Quantitative Archaeology and Archaeological Modelling

Cesar Gonzalez-Perez
Patricia Martin-Rodilla
Martín Pereira-Fariña *Editors*

Discourse and Argumentation in Archaeology: Conceptual and Computational Approaches

Springer

Quantitative Archaeology and Archaeological Modelling

Quantitative approaches and modelling techniques have played an increasingly significant role in archaeology over the last few decades, as can be seen both by their prominence in published research and in university courses. Despite this popularity, there remains only a limited number of book-length treatments in archaeology on these subjects (with the exception perhaps being general-purpose GIS). 'Quantitative Archaeology and Archaeological Modelling' is a book series that therefore responds to this need for (a) basic, methodologically transparent, manuals for teaching at all levels, (b) good practice guides with a series of reproducible case studies, and (c) higher-level extended discussions of bleeding edge problems. This series is also intended to be interdisciplinary in the analytical theory and method it fosters, international in its scope, datasets, contributors and audience, and open to both deliberately novel and well-established approaches.

Cesar Gonzalez-Perez • Patricia Martin-Rodilla •
Martín Pereira-Fariña
Editors

Discourse and Argumentation in Archaeology: Conceptual and Computational Approaches

Springer

Editors
Cesar Gonzalez-Perez
Incipit, CSIC
Santiago de Compostela, Spain

Martín Pereira-Fariña
Facultade de Filosofía
University of Santiago de Compostela
Santiago de Compostela, Spain

Patricia Martin-Rodilla
Facultade de Informática
Universidade de A Coruña
A Coruna, Spain

ISSN 2366-5998 ISSN 2366-6005 (electronic)
Quantitative Archaeology and Archaeological Modelling
ISBN 978-3-031-37155-4 ISBN 978-3-031-37156-1 (eBook)
https://doi.org/10.1007/978-3-031-37156-1

This work was supported by COST Action "Saving European Archaeology from the Digital Dark Age" (SEADDA), CA 18128, https://www.cost.eu/actions/CA18128/ (CA 18128) and by grant PID2020-114758RB-I00 funded by MCIN/AEI/10.13039/501100011033

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Paper in this product is recyclable.

Foreword

How do archaeologists come to establish and validate accounts of the past, based on their encounters with its material remains as mediated by fieldwork, collections, and years of study and toil? How do they justify claims they make, and on what grounds do they accept or reject claims made by others? How do they reach good decisions as they investigate, construct, curate, and communicate the archaeological record? What are archaeological facts, and how do they come to be accepted as such? What are the traits of sound archaeological syllogisms? And, more generally, what is archaeological knowledge? Where can we find it, and in which forms does it manifest itself? How can it be captured, represented, and analyzed? How is it communicated, debated, and evaluated? Is there "good" and "bad" archaeological knowledge, and how can we tell them apart? Which factors are at play in knowledge-making, and in knowing? What are the implications and stakes of archaeological knowledge, and the ways it comes into being?

Few archaeologists spend much time reflecting directly on this Pandora's box of vexing questions. Yet many of them, prompted by engaging with the trans-disciplinary perspectives in this exciting volume on the use of computational approaches to discourse and argument analysis in archaeology, are central to methodological aspects of archaeological research, and to the acquisition of archaeological expertise. For one thing, competent archaeologists should surely be able to reason on the validity of an archaeological study in their area of expertise, and, beyond that, to produce research findings substantiated by persuasive arguments, supported by reliable evidence, and consonant with accepted knowledge in their field. On the other hand, scholars of archaeological theory, as well as those concerned with policies, decision-making, and interventions related to the preservation of archaeological heritage, its multiple and often conflicting socioeconomic, cultural and symbolic uses, and the future of archaeological work, need also to grapple routinely with questions related to the factors under which archaeological knowledge is produced, the felicity conditions under which archaeological facts can be deemed to be acceptable, and the status, impact, and repercussions of resulting knowledge for contemporary societies. In almost all aspects of archaeological work, researchers and professionals are inevitably entwined in knowledge-laden activities,

as they engage with the body of scholarship in their area of expertise; as they identify research topics and questions; as they collect, represent, and analyze evidence from archaeological fieldwork and collections; as they develop identifications, classifications, descriptions, explanations, and, more generally, accounts of the material record of humanity and its implications for past societies and cultures; as they produce archaeological reports, catalogs, databases, monographs, articles, and conference papers; as they debate and come to conclusions on the validity of research ideas and findings, and on deliberations on the management and use of archaeological heritage, be it in scholarly publications, administrative and policy venues, or in informal interpersonal settings including online communications; and, last but not least, as they address the historical and contemporary misuses of archaeology by political and state actors, the appropriation of research agendas and heritage policies by dominant ideologies and sectarian and economic interests, and of archaeologically manifested phenomena by sensationalism, pseudo-science, and irrationalism.

We might assert, paraphrasing Bruno Latour, that archaeology, not unlike experimental science, "has two faces: one that knows, and one that does not yet." The latter is of relevance here. It offers a view of the discipline not as "readymade science" with its middle-range theories and accounts of particular sites, cultures, periods, artifact types, etc., but as a "science in the making": a domain where archaeological knowledge, as an object (manifested in the representations of ideas in texts, visual representations, data structures, and the like), is examined in its articulation with archaeological *knowing* or *knowledge-making* as an activity, ripe with "uncertainty, people at work, decisions, competition, controversies." It is precisely in this domain of archaeological activity where the Pandora's box of our initial questions is primarily located.

Studying how archaeologists establish ideas, facts, and assertions from their encounters with the material remains of the past, from the translation of the material record of features and finds in the field into an informational record made of descriptions, data points, visualizations, enmeshed with identifications of sites, archaeological contexts, artifacts, types and assemblages in the excavation report, and further developed into typologies, seriations, and other manifestations or archaeological systematics, as well as into synthetic accounts and interpretations, explanations, and theories in scholarly publication, has been a fruitful way to approach archaeology "in the making." From publications such as Mike Edgeworth's fascinating ethnography of the "acts of discovery" in an unnamed excavation in Britain, to the fertile qualitative investigations of diverse aspects of archaeological information work in northern Europe by Isto Huvila, and the multi-sited study of archaeological curation across different stages in the formation of four North American archaeological collections in Sarah Buchanan's insightful doctoral dissertation, the study of archaeological practices and knowledge work has emerged as the pursuit of an growing trans-disciplinary community of researchers concerned with making sense of the agents, processes, settings, mediating tools, and objects of archaeology "in the making."

A central aspect of "archaeology in the making" concerns how archaeological data, facts, and assertions related to them are represented in different genres of representations, and how such representations – from descriptive records, lists, and catalogs to research publications – underlie different modes of archaeological knowledge production. As I argued in an earlier manuscript (Dallas, 2016), we owe a seminal, and perhaps the first, systematic attempt toward a theorization of these questions to the still under-appreciated intellectual contribution of French Classical archaeologist and information scientist Jean-Claude Gardin. A pioneer of computational analysis in archaeology in the 1950s, he was initially preoccupied with the development of analytical "codes" or vocabularies for the formal description and classification of archaeological artifacts, culminating into the development of his Syntol free structure indexing language, a means for representing the content of documents through n-place predicates expressible in a machine language. Drawing critically from fields as diverse as documentation, classification theory, material culture studies, structural linguistics, argumentation theory, and philosophy of science, in his "Document analysis and linguistic theory" (1973), Gardin then expands his earlier attempts to account for the intellectual content of archaeological documents through term indexing by an added emphasis on their syntax and semantics, noting that "the boundary between syntax and semantics becomes so fuzzy that it is not possible any more to regard syntax as independent nor to confine semantics to an interpretative function."

This is the foundation of Jean-Claude Gardin's seminal contribution to the theory of archaeological argumentation and discourse, translated into English as *Archaeological constructs: an aspect of theoretical archaeology* (1980). The book is a formidable theoretical construct in its own right. In the first chapter, it outlines Gardin's "iterative model" linking the acquisition of archaeological materials with their annotation and consequent generation of propositions, and offers examples of what he calls a "logicist analysis" of processes of cataloging, classification, pattern recognition, and historical inference that constitute the "lifecycle" of archaeological knowledge process. He then goes on to analyze processes relevant to the construction of two very different kinds of archaeological publications: "compilations," such as finds catalogs or excavation reports, typically concerned with material remains of the past and their attributes, and "explanations," such as synthetic monographs and interpretative accounts of ancient societies, their history, and mode of life. In his analysis, he castigates the failure of traditional archaeological publication in the narrative genre to attend to methodological rigor, theoretical frugality, and clarity, even often violating sound reasoning. As an alternative, he advocates the "condensation" of archaeological scholarly prose through a process of schematization, taking the form of an ordered tree of logical inferences using *modus ponens,* and operating on a lexicon of structures of symbols representing propositions – in other words, an inference tree.

But then, Gardin adds the following qualification: "I am not proposing a new handbook on archaeological theory, from which students can learn the techniques of observation and interpretation [. . .] my goal is an analysis of the mental operations carried out in archaeological constructions of all sorts, from the collecting of data to

the writing of an article or book in published form." While his action-oriented, even polemical, advocacy of a mode of archaeological communication based on formal reasoning is undeniable, he notably advances also a salient approach to representing and understanding the way actual archaeological argument unfolds in practice: a way to make archaeologists "more aware of the empirical or social limits of our interpretations" – what he calls "a practical epistemology" of archaeological knowledge. Adopting Stephen Toulmin's criterion of "reasonableness," he advocates an archaeology whose propositions and theories, as represented in its publication practices, stand the test of reason, but also intends his logicist schematization as a means to "to gain a deeper understanding of what our interpretive writings 'are', as symbolic constructs; we also wish to evaluate what those constructs can 'do', in the universe of discourse under study."

The most notable methodological contribution of Gardin's theorization of archaeological argumentation concerns archaeological publication. His method of re-expressing traditional archaeological argument in terms of a lexicon of symbols and a set of argumentation operations has been adopted by a limited number of studies. Among them, ethnoarchaeologist Valentine Roux's Arkeotek project goes beyond logicist schematization to address the interdependence between archaeological data constitution on the one hand and scholarly argumentation on the other. Its hypertext-based "Scientific Constructs and Data" model provides for integrating archaeological argumentation structure with descriptive archaeological data. Further work demonstrates the possibility of modeling the logicist schema of scholarly reasoning as a formal ontology. In a parallel development, the UK Archaeology Data Service's *Internet Archaeology* journal featured, as early as 1997, a similar ability of offering interactive access to archaeological studies that allowed simultaneous access to scholarly claims and supporting data: a non-lasting experiment which, nevertheless, still goes beyond the current state-of-play in research data publication. Such attention to the structure and content of archaeological scholarly communication, and its reliance on the propositional content and structure of publications, is self-evidently justified on pragmatic reasons of allowing better access to and evaluation of claims made by archaeological research.

Yet, dealing with argumentation and discourse in archaeology makes the case for accounting, beyond methodology, for ontological, epistemological, and axiological considerations. In other words, when we consider archaeological knowledge "in the making" as a worthy subject of study, we need to decide on questions of existence, knowledge, and values. As regards ontology, most archaeologists would agree that their domain of reference – including material remains of past human activity and past people – exists, or has existed, independent from our knowledge of it; that it consists of differentiated objects and structures – be it natural or social – which have powers and ways of acting that contribute to the production of events; that apart from actual objects accessible directly to experience, this external world is also composed of latent, underlying entities and relations between observable entities, yet such relations may be contingent rather than necessary; but also that, unlike natural objects, social particulars such as a specific action, an artifact, or an archaeological

culture are dependent also on categories accessible only within our own interpretive frame, even if we still admit that they exist regardless of our specific interpretation of them. At the epistemological level, on the other hand, many (but not all) workers in the field will admit that archaeological knowledge is theory-laden, socially constructed, and historically situated; therefore, what we accept as true today may be falsified tomorrow, and "thought collectives" (to use Ludvik Fleck's useful term) adopting different theoretical premises may legitimately have conflicting views of what constitutes knowledge on a given subject; that there are both continuities and discontinuities in the evolution of archaeological knowledge; and that the production of archaeological knowledge is a social practice, and therefore social relations, context, and interests, as well as the ways in which archaeological knowledge is communicated (typically, through historically sanctioned genres of information carriers), influence its content. Finally, at the axiological level, most archaeologists would adhere to the idea that archaeological research should be critical of its object of inquiry, and that the understanding of archaeological phenomena entails viewing them critically; some would also add that archaeological practice should be emancipatory, and adhere to values of social justice and an ethics of care.

Readers with an interest in the philosophy of science may recognize that this set of ontological, epistemological, and axiological positions is aligned with a critical realist account of the human sciences (and, in fact, derived directly from Andrew Sayer's account of critical realist assumptions): a transcendental realist ontology, a constructivist epistemology, and a value-laden, reflexive axiology. In tandem, a critical realist account conceives the process of archaeological explanation – one common objective of archaeological argumentation – as consisting of the identification of some past human activity or phenomenon to be explained and its resolution into elements, re-description of these elements in the theoretical language of archaeology (or the approach to archaeology espoused by the researcher), a retroductive attempt to describe the likely structural conditions (such as causal mechanisms, material-semiotic rules, etc.) and tendencies involved, and, finally, a process of elimination of alternative causes, or explanations. Of course, not all archaeological research aims at explanation: in fact, the reliance of archaeological knowledge related to social aspects of past reality on categories (kinds) that can only become accessible through human cognition – those which, in a more clearly constructivist vein, have been called "interactive kinds" by philosopher Ian Hacking – on the shared scholarly language of the epistemic community in which an archaeological study is situated, makes it clear that words used for identification or assignment of properties of archaeological entities have consequences on the content of archaeological knowledge. In other words, far from being the result of menial or mechanical work with limited value as knowledge, archaeological descriptions, such as those found in field recording sheets and collections databases, do matter.

This has an interesting implication on what we consider as the scope of archaeological argumentation. Clearly, a causal syllogism connecting an archaeological phenomenon to likely causes, or a justification provided for some intervention concerning the protection and use of an aspect of the archaeological heritage,

belongs within the purview of argumentation. But what about a finds database? What about the identification of some archaeological feature, its assignment to some particular function, provenience, or cultural period, in a catalog without explanatory aspirations? What about the broad range of visualizations often included as part of archaeological reports and publications? What of the illustrations – figures, photographs, diagrams, models – often accompanying archaeological texts? Are we to assume that they play no role in archaeological argumentation, and, if so, that they are not involved in knowledge production?

The last statement points to an interesting conundrum: pragmatically, the very inclusion of visualizations and illustrations within archaeological documents indicates that they contribute to knowledge production. If we were to accept that they do not participate in argumentation, then we would need to posit other rhetorical modes of archaeological knowledge beyond argumentation. But, in fact, it should not surprise us that no archaeological document consists solely of propositions linked together to form an argumentation structure. The most lucid exposition (pun intended) of this is provided by Gavin Lucas in his recent *Writing the Past* monograph, where he demonstrates how *argument* not only co=exists but in fact cooperates in the very same text toward the archaeological knowledge construction with instances of all three alternative rhetorical modes systematized as early as the nineteenth century in the context of rhetoric and composition studies: *narrative,* presenting a story unfolding through time through the involvement of actors and events; *description,* involving the presentation of qualities and attributes of some observed object or event; and, *exposition,* explaining or clarifying a topic or issue.

How, then, different archaeological communication objects mobilize different rhetorical modes, and how they are articulated in reports and publications to construct archaeological knowledge, is a fascinating topic. Going beyond rhetorical modes, the example of archaeological visualizations which I had the opportunity to reflect upon a few years ago in an interesting conference session on "Visualization as analysis in archaeology," which provides good insights on how a site section and "hermeneutic matrix" diagram may act as an exposition of the temporality and longevity of each excavation cut; or, how a dynamic virtual reconstruction of the Antikythera mechanism captures performative knowledge, and supports a plausible explanation, about the function of the mechanism; and, more generally, how archaeological visualization constitutes an objectual epistemic practice rather than being merely an act of display; and an archaeological 3D visualization can act as an "epistemic contract" (borrowing Harold Garfinkel's identification of the transcript of an outpatient clinic interview as "therapeutic contract" rather than as "actuarial record"), made to support the generation of knowledge claims in further steps of the interpretation ladder, rather than to represent faithfully "what the sensor saw."

This edited volume is not an archaeological study. It is, rather, a collective work *about* archaeology as a field of knowledge and as a practice of knowledge-making. It offers a shared foundation useful to archaeologists curious about the conditions of archaeological knowledge production and the potential of computational approaches for opening new paths for communicating and validating archaeological research,

computer scientists from the fields of natural language processing and argumentation support, information researchers interested in archaeological practices and knowledge work, anthropologists and sociologists of science, and others interested in how archaeologists produce knowledge through argumentation "in use." In the spirit of the agonistic nature of argument, the volume accommodates diverse, and in some cases dissonant, conceptualizations and computational approaches to argumentation and discourse, ranging from archaeological to computational, from normative to accommodative, from pragmatic to illustrative, from synthetic to highly focused, and from instrumental to critical. It provides useful insights, and stimulates ample reflection toward new questions. It is unique in combining critical and theoretical accounts of archaeological discourse and knowledge work, and overviews of key computational approaches to discourse and argument analysis, with examples of specific applications to the formal representation of archaeological knowledge, ranging from the identification of topics through computer-assisted recognition of historical names and common descriptors, to formal conceptualizations that allow the articulation between the domain of archaeological discourse which archaeological texts inhabit, and the domain of past human activity which such texts refer to.

Reiterating the core thesis he originally advanced in *The Uses of Argument*, Stephen Toulmin admits to "a single, deeply held conviction: that, in science and philosophy alike, [people] demonstrate their rationality not by ordering their concepts and beliefs in tidy formal structures, but by their preparedness to respond to novel situations with open minds—acknowledging the shortcomings of their former procedures and moving beyond them. Here again, the key notions are 'adaptation' and 'demand', rather than 'form' and 'validity'." In a similar vein, the dynamic nature, historicity, and pragmatic situatedness of archaeological argumentation are acknowledged across this volume. In diverse ways, different chapters address the content of archaeological argumentation, offer methods and examples to identify its subject-matter computationally and to represent formally its logical and procedural structure, and offer insights on the conditions under which particular claims are (and should be) accepted. They account for the reliance of archaeological argumentation on communicative processes, set in motion by archaeologists in conversational semiotic activity governed by historically situated systems of signification. Furthermore, they also engage with the dependence of archaeological discourse on reference to "things-in-the-world" – empirically manifested aspects of the archaeological record, persons and collectivities, objects, places, and events, as well as conceptual entities comprising the subject-matter of arguments. Finally, they illustrate how discourse "in use" hinges on the pragmatic dimensions of archaeological knowledge work – affiliation to thought collectives (to use Ludvik Fleck's salient notion) and communities of interest with their shared communicative codes and accepted knowledge, presuppositions, norms, motivations, affects, and future stakes – which underpin the discursive activity of archaeologists as they respond and adapt to a changing field of epistemic, ethical, political, socioeconomic, and cultural challenges. Reaching beyond epistemological, methodological, and axiological considerations on the nature, poetics, and politics of archaeological

knowledge, argumentation, and discourse, which have been the focus of numerous earlier contributions (from Jean-Claude Gardin to Alison Wylie, Rosemary Joyce, and Gavin Lucas, to name but a few), this volume provides a pragmatically useful body of knowledge on the relevance, critical context, methods, and practical applications of discourse and argument analysis technologies as tools to represent, analyze, and reflect on archaeological knowledge and its production, aptly demonstrated through salient case studies of computational approaches.

At a time when the representation of the archaeological record and the production of archaeological knowledge is increasingly mediated by digital research infrastructures and associated standards, tools, and procedures, and when the promises of deep learning and artificial intelligence assume renewed impetus across the disciplines, the task of understanding archaeological discourse and argumentation as knowledge work becomes an urgent undertaking. This volume addresses consequential issues and offers examples of promising computational approaches for *representing* the dynamic structure and situated process of archaeological argument, and its discursive and pragmatic underpinnings in past and contemporary realities. It opens important additional questions, contributing to the emergence of an important interdisciplinary subfield bridging archaeological theory and method with computational approaches to meaning and argument analysis. Most importantly, it also provides a springboard for *intervening,* by mobilizing the archaeological community to act toward the use of computational technologies to enable reflexive, critically informed, and relevant approaches to the production, publication, epistemic validation, and use of archaeological knowledge, adapted to the demands and challenges facing contemporary societies, and the planet.

Faculty of Information, University of Toronto, Toronto, Canada Costis Dallas

Reference

Dallas, C. (2016). Jean-Claude Gardin on archaeological data, representation and knowledge: Implications for digital archaeology. *Journal of Archaeological Method and Theory*, *23*(1), 305–330. https://doi.org/10.1007/s10816-015-9241-3

Preface

Most of the knowledge that we produce in archaeology comes from careful argumentation from basic premises to elaborate conclusions. Initial premises include descriptions of finds, features, sites, and landscapes, while conclusions range from settlement patterns to trade routes or social organisations. In this regard, most archaeological texts constitute discourses aiming to persuade the reader to accept a series of conclusions based on some initial premises, often factual and evidentially supported. Whether or not an archaeological text is capable of persuading its readers and thus advance the state of the art in the field depends on the quality of the chosen premises as well as the robustness of the subsequent argumentation. Therefore, paying attention to discourse and argumentation in archaeology constitutes a crucial aspect of meta-research.

Language technologies have evolved rapidly over the last 10 years, and today we can process natural language on a computer with relative ease, at least for some well-defined purposes. The conceptualisation of discourse and argumentation has advanced significantly as well, together with applied approaches. Although the importance of discourse and language in archaeology has been pointed out by many authors, there is no comprehensive work to date that presents a panoramic view of argumentation and discourse approaches and technologies in archaeology. In this book, we aim to provide this.

Audience and Objectives

This book is aimed at archaeologists with an interest in language, discourse, and argumentation, and specifically on how archaeological conclusions are obtained through argumentation processes. In particular, researchers in archaeology can find the book useful to gain a better understanding on how argumentation can take us from premises to conclusions and learn how to do it better. Lecturers and students of archaeology can use the book to learn specific conceptual approaches and

computational approaches to discourse and argumentation analysis for archaeological texts.

All in all, the book aims to provide a comprehensive overview of conceptual approaches and computational techniques for argument analysis in archaeology. It does so by building slowly from scratch, starting with introductory topics and progressing towards advanced and more specialised issues. Also, the book unites theory and practice, providing a comprehensive panorama of conceptual approaches and computational techniques.

The book starts with the basic foundations of discourse and argumentation analysis, introducing the main goals of discourse analysis, presenting different approaches to what an argument is, and concluding with cutting-edge and state-of-the-art technologies for the fully automatic analysis of texts. In addition, the book tackles different contexts where archaeological discourses are found, from data collected during fieldwork to archiving of excavation reports or court resolutions on heritage-listed items.

The book also presents an updated review of approaches and methods related to natural language processing and text mining that are applicable to archaeological settings, and at multiple linguistic levels such as lexical, grammatical, and discursive. Also, the book proposes some methodological approaches for the analysis of argumentative strengths and weaknesses in archaeological texts based on Toulmin's schemes.

Finally, the book considers different degrees of formalisation in discourse analysis, from critical Foucauldian approaches to the more quantitative computational analytics, and takes into account the social dimension of archaeological discourse production.

Book Structure

This book is organised into two major sections: Conceptual Approaches and Computational Techniques. A preface provides a general introduction, and a final chapter offers some speculations as to what the future of discourse and argumentation in archaeology may look like.

The first section, Conceptual Approaches, contains a collection of contributions from different foundations and perspectives, offering a comprehensive overview of the discursive and argumentative phenomenon in archaeology and its ramifications. In Chap. 1, Martín Pereira-Fariña presents the fundamentals principles of discourse analysis and three different theoretical approaches of how arguments can be represented, summarising the process to transform raw data into an annotated corpus that allows us to draw conclusions anchored in how language is used in context. In Chap. 2, Stephen Stead deals with the issue of documenting the argumentation in a discourse so that it can interoperate with other sets of data. In Chap. 3, Michael E. Smith offers a historical journey through different stages and degrees of importance attributed to the study of archaeological argumentation, analyses some reasons for

the low level of attention that is paid to argumentation in archaeology today, and presents a methodological proposal based on argument strengths and weaknesses. In Chap. 4, Alejandro Sobrino and Beatriz Calderón introduce a theoretical framework for the analysis of causal linguistic structures related to culturally relevant elements, acknowledging that causality can be linguistically expressed in multiple ways, and showing how this issue can be tackled.

In Chap. 5, in turn, Cesar Gonzalez-Perez focuses on what archaeological texts talk about and presents an approach to connect the argumentation in the discourse with the underlying ontological elements in the world, using a referential device named ontological proxies. In Chap. 6, Isto Huvila takes on a more sociological, anthropological, and critical nature to archaeological discourse and reflects on discourses in archaeology as situated in their social context of production, including an analysis on the role of different agents and the impact of new ways of discourse production such as social networks or other techno-mediated mechanisms. In Chap. 7, Cesar Gonzalez-Perez, Martín Pereira-Fariña, Patricia Martín-Rodilla, and Leticia Tobalina tackle the issue of vagueness in archaeological discourses and present a conceptual framework to capture and manage vague information from the field to the text. Finally, in Chap. 8, Jeremy Huggett uses a multimodal approach to extend discourse analysis in archaeology beyond the mere text.

The second section, Computational Techniques, provides a sample of some algorithmic approaches that have proved useful to deal with discourse and argumentation in archaeology. In Chap. 9, Patricia Martín-Rodilla offers an introductory overview of how computer-based processing of natural language has been applied to archaeological texts, and what major lines of work exist today. In Chap. 10, Holly Wright, Tim Evans, and Katie Green deal with the natural language processing of lexicon in archaeological texts from the perspective of a large digital archive, showing how these techniques are useful for information extraction for researchers. In Chap. 11, Alex Brandsen deals with text mining at the lexical, grammatical, and discursive levels, as well as machine learning applied to archaeological texts. In Chap. 12, John Lawrence, Martín Pereira-Fariña, and Jacky Visser go beyond the discourse itself to explore the mining and analysis of arguments from plain text, with a special focus on argument analytics and result dissemination. Lastly, in Chap. 13, Maria Elena Castiello provides an approach to processing the vagueness that is inherent to archaeological language in a site modelling context.

For those readers who have a special interest in a particular topic, the book admits a theme-oriented reading in addition to a linear sequence of chapters. Chapters 2, 4, and 3 in Part I, as well as Chap. 12 in Part II, deal with argumentation and different approaches to understanding how people argue to defend their standpoints. Chapters 5 and 7 in Part I, as well as Chaps. 9, 10, and 11 in Part II, deal with lexical,

grammatical, and semantic language processing. Finally, Chaps. 4, 6, 7 and 8 in Section I, as well as Chap. 13 in Part II, deal with language as used in context, including social aspects, vagueness, and multi-modality.

Enjoy reading!

Santiago de Compostela, Spain — Cesar Gonzalez-Perez
A Coruna, Spain — Patricia Martín-Rodilla
Santiago de Compostela, Spain — Martín Pereira-Fariña

Acknowledgements

The editors wish to thank the authors of the chapters of this book for their generous contributions, as well as the Springer staff who guided and helped us throughout the publication process.

The editors must acknowledge the contributions and support of the following grants towards the preparation of this book: project "Heritage 3.0: Argumentation and Conceptual Modelling for Enhanced Cultural Heritage Participation and Management Policies" (ACME), grant number PID2020-114758RB-I00 funded by MCIN/AEI/10.13039/501100011033; project "Deflationist Views in Ontology and Metaontology", grant number PID2020-115482GB-I00 funded by MCIN/AEI/10.13039/501100011033; project "Saving European Archaeology from the Digital Dark Age" (SEADDA), grant number CA18128 funded by EC COST Actions; and *Consellería de Educación, Universidade e Formación Profesional* (accreditation 2019-2022 ED431G/01, ED431B 2019/03); and European Regional Development Fund, which acknowledges the CITIC Research Centre in ICT at the University of A Coruña as a member of the Galician University System.

Contents

Chapter 1
Introduction to Discourse Analysis and Argumentation Theory

Martín Pereira-Fariña

Abstract Discourses analysis is an explicit and systematic study of the structures, strategies and manoeuvres of texts or talks in terms of a given theoretical framework. The current stage of computational technologies allows us to tackle this task from different perspectives. Along this chapter, I explore how an argument can be characterised and analysed from three theoretical perspectives (logic, pragmatic and cognitive). Each of these approaches lead us to different types of discourse analysis, emphasizing different angles of the same text, which shows the richness of this analytical framework. After that, I describe the main steps for transforming raw text into an annotated corpus, essential to draw any reliable conclusions from it. Annotation is a complex task, essential for a good quality analysis of discourse, but it can be split into doable steps. Finally, the chapter concludes with some ideas for the exploitation of these results and how they can be disseminated.

Keywords Argumentation · Annotation · Corpus creation · Discourse analysis · Ontology

1.1 Introduction

It is 4 pm on a cold day in February. Two senior archaeologists are discussing about the future of The Cave of Altamira,[1] a set of charcoal drawings and polychrome paintings that constitute one of the firsts masterpieces in the history of mankind. Sitting in front of each other, together with a moderator, they debate a question

[1] World Heritage Site by UNESCO located in Santillana del Mar (Cantabria), North of Spain. For more information: http://www.culturaydeporte.gob.es/mnaltamira/en/home.html

M. Pereira-Fariña (✉)
Department of Philosophy, University of Santiago de Compostela, Santiago de Compostela, Spain
e-mail: martin.pereira@usc.gal

C. Gonzalez-Perez et al. (eds.), *Discourse and Argumentation in Archaeology: Conceptual and Computational Approaches*, Quantitative Archaeology and Archaeological Modelling, https://doi.org/10.1007/978-3-031-37156-1_1

that has been floating around Altamira during the last 20 years: should be the Cave opened for public access or just for experts with research purposes? Experts have mixed opinions:

- **Researcher 1 (R1)**: *I think the question is basically that the Cave of Altamira should be opened because it is obviously a place, to say simply, that everyone has the right to visit. Thus, from that principle, I think that all that can be negotiated, discussed, and talked about, it's under what conditions and, above all, for what, that is, what is the benefit of opening it, right? Starting from the principle that heritage, everyone has the right to access it, then, considering the problematic and risky conditions that the cave has, from there one can think of possible restrictions and criteria to restrict and so on, but from the outset, I mean, it must be open. That is my position.*
- **Researcher 2 (R2)**: *I think not, precisely for the same reason; because everyone has the right to access heritage, but if everyone accesses heritage, that heritage is destroyed, isn't it? So, I think that access should be restricted to experts, let's say, and to researchers, and in fact I think that there was even a more or less exact reproduction and, well, that for tourism purposes or just for dissemination, I think that it could therefore do the job quite well. And basically, that is my point. If you want, we can go into more detail, but...*
- **Moderator (M):** *Only experts, you say, should access to the cave.*
- **Researcher 2 (R2):** *Yes, indeed, just researchers. Let's say, people who are in research centres and it is essential, come on, that they enter to check certain things, for example... I don't know.*

We can easily appreciate how each position is argued and how a certain common background is presupposed. So, how are both discourses elaborated? Are they talking about the cave as a physical object or a social object? What type of reasons do they use for supporting their corresponding positions? What is the connection between the researchers and the cave? Are they considering themselves as experts or as regular visitors? Could they have any potential conflict of interest? What is the context in which this debate is happening?

This is just not a scientific debate, but it also has a social impact. Understanding and evaluating it requires to unpack the connections between language, reality, and speakers. Discourse analysis tackles precisely this question, as it is defined by Paltridge (2012, p. 2):

> Discourse analysis examines patterns of language across texts and considers the relationship between language and the social and cultural contexts in which is used. Discourse analysis also considers the ways that the use of language presents different views of the world and different understandings. It examines the use of language is influenced by relationships between participants as well as the effects the use of language has upon social identities and relations. It also considers how views of the world, and identities, are constructed through the use of discourse.

Therefore, discourse analysis does not study the language itself, but the language in use (Gee, 2011). It is a broad and interdisciplinary field, connected with other disciplines such as semiotics (Eco, 1979), linguistics (Serrano, 1983) or communication

studies (Chandler, 2003). There are two main methodological approaches which are differentiated by their respective goals (Gee, 2011):

- Descriptive: It aims to understand how language works in different communicative situations: what are the topics of discussion, what is the grammar applied to produce meaning, what are the different stylist resources of manoeuvres to produce meaning, etc.
- Critical: It aims to intervene in social, political or cultural problems and controversies and provoke changes in the world based on studying how language works.

In this chapter, I will focus on a descriptive type of discourse analysis which main goal is to unpack argument structures in a given discourse and, eventually, the folk ontology underpinning a specific discourse. The goal is to provide robust theoretical and methodological grounds for understanding the different methods of reasoning and how knowledge is produced in a field such as cultural heritage (Lucas, 2019). Discourse analysis, following the spirit of archaeological stratigraphy, allows us to identify how the different layers that constitute the structures of meaning (what are the internal elements of the text and how they are organised) and interaction (how speakers take part in the discourse).

This chapter is organised as follows. Section *What is an argument?* Introduce the three different views about the notion of argument. Section *Designing and Annotation Campaign* describes how to design and carry out a concrete study in discourse analysis. Section *Exploiting the results* provides some ideas for the communication and dissemination of the results extracted from the study. Finally, this chapter concludes with some reflections on the impact that discourse analysis can have in the field of archaeology and cultural heritage studies.

1.2 What Is an Argument?

R1 says *"I think the question is basically that the Cave of Altamira should be opened"* while R2 replies *"I think that access should be restricted to experts"*. Both speakers maintain opposed positions with respect to the Cave, which can be reconstructed as assertions as follows:

- **R1**: *The access to the Cave should be opened to everybody.*
- **R2**: *The access to the Cave should be restricted to experts.*

These are not arguments; these are just assertions. So, what does we need to have an argument? Generally speaking, an argument requires at least another assertion that play the role of support. A more specific definition is highly dependent on how the relationship between both statements is conceptualized. Next, I will describe three different approaches:

- Logical approach: An argument is a sort of linguistic entity where a statement, named conclusion, is supported by one or more statements, named premises (Salmon, 1984). Logic studies the connection between premises and conclusion in order to determine when it is correct and when is not; i.e., the rules and principles to determine the validity of the argument.
- Pragmatic approach: An argument is a particular type of speech act where a speaker has the intention to support a specific statement, the conclusion, by means of another statement or a set of them, the premises (Janier & Reed, 2017). Argument theories based on Speech Act Theory (Austin, 1989; Searle, 1965) aims to identify when a speaker intends to make an argument; determining its validity is a secondary issue.
- Cognitive approach: An argument is a cognitive category where the linguistic expression is acting as a sign vehicle of a specific relation of support between two or more mental representations where one is the conclusion and the other or others are the premises (van den Hoven, 2015; Searle, 1965). This is the most ambitious approach, because it aims both to identify the structure of the argument and its strength. The key point of strength is not logical validity but acceptability; i.e., whether the argument has convinced its addressee or not (Mercier, 2012).

In the following subsections, I will describe in detail each of these views and I will illustrate how they can be diagrammatically represented for their computational analysis.

1.2.1 Logical Approach

The logical analysis of a discourse fragment entails three basic steps (Salmon, 1984): (i) checking whether that text is an argument or not; (ii) distinguishing between premises and conclusion; and, (iii) if the argument is not complete, adding the hidden or presupposed premises. Thus, let's consider again the main positions expressed by R1 and R2:

R1: *I think the question is basically that the Cave of Altamira should be opened* ***because*** *it is obviously a place, to say simply, that everyone has the right to visit.*

R2: *I think not, precisely for the same reason;* ***because*** *everyone has the right to access heritage, but if everyone accesses heritage, that heritage is destroyed, isn't it?*

Arguments rarely appear in a stereotypical way (a premise per line, and horizontal line and the conclusion below it) in natural language discourse. Usually, they appear disorganised and hidden in the middle of the discourse, accompanied by non-argumentative fragments. So, step (i) aims to recognise argumentative text among non-argumentative one. We must start looking for certain linguistic particles or phrases that indicates the presence of arguments. Some typical expressions are "therefore", "hence", "consequently", "so", "it follows that", "since", "for",

"because", etc. In the example, both **R1** and R2 use "because" (in bold) and this indicates that there is an argument there.

Step (ii) consists of identifying the premises and the conclusion of the argument and reconstructing the propositions expressed by them. For distinguishing premises from conclusion, we can use linguistic markers again. Particles such as "therefore", "hence", "consequently", "so" or "it follows that" indicate that the conclusion is going to be introduced; particles such as "since", "for" or "because", indicate that what is following are premises. In the previous example, both **R1** and R2 use "because", which gives us a delimitation mark to split the text following this pattern "<conclusion>*because*<premise(s)>". R2, in addition, uses the linguistic marker "but" (underlined), which usually indicates that a new premise is added to the argument. Next, I reconstruct the argument structure of both speakers (Table 1.1).

The next step is reconstructing the propositions. This is a problematic notion in philosophy (Richard, 2013). For the sake of simplicity, we assume here its minimal definition: a proposition is what is expressed by a statement, and it has a truth-value (it is true or false). Next, we show the simplest reconstruction of the argument and the propositions by R1 and R2, removing epistemic verbs and any other linguistic elements not necessary to make clear its main contain (Table 1.2).

However, both R1 and R2 seems to be incomplete, there is a lack of connection between the premise and the conclusion. Step (iii), following Salmon's methodology, consists in the reconstruction of hidden premises. Thus, R1 is presupposing a link between "the right to visit a place by everyone" and "The Cave of Altamira should be opened"; therefore, we need an additional premise (a conditional) to make this connection: "If the Cave of Altamira is a place that everyone has the right to visit, then it should be opened". In the case of R2, the additional premises are "The Cave of Altamira is a heritage site" and "The Cave of Altamira should not be

Table 1.1 Reconstruction of the argument structure of R1 and R2 arguments

R1	R2
The Cave of Altamira is obviously a place that everyone has the right to visit.	The Cave of Altamira will be destroyed if everyone has the right to access it. If everyone accesses heritage, that heritage is destroyed.
The Cave of Altamira should be opened.	The access to the Cave of Altamira should be restricted to experts.

Table 1.2 Reconstruction of the propositions of the R1 and R2 arguments

R1	R2
The Cave of Altamira is a place that everyone has the right to visit	If everyone accesses heritage, then that heritage is destroyed If everyone has the right to access the Cave of Altamira then, it will be destroyed
The access to the Cave of Altamira should be opened to everyone	The access to the Cave of Altamira should be restricted to experts

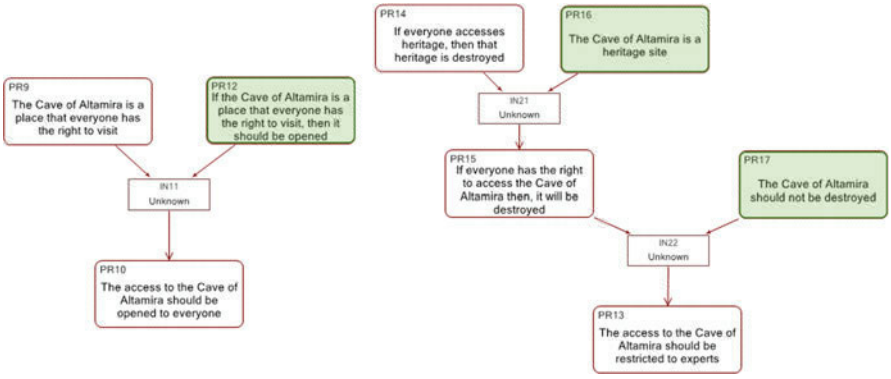

Fig. 1.1 Full reconstruction of R1(left-hand side) and R2 (right-hand side) arguments. In green, the hidden premises that have been added. The nodes content premises and conclusions and the arrows always point to the conclusion

destroyed". Figure 1.1 shows the full reconstruction of R1 and R2, including hidden premises, by means of a diagrammatic representation using LogosLink.

Logical approach considers arguments as single and autonomous units that must be fully reconstructed to be evaluated. The two basic types of arguments are: (i) deductive arguments; and (ii) inductive arguments. Deductive arguments are demonstrative (Salmon, 1984); therefore, if the premises are true and the argument is valid, then the conclusion is necessarily true. However, it does not provide new information because the information in the conclusion is already implicit in the premises; in other words, the conclusion only makes explicit information that was already in the premises. Inductive arguments are not demonstrative (Black, 1967); therefore, premises only provide a degree of support or confidence ore even probability to the truthfulness of the conclusion. However, it provides new information which is not included in the premises.

R1 is reconstructed as a deductive argument, since the conclusion is just the consequent of the conditional that can be inferred because the antecedent is asserted as a premises. R2 is an inductive one, since it is adding new information, such as "experts" also entails that "the Cave should not be opened to everyone".

Logic is a very well-defined methodology for evaluating the quality of arguments. Different types of logics (propositional logic, first-order logic, etc.) allow us to evaluate different type of arguments. However, deductive or even inductive arguments are very rare in natural language discourse because we have to deal with incomplete information and uncertainty in many everyday situations. Moreover, reconstructing arguments in this way usually requires a lot of presuppositions and extracting implicit information that cannot be easily derived from the original text. Finally, it does not allow us to capture the dynamics of the debate and complex argumentative structures cannot be analysed.

A more flexible framework under a logical approach is the Periodic Table of Arguments (PTA) (Wagemans, 2016, 2019). It focuses on the study of arguments

in natural language by means of a step-by-step method for identifying arguments, including more types than deductive and inductive.

1.2.2 Pragmatic Approach

Usually, arguments are elaborated during a communicative interchange, in a dialogue. In such as circumstances, any speaker pursues a specific goal: either justify him or herself or persuade others (Mercier & Sperber, 2018). To achieve this goal, speakers use different linguistic structures and argument structures. From a rhetorical point of view, if a speaker uses rational arguments, he must prove the truth of his premises and the audience will accept the truth of the conclusion (Perelman & Olbrechts-Tyteca, 1973).

Speech act theory is the general frame upon which this approach is built up. A speech act is the production of a linguistic instance, an utterance, under specific circumstances (Searle, 1965). The illocutionary act is the minimal unit of linguistic communication, and it comprises two components (Searle & Vanderveken, 1989): (i) an illocutionary force; and, (ii) a propositional content. For example, "Open the window!" and "Could you open the window?" are two utterances with the same propositional content (i.e., 'you should open the window') but with different illocutionary forces: the former is an 'order', and the latter is a 'request'. Currently, there is not a fixed catalogue of illocutionary forces, although some of them are widely accepted such as *assertion* or *questioning* (Searle & Vanderveken, 1989).

In this section, I will introduce *Inference Anchoring Theory* (IAT) (Reed & Budzysnka, 2010; Janier & Reed, 2017), which main goal is to describe and capture dialogical aspects of argumentation; and Pragma-dialectics (van Eemeren & Grootendorst, 1984, 2004), a normative approach for the development of a rational conversation.

IAT presupposes that the analysis of dialogical interactions allows us to extract the argument form of a discourse, since linguistic argumentative indicators (such as, 'therefore' or 'because') are not as common in spoken language as in written texts (Janier & Reed, 2017). The sequence of interventions during a dialogue also conveys the structure of the argument that the speaker wants to elaborate. Thus, IAT argument analysis requires the following steps: (i) segmenting the utterances of each speaker into argumentative units; (ii) identifying the illocutionary forces and reconstructing the propositional content of the argumentative units; and, (iii) unpacking and reconstructing the argumentative relations between the propositions. Figure 1.2 shows the diagrammatic analysis of a fragment of the first interchange between R1 and R2 using IAT framework and OVA+ (Janier & Reed, 2017), a web annotation tool specifically developed for IAT analysis.

As can be observed in Fig. 1.2, the fragment of the dialogue between R1 and R2 is represented as a graph composed by three main sections: (i) the right-hand side, where we capture the dialogical structure and it comprises both the utterances from each speaker (locutions) and the relevant moves between them (transitions);

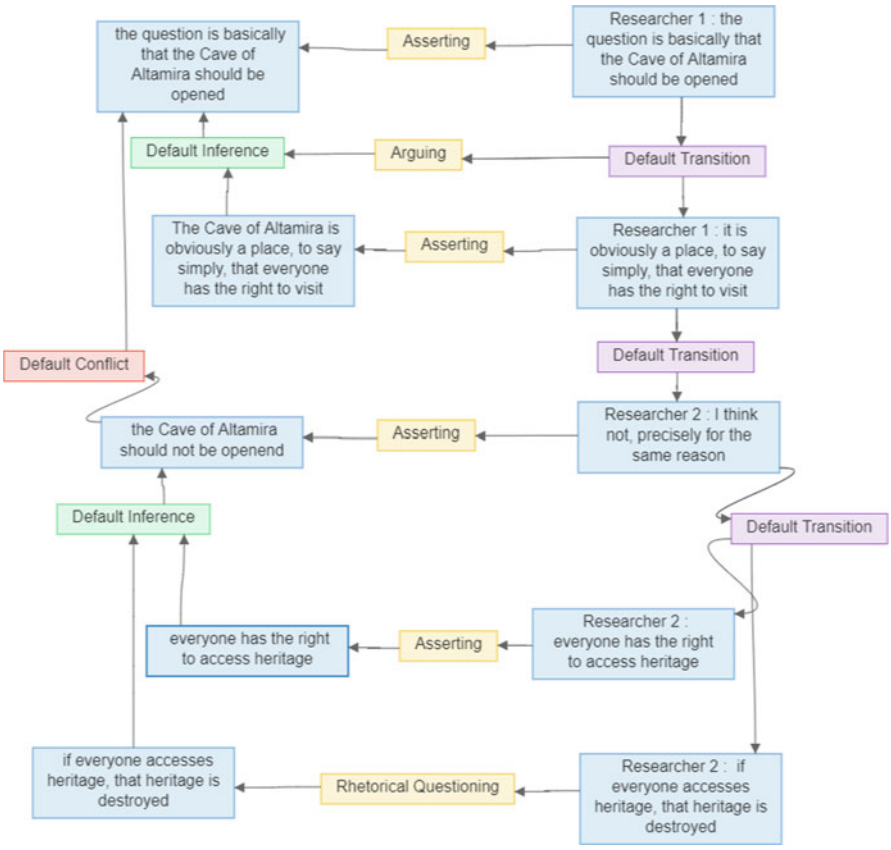

Fig. 1.2 IAT analysis of dialogue between Researcher 1 and Researcher 2

(ii) the middle side, that contains the illocutionary forces representing the speaker's communicative intentions; and, (iii) the left-hand side, which represent the argument structure.

This analysis presents several relevant differences with respect to the logical one. Firstly, it captures both the utterances (the actual statement that was said by the speaker) and their propositional content (with the minimal possible reconstruction), which allows us to keep track about what was actually said by each speaker. Secondly, it shows the dynamic of the dialogue and the turn taking among the participants. Thirdly, the disagreement between both speakers is explicitly captured by the "Default conflict" node in the left-hand side, which indicates that an already said proposition is being neglected. Fourthly, it also gathers the intentions of the speakers through the illocutionary forces, which can come from both the utterances and the turn taking. Figure 1.2 contains four illocutionary forces, although IAT defines more than 20 different ones (Janier & Reed, 2017): (i) "Asserting", which indicates that a speaker just made a statement; (ii) "Arguing", which indicates that

a speaker has the intention to support a claim; (iii) "Disagreeing", capturing the speaker intention of rejecting a statement that has already been said; and, (iv) "Rhetorical question", which shows that a speaker has made a claim but formulating it as a question, so no answer is required.

Thus, from the perspective of Discourse Analysis, we can get a deeper understanding on how argumentation is elaborated using the pragmatic approach rather than the logical one. Its main weakness is the lack of a systematic methodology for evaluating the strength of the argument, something that logical approach provides.

The other mentioned pragmatic approach, Pragma-dialectics (van Eemeren & Grootendorst, 1984, 2004), is, essentially, a normative model where any argumentative exchange is taken as an instantiation of the ideal model of a critical discussion which goal is a reasonable resolution of difference of opinion. This conversation is guided by a set of rules, named "dialogue protocol" (that should be captured by "Transitions"), to achieve the proposed goal; the violation of any of these rules will constitute a fallacy.

Pragma-dialectics establishes three basic components for a rational conversation: (i) setting the roles of participants, basically protagonist (who argues in favour of the standpoint) and antagonist (who argues against the standpoint); (ii) going through the four stages of the discussion (confrontation stage, opening stage, argumentation stage and concluding stage); (iii) evaluation if any of the 15 rules of critical discussion (spread along the different stages) were violated.

The analysis of an argument within this framework requires five different steps: (i) identifying the standpoints of the discussion, which is composed by a proposition and the illocutionary force (the attitude of the speaker with respect to that proposition); (ii) recognizing the protagonist and the antagonist, assigning their respective standpoints; (iii) agreement on the shared propositions that establishes the common ground of the speakers; and, (iv) identifying the argumentative structures used by the speakers during the discussion, which include both argument schemes and critical questions (Walton et al., 2008). A deeper analysis of this framework is out of the scope of this paper, since it requires the analysis of the full dialogue; however, from the perspective of Discourse Analysis, it is a very valuable framework.

1.2.3 Cognitive Approach

The last approach to the nature of arguments that I will explore in he is considering argumentation as a mental process (van den Hoven, 2015) or a cognitive activity (Mercier & Sperber, 2018). An argument expressed in natural language (written, spoken, trough images, etc.) is not the argument itself but the representation of a mental process. Therefore, understanding or interpreting an argument always entails the reconstruction of the corresponding mental process.

Thus, any linguistic argument is a sign vehicle of the mental process and, therefore, it must be analysed as a semiotic entity (van den Hoven, 2015): its textual part is a sort of *representamen* which stands for the argument itself –the

object– which is a cognitive entity with a particular goal. As hearers, we reconstruct that connection between the textual argument and the argument itself through the *interpretant* (Peirce, 1958; Chandler, 2003).

This semiotic conception of arguments presupposes two main types of relationships (van den Hoven, 2015):

– Mimesis: The textual argument is a perfect imitation of the mental process of argumentation.
– Diegesis: The propositions constituting the argument convey a specific interpretation and evaluation of the world.

Both relationships are the constituent parts of the named 'discourse world'. It comprises the background, presuppositions, commitments, beliefs or desires of each speaker –shared or not– and, therefore, it plays a major role in the reconstruction of the argument for its understanding and evaluation (Mercier & Sperber, 2018). Under this approach, the intention of the speaker of making an argument is not enough to have an argument, it also requires to be recognised as that by the hearer. Therefore, arguing is, essentially, a social activity (Mercier & Sperber, 2018).

From our point of view, this is the richest framework for modelling argumentative discourse also it is the most complex one. To the best of my knowledge, there is not still a fully developed framework for that. IAT/ML (Gonzalez-Perez, 2020), that combines IAT with conceptual modelling (Gonzalez-Perez, 2018), is a theoretical approach under development grounded on this conception of argumentation.

IAT/ML defines four basic steps to carry on a cognitive analysis of an argumentation: (i) setting the initial discourse world of the participants in the conversation by means of a conceptual modelling language; (ii) identifying the chunks of texts that are acting as a sign of a mental process of argumentation (following linguistic indicators, grammar structures, images, etc.); (iii) reconstructing the argument mentally elaborated by each speaker using the contextual information and foreknowledge available for the analyst (which might significantly vary between analysist); and, (iv) evaluating whether the result of the interaction requires any change in the discourse world.

Figure 1.3 shows a reconstruction of the discourse world (ontology) underlying the debate between R1 and R2 about the Cave of Altamira using ConML, a conceptual modelling language.[2] Each node represents a discourse entity, such as the "Cave of Altamira", which appears in a central position since it is the main entity discussed in the debate. Both speakers know that the cave is the support of the prehistorical paintings, but they disagree with respect to the "RightOfUse", which appears twice taking two different values: once as "Experts may access" and other as "Everyone may access". Each edge defines a directed connection between entities, such as between the "RightOfUse" and two different groups of people, "Experts" from one side and "Everyone" from the other side.

[2] http://www.conml.org/default.aspx