still, you should be aware that analysing a text with IAT/ML can take a long time and need multiple iterations. Informal experience suggests that an ontology can be developed in a few hours, argumentation analysis can proceed at about 300 words per hour, and an agency analysis can be completed in a few hours as well. In any case, bear in mind that you will be slower the first few times using IAT/ML, and you will gain much speed as you get more comfortable with the methodology and the tools.

Practical Advice for Adoption

If you are considering the adoption of IAT/ML, please start by reading about it. This document is the best place to begin, but you are encouraged to go further and explore the Frequently Asked Questions (FAQ) section on the www.iatml.org website. These are updated regularly and provide simple answers to many common questions. In addition, you should have a look at the *IAT/ML Analysis Process Guidelines* document, which describes the methodology in full, also available from the same website. Finally, you are welcome to browse through the *IAT/ML Technical Specification* document if you want to know the formal details of the methodology.

In addition to reading, you are welcome to download and install LogosLink. You can find it on www.iatml.org/LogosLink. Experimenting with the tool on your computer will give you a flavour of how analysis may take place and work as a test of its suitability for you as a tool.

In particular, you may want to try out LogosLink's corpus management features, as it is likely that you will be interested in analysing collections of documents. Setting up a corpus with LogosLink is straightforward, but you can always check the User's Manual on www.iatml.org/LogosLinkDoc/UsersManual.

If you are planning to adopt IAT/ML and LogosLink in a group of people or an organisation, then you should consider some training. Experience shows that a 3-day training workshop can give you a crucial boost to start working without much of the fuss and confusion that are common when a new technology is adopted. Contact us for details on workshops tailored to your needs.

Finally, you are encouraged to carry out a small pilot project before you tackle larger or critical ones. For example, you can set up a small corpus, do some ontological, argumentation and agency analysis of a few documents, and use the results as a baseline to estimate the amount of effort that would be required to analyse your actual corpus. This will avoid surprises and allow you to plan ahead.

Alternatives

IAT/ML is not the only possible solution to the problems described in *Background and Motivation*, p. 4, although it's possibly the best one. In this section, other approaches are described, each together with some pros and cons as compared to IAT/ML.

IAT

The Inference Anchoring Theory (IAT) [4], [6] is a well-known argumentation analysis approach. IAT allows you to construct argumentation models based on very similar concepts as IAT/ML, such as locutions, propositions and illocutionary forces. IAT has a longer tradition than IAT/ML, and has been successfully used by some interesting projects. The online OVA tool (<https://arg-tech.org/index.php/ova/>) is based on IAT.

However, IAT and OVA are limited to argumentation modelling. You cannot develop ontologies or do agency analysis with IAT, and you cannot use denotations to connect argumentation to ontologies. In addition, IAT doesn't support corpus management. Use IAT if your only need is to develop argumentation models.

ConML

ConML [2], [3] (www.conml.org) is a conceptual modelling language especially designed to cater for the needs of the humanities and social sciences. By using ConML and the accompanying tool Bundt, you can develop ontologies from texts.

However, ConML is limited to ontological modelling. You cannot use it to develop argumentation or agency analysis, and there is no support for corpus management. Use ConML if your only interest is developing ontologies.

Critical Discourse Analysis

Critical Discourse Analysis (CDA) has a very long tradition and a rich body of literature. It has been used to provide critical insight into discourses of many kinds, and its applicability is beyond doubt.

Like in the previous cases, CDA provides no support for ontological or argumentation modelling, so you are limited to one kind of analysis only.

In addition, CDA has been strongly criticised for being too subjective, as most of the available guidance provides little methodological support and relies mostly on the vague sensations that texts produce on the analyst. IAT/ML mitigates this by integrating some CDA techniques with ontological and argumentation analysis, so that a grounding is provided (in the form of ontologies and argumentation models) before questions are to be answered.

RST

The Rhetorical Structure Theory (RST) [5], [7] is a descriptive approach to discourse structure and coherence relations. It has a long tradition in linguistics and has been used to explain relationships between discourse elements as well as to support operations on discourse such as text generation and summarisation.

RST is not particularly oriented towards ontological, argumentation or agency analysis, but to describing the relationships between discourse elements. It can work very well as a computational infrastructure for text processing, but perhaps not as well for discourse analysis.

Artificial Intelligence

Artificial intelligence (AI) and machine-learning (ML) methods are very popular in the fields of corpus linguistics and natural language processing. They can process large amounts of text and produce results very quickly with relatively little effort from users.

However, AI and ML approaches are heavily limited regarding what they can do. Computers do not understand the semantics, pragmatics and social context of a discourse, and despite the very sophisticated techniques that they employ, they can only work by manipulating the literal words and symbols in the text. The IAT/ML analyst, on the contrary, creates models (ontological, argumentation and agency) from the text, thus adding value and generating new knowledge, something that today's computers alone cannot achieve.

AI and ML techniques may be added in the future to assist human analysts at some specific tasks within IAT/ML and LogosLink.

No Particular Approach

Although it may sound odd, many projects tackle discourse analysis without any specific approach or theory. Texts are simply read multiple times and conclusions derived by using the analyst's intuitions, previous background and hypotheses.

Although this can work for quick and informal work, a rigorous analysis needs a cohesive and complete methodology, ideally supported by software tools. Learning IAT/ML may seem daunting at first, but the invested time and effort are easily recouped as soon as you start producing results.

Conclusion

IAT/ML is a method for the joint analysis of discourse that uses a combination of ontological, argumentation and agency perspectives. It is based on Inference Anchoring Theory (IAT) and the ConML conceptual modelling language.

As opposed to other approaches, IAT/ML is systematic and rigorous, so that you or anyone else can always trace a conclusion back to its sources and assess how well founded it is. IAT/ML is also strongly trans-disciplinary, being based on principles from linguistics, philosophy of language, argumentation theory, conceptual modelling, sociology and ontology engineering. This provides a wider and richer conceptualisation of discourse and a better-informed array of analytical techniques.

By using IAT/ML and the accompanying tool LogosLink, you can manage your corpus, carry out different kinds of analysis, produce automated analytics, and derive new knowledge from your texts that would otherwise be very difficult to obtain.

For additional information on IAT/ML, please visit www.iatml.org.

Acknowledgements

Very special thanks to the ACME team, especially Beatriz Calderón-Cerrato, Martín Pereira-Fariña and Patricia Martín-Rodilla. They are the co-creators of IAT/ML.

Special thanks to Beatriz Calderón-Cerrato for being brave enough to adopt and use IAT/ML in her work while it was still in its infancy.

Thanks also to Amelia Sanz Cabrerizo, Ana Fernández-Pampillón Cesteros, Ana García García, Ana Yáñez Vega, Chris Reed, Guillermo Rojo, Javier García Ferragud, Jean Wagemans, Kasia Budzyńska, Nacho Vidal Liy, Pablo Gamallo Otero, Teresa Abejón Peña and Victoria Escandell Vidal for their comments, tips and feedback on IAT/ML and LogosLink.

Thanks also to the staff at Incipit CSIC for their help during the ACME project, especially Felipe Criado-Boado and David Barreiro, for their insightful comments and ongoing support.

And thank you to the participants in the many discourse analysis workshops that we have taught over the years. Your questions and learning experiences are a great source of feedback for us.

Finally, thanks to the anonymous participants in the many interviews and focus groups that were carried out to gather discourses to be analysed with IAT/ML.

References

- [1] CBS News, “Transcript: President Bill Clinton on ‘Face the Nation,’ September 20, 2020,” *CBS News*, Sep. 20, 2020. <https://www.cbsnews.com/news/transcript-president-bill-clinton-on-face-the-nation-september-20-2020/> (accessed Apr. 06, 2022).
- [2] C. Gonzalez-Perez, “A 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, 2012, pp. 396–401.
- [3] Incipit CSIC, “ConML Technical Specification,” Incipit CSIC, 2020. [Online]. Available: <http://www.conml.org/Resources/TechSpec.aspx>.
- [4] M. Janier, M. Aakhus, K. Budzynska, and C. Reed, “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. Mohamed and M. Lewinski, Eds. College Publications, 2016.
- [5] W. C. Mann and S. A. Thompson, “Rhetorical Structure Theory: Toward a functional theory of text organization,” *Text - Interdisciplinary Journal for the Study of Discourse*, vol. 8, no. 3, 1988, doi: 10.1515/text.1.1988.8.3.243.
- [6] C. Reed and K. Budzynska, “How Dialogues Create Arguments,” in *ISSA Proceedings 2010*, 2010, [Online]. Available: <http://rozenbergquarterly.com/issa-proceedings-2010-how-dialogues-create-arguments/>.
- [7] M. Taboada and W. C. Mann, “Rhetorical Structure Theory: looking back and moving ahead,” *Discourse Stud*, vol. 8, no. 3, pp. 423–459, Jun. 2006, doi: 10.1177/1461445606061881.