CHARACTERISING RHETORICAL ARGUMENTATION

By

Floriana Grasso

SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
AT HERIOT-WATT UNIVERSITY
ON COMPLETION OF RESEARCH IN THE
DEPARTMENT OF COMPUTING & ELECTRICAL ENGINEERING
JUNE 2003.

I hereby declare that the work presented in this thesis was carried out by myself at Heriot-Watt University, Edinburgh, except where due acknowledgement is made, and has not been submitted for any other degree.

Floriana Grasso (Candidate)

Dr. Alison Cawsey (Supervisor)

Date

Abstract

The study of argumentation has undergone periods of decadence alternate with periods of renaissance. Currently, we are witnessing one of the latter, as proved by research flowering in many issues and aspects, both in Philosophy and in Artificial Intelligence. The work presented in this thesis is concerned in particular with "rhetorical argumentation", which on the other hand has enjoyed consideration to a somewhat lesser extent. By rhetorical arguments we denote arguments which are both heavily based on the audience's perception of the world, and concerned more with evaluative judgments than with establishing the truth or otherwise of a proposition. Rather than a logic focus on argumentative reasoning, or a pure computational linguistic focus on modelling discourse which happens to be argumentative, we place ourselves halfway and specifically focus on the characterisation of rhetorical argumentative discourse. Methodologically, we do this by explicitly drawing upon the philosophy of argument.

The thesis proposes an all encompassing framework for the formalisation of rhetorical argumentation, inspired by a well established philosophical theory, the New Rhetoric. The thesis starts by giving a formal definition of rhetorical argument, and rhetorical reasoning. It proposes a model for a rhetorical dialogue game between two partners. It provides a characterisation of the game partners, in terms of their mental attitudes. The model is then applied to a particular task and domain, that is the provision of health education on nutritional issues. An ontological analysis of the knowledge required to perform the task is proposed, based on a psychological model of behavioural change. Finally, the thesis proposes an architecture integrating the various components, informed by the formal model defined.

The work is admittedly biased towards the generation, rather than understanding, of argumentative discourse, and we see, and indeed we built, our formalisation to be comfortably used in a system for generating argumentative discourse based on planning. At the same time, we see our work as a contribution to typical pragmatic issues of providing a characterisation of what the speaker has to take into account when producing a "good" argument, and therefore what the hearer can ascribe to the speaker. The framework is well grounded on theoretical foundations, and indeed our proposal coherently integrates work in the philosophy of argument, in linguistics, in social psychology and in artificial intelligence.

Acknowledgements

People are divided into two categories: those who always *have* to read the acknowledgement section, and those who cannot be bothered. I belong to the former. These words are dedicated to all the kindred spirits who will read them.

The enterprise of a PhD is not worth the paper on which the thesis is written, if it does not teach something which goes beyond the actual content of the work itself. In fact, somebody would say it is not worth, fullstop. I would like to thank here all those who contributed to this “added value”. No, I’ll rephrase this. To the “main value”.

First and foremost I am grateful to Alison Cawsey. For showing me that research is not only high blue skies, but also rigour and critical work. For teaching me how to write in English (all the Italianate sentences here are entirely my fault). And for patiently adapting her concept of a deadline to mine.

Jean Carletta and Keith Brown, my examiners, for the most enjoyable discussion on my work.

Alan Gibbons, Paul Leng, and all of the Department of Computer Science of Liverpool University, that I joined half way through my PhD: for giving me the opportunity to finish this work while working there. A special thank to Mike Wooldridge, for giving me a reason to get this stuff to completion (no, it’s not the new laptop, Mike!).

Yoram Kalman, Vered Yatsiv, and all the staff at K.I.T.

Fiorella de Rosis, for disclosing me the captivation, the enjoyment and the reward of research. For her dear friendship. And for always believing, despite all odds, that my ideas were worth to be pursued.

Chris Reed: a colleague, and a friend. For involving me into the almost ”mystical” experience of the Symposium at Bonskeid House. And Giuseppe Carenini, the third cavalier of the CMNA workshop series.

Those with whom I have had occasional (or not so occasional) chats about my research: Michael Gilbert, Leo Groarke, Jaanne-Maike Gerlofs, and Corin Gurr (*aka* the infamous ”G” group), Henry Prakken, Leila Amgoud, Tim Norman, Ehud Reiter, Valeria Carofiglio, Nadia De Carolis, Helen Gigley, Cécile Paris, Peter McBurney, Chris Mellish, Trevor Bench-Capon, Cristiano Castelfranchi, Joel Katzav, Dianne

Berry, Alistair Knott, Massimo Poesio, Simon Parsons, Daniela Carbogim, Grahame Hirst, Doug Walton, Joris Hulstijn, Rino Falcone, Michael Bacharac. And, last but first of all, Bill Mann.

Ray Jones, who always seemed to enthuse over the idea of arguing with patients (!), and the staff at the Public Health and Human Nutrition Departments of the University of Glasgow, especially Catherine Hankey and Sophie Tessier.

Diana Bental, with whom I shared the office at Heriot-Watt: for her sense of humour, and her exquisite kindness.

Manuel Trucco, who, being the first specimen I have encountered of an Italian lecturer in UK, helped me understand the often mistifying ways of British academia, and who forced me to stand before a class, albeit small, and teach for the very first time.

The Italians in Edinburgh: Francesca, Tiziana, Gianluca, Stefania, Nunzio, Mariangela, Andrea, Mimmo. And Tim, who surrended to the Italian invasion of his flat by deciding to learn Italian and to find an ally in one of the enemies. For the laughters, the chats until dawn, *caffé*, tequila and movies.

The Italians in Liverpool: Annarita, Felice, Gianni, Victoria, Edward, Giuseppe, Michelle, Maresa, Gianluca, Momo, Riccardo, Nunzia, Marco, Kate, Leonardo, Carlo, Eugenia, Lucia, Mimmo, Massimo, Pio, Chiara, Sergio, and whoever else I've undoubtedly forgot. And the forty plus people who turned up for my birthday pizza: I know you only came for the guest, but it was fun nevertheless.

The other *emigranti*, especially Nicola, Marina, Daniela, Roberto, Antonella.

Ken, for such an incredible variety of things, from fixing my Linux problems, to taking pictures at my wedding, not to mention introducing me to the most *squisita* Chinese cuisine.

Jeff: so, the thesis is over. *Ormai*.

Marialuisa, for being there.

Francesco: life in Edinburgh, and life in general, would have been so much different without him.

Valentina: for making Liverpool a little piece of home, of family, of things usual and warm.

Laura, Nicola, Matteo, Isabella: for giving me a perspective of life, in which, thank goodness, my work is so insignificant.

Edinburgh, for always smiling at me.

Liverpool, for always making me smile.

All my family, and in particular Mum, Dad and Cesare: for their love and their incessant support, no matter what I was up to, or how hard it was for them.

My husband Michele. Words are so inadequate, to even start to explain.

And ALL the ones who bore with me everytime I would swear this was *just impossible*.

While in Edinburgh, this work had the financial support of EPSRC.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

> Plenty of things in real life are superlatively uninteresting; so that it is one-half of art to select from realities those which contain possibilities of poetry.
>
> - Honoré de Balzac, *The Human Comedy (The Message)*

1.1 The Problem

Consider the task of providing information: the information seeker has a particular information need, formulates the appropriate query to an information provider, whether human or not, and (hopefully) obtains the desired piece of information. An efficient and efficacious information provider would also be able to tailor the reply to the individual seeker, as both intuitively and experimentally it is accepted that personalised information is preferred and better understood.

Suppose now that the information provider needs also to convince the information seeker of the validity of the information provided. The hypothesis, in other words, is that the receiver of the message cannot be assumed to blindly believe anything communicated, but needs some reinforcement to the claims before accepting them. This may happen for several reasons, most importantly because the information seeker is, in one way or another, an "expert" in the domain at hand, having a pre-existent set of well supported beliefs.

Suppose, furthermore, that the information provider has an agenda of its own, which is not only concerned with convincingly informing the user, but also with trying to change the information seeker's attitude, or even behaviour, towards the topic under discussion. Many factors potentially make the task complicated: the behaviour suggested may be perceived as not appropriate, may have harmful side effects, or may simply be unpleasant or hard to achieve.

Finally, suppose that there is no explicit information seeking from the interlocutor's side at all: the information provider takes the chance of an occasional contact in order to pursue its hidden agenda to get information across that could hopefully provoke the desired change.

This is an example of how things may go terribly wrong[1]:

Provider: *I would advice at least 6 slices of wholemeal bread per day.*

Seeker: *Gulp. It's the fibre polis. Do you have any idea how long it would take to eat six slices of said bread? Can you be sure of your recommendation without knowing my average intake of fibre and complex carbohydrates over a given week? Perhaps this is a test. Why should I take unsolicited advice from a virtual nutritionist of unknown provenance? Do you work for a flourmill?*

The research reported in this thesis, by and large, concerns the problem of understanding why situations like the one in the dialogue above happen. Or, in general, what it is that people use in order to change other people's mind, apart from coercive means of course, and why they are, or are not, successful. It is clear that we are talking about argumentation here. It is also clear, however, that we are *not*, or not only, talking about logical demonstration, or proving that a standpoint can be held by reasoning from a set of premises and axioms. Naturally occurring argumentation seems to be one in which soundness is not necessarily the first priority: arguments are judged to be more or less *successful* when their premises make the conclusion plausible to a greater or lesser degree.

It appears that humans are very comfortable with this type of argumentation, and that there are mechanisms in place to deal with plausibility in a very subtle and sophisticated way: not many people would complain about being *deceived* by TV advertisements promising a glamorous life if washing powder so-and-so is used. In fact we all are prepared to read between the lines, to capture the main message, and perhaps to recognise fallacious, if not deceiving, arguments. Nevertheless, the advertisement world, to continue with the analogy, relies heavily on the messages produced, which are typically targetting a particular audience, emphasising only a few aspects of the product to promote, encouraging behaviours, creating trends. In one word, *rhetorical*.

By rhetorical argumentation we mean argumentation which is both heavily dependent on the audience it is addressed to, and is concerned with evaluative judgements, rather than logical proofs. Despite the fact that this seems to be the most widely used form of argument, Artificial Intelligence work in argumentation seems to have paid attention to rhetorical argument to a somewhat lesser extent: research has mainly concentrated either on logical or quasi-logical approaches to argumentation. The focus has

[1]The exchange is taken from a corpus of real dialogues, which will be introduced in Ch. 5.

typically been on reasoning and negotiation, or on discourse modelling approaches, with argumentative discourse as a byproduct, or a special case, of more general discourse structures. This is especially true for research in multi-agent negotiation, and general communication protocols, which usually assumes a rather “safe” environment, where only a limited form of deception, if at all present, can take place. The lack or rhetorical abilities, however, limits considerably, in our opinion, the credibility, and perhaps the usefulness, of electronic agents, which, especially in marketplace scenarios, should be expected to deal with all sorts of persuasion means. To design useful agents in this respect, one has to first address the broader issue of how these “rhetorical abilities” could be more precisely defined and described.

1.2 The Approach

Our interest in rhetorical argumentation is therefore motivated by its endemic character. But we are also motivated by the complementary goals of building on the long history of work in the more philosophical aspects of argumentation and rhetoric, and developing models amenable to implementation. The former goal meant that we specifically posed ourselves the aim of drawing insights from theories coming from the philosophy of argument. The latter goal meant that we always maintained a perspective on implementation, critically evaluating the abstract theories coming from philosophy with respect to their computability.

Our research goal is therefore to precisely characterise rhetorical argumentation. This has been addressed by identifying a suitable, philosophical theory of argumentation, the *New Rhetoric*, and translating its insights into:

1. a semantic for rhetorical argumentation;
2. a model of rhetorical dialogue games;
3. a mental model of an agent able to argue rhetorically;
4. an annotation scheme for analysing naturally occurring rhetorical dialogues.

The result of this investigation is therefore a novel, complete formal framework for rhetorical argumentation. It is complete, because we analyse the problem from all perspectives, starting from the study of the very process of arguing rhetorically, and continuing with the study of argumentation as a two party process, with a formalisation of a mental model of each of the two parties, with special attention to natural discourse.

Furthermore, we analyse, as a way to exemplify our work, a possible application scenario, by collecting and annotating real case examples from that scenario, exploring the ontological dimension of applying our theory to the scenario, and proposing an architecture for a possible implementation of the model.

We report this investigation in this thesis.

1.3 Outline of the Thesis

The thesis is structured in such a way that each aspect of the problem is treated separately. The various Chapters are designed to be self-contained, in the sense that each introduces one particular issue, surveys relevant literature, and shows the solution adopted, with some final comments on the method.

This thesis is divided into two parts, reflecting the twofold interest that has driven us throughout our research: a more philosophical one, and one closer to artificial intelligence issues. The first part, from Chapter 2 to 4, sets up the theoretical framework. In each of the Chapters in this part, after the introduction, a Section on philosophical backgrounds surveys the theories which underpin our model. The second part, from Chapter 5 to 7, explores the more practical aspect of applying the theoretical framework to a specific scenario. In more detail, the Chapters, besides the current one, are structured as follows.

Chapter 2 analyses the general notion of rhetorical argumentation, starting with some historical traits on the realm of rhetoric, and the way it has developed over the years, and describes the angle from which it has been seen in the work presented here. The Chapter shows how the framework in the thesis draws upon a specific theory in the theory of argument, the New Rhetoric, and gives the basic definitions that constitute the framework.

Chapter 3 looks at the dialogical dimension of rhetorical argumentation, by concentrating on how scholars have modelled argumentative dialogues and dialogue games, and proposing a notion of rhetorical dialogue game, based on work in informal logic.

Chapter 4 further explores the dialogue dimension, by tackling the issue of modelling the mental attitudes of the dialoguing partners in a rhetorical argumentation discourse.

Chapter 5 moves slightly away from the theoretical aspects of the work, in order to give a more concrete grasp on the problem at hand. An application domain is presented, in the area of health education, that presents ideal features for implementing and testing the framework proposed. This is demonstrated by describing the result of an experimental study conducted via e-mail with a sample of recruited participants.

Chapter 6 begins the investigation towards a possible implementation of the framework, by concentrating on the representation of knowledge: ontological issues are considered in the task of reconciling different, and possible heterogeneous aspects of the problem into a unifying knowledge model.

Chapter 7 presents the architecture of a system meant to encapsulate all the key aspects of the framework presented in the proposed domain of application.

Finally, Chapter 8 presents some reflections on the work carried out, as well as perspectives for a future broadening of the research scope.

A collection of Appendices shows in more detail the experimental and modelling work which is exemplified in the thesis.

Chapter 2

A Theory of Rhetorical Argumentation

> [...] "philosophy", as a well-defined and homogeneous domain of activity exists as little as does "science". There are the words, there are even concepts, but human existence shows no trace of the boundaries the concepts imply.
>
> - Paul Feyerabend, *Three Dialogues on Knowledge*

Over the years, the study of argumentation has undergone periods of decadence, alternating with periods of renaissance. Currently, we are witnessing one of the latter, as proved by research flourishing in many areas, both in philosophy and in artificial intelligence. The work presented in this thesis is mainly concerned with one specific sort of argumentation, that is rhetorical argumentation, which on the other hand has enjoyed consideration to a somewhat lesser extent. The present Chapter aims at providing an overview of the main issues in argumentation theory and rhetoric which are relevant to the work presented here, well aware that a comprehensive survey would be a hopeless task. We introduce the New Rhetoric, a philosophical theory that has been computationally modelled in this thesis work. The Chapter also presents an *excursus* on the ways in which research in rhetoric has found application in artificial intelligence, and especially in computational linguistics. Finally, the Chapter introduces a new formal framework for rhetorical argumentation, which draws on the New Rhetoric, and which underpins the computational model presented in this thesis.

2.1 Historical Traits

A survey on relevant works in argumentation theory cannot ignore classical works. Important more modern works will also be surveyed in what follows, and the New Rhetoric will be briefly introduced, while a more comprehensive treatment will be done in Sect. 2.2.

2.1.1 Classical Tradition

The fathers of Rhetoric[1] are traditionally believed to be Corax and Tisias, in Syracuse, Sicily, about 500 years before Christ, when a series of trials was put in place to claim back land that the tyrants of the town had confiscated. Corax and Tisias acted as jurisprudential advisers, teaching the art of effective speech, and their strategy was famously based on the perception that "seeming true" is more important than "being true". At about the same time, again in Sicily, another rhetorical school of thought was flourishing, called *psychagogy*, based on antithetical reasoning and *polytropism*, that is the ability to find different types of argument for different types of audience. From its very origin, therefore, rhetoric has been characterised by its communicative dimension, and by its emphasis on adapting the speaker behaviour to the particular audience. Indeed, the technique of contradicting by means of antitheses, showing how an argument could be seen from different points of view, was at the basis of the Sophists' studies in Greece, about 50 years after Corax and Tisias. Famous, in this context, is the work by Gorgias, a Sicilian moved to Athens, who merged rhetoric and poetry in a unique style, and who first recognised the existence of *fi gures* of speech.

Fiercely opposing Sophism, Plato distinguished "rhetoric", in his view a mere persuasion exercise which disregards the content of the speeches, from "dialectic", the art of good discussion aimed at analysing specific subjects, by decomposing discourse into its basic elements. This attack on the Sophists was in retrospect quite unjustified: Gorgias, for instance, on the contrary recommended the rhetorician to gain deep knowledge of scientific, jurisprudential and philosophical arguments. Plato's view would, however, mark forever the art of rhetoric, which has acquired, with the centuries, the reputation of being just a set of manipulative discourse techniques.

The most important theorist of rhetoric, however, remains Aristotle, who gives constitutive definitions of the species of argumentation, in his three most influential treatises: *(Prior and Posterior) Analytics*, *Topics* and *Rhetoric* (Aristotle, about 350 BC 1964, about 350 BC 1928, about 350 BC 1926). In the first, Aristotle characterises what we now call "demonstrative" arguments, that is arguments whose objective is to establish the "truth or falsity" of a certain standpoint. In *Topics*, "dialectic" arguments are described, in which the hypothesis of using premises which are evidently true is relaxed, and whose aim is to obtain that the standpoint is generally "acceptable", rather than universally true. Finally, in *Rhetoric*, Aristotle reflects on arguments that will persuade *in the given circumstances*, by appealing to one particular audience's set of beliefs, and whose aim is to achieve persuasiveness to that audience, rather than general acceptability.

[1] Material in this Section draws on various sources, among which most notably Cattani (1990), van Eemeren *et al.* (1996), Garavelli (1990), Grassi (1980), Lo Cascio (1991), and Murphy (1974).

Rhetoric in Rome has its major development in the first century before Christ, with the first work in Latin, *Rhetorica ad Herennium*, by some Cornificio (though erroneously attributed to Cicero for some time). The work establishes the Latin terminology in the field, and defines the organisation of rhetorical discourse as comprising four parts: *inventio*, or collection of the arguments, *dispositio*, their arrangement, *elocutio*, their expression in the language, and *pronuntiatio*, their utterance. The major scholar of the Roman rhetoric was of course Cicero, who defended rhetoric against many polemics, most notably those of the Stoicists. Cicero's *De Oratore* will be a fundamental treatise both in classical and Medieval rhetorical studies, placing rhetoric in a central place between the two other *artes* of the *Trivium*: grammar and dialectics. We have however to wait until Quintilian's *Institutio Oratoria*, in the first century A.C., after the Roman Republic gave place to Imperialism, for a systematic summary and comparison of the previous rhetorical doctrines, re-elaborated with a pedagogical emphasis, in the typically pragmatic Roman approach. The treatise, though not particularly original, will have a fundamental impact to the definition of the field up to the present day, not only when thinking of the virtues of rhetoric, but, perhaps especially, when thinking of its vices (notorious are the suggestions given to the orator who is asked to defend a false position).

Ancient rhetoric was passed almost intact to Medieval scholars, although the emphasis was somewhat different, as well as its fortune, as the Fathers of the Church expressed some diffidence towards its subtleties. Most notably, Augustine attempted in the fifth century, to combine rhetorical tradition and the message of Christianity. Despite advocating a *sermo humilis* (a "humble sermon"), Augustine, a rhetorician by vocation and studies, does not underestimate the importance of the figures of speech, used with various intensity depending on the purpose: teach, praise or motivate.

From the Middle Ages onwards, we can only talk about variations to the classical tradition, addressing more specific problems. The thirteenth century witnesses the triumph of the *ars grammatica* as the basis of all other *artes*. Dialectics flourishes until the end of the fourteenth century as the art of discussing to reach a conclusion on two positions which are contrasting but equally plausible. This gave raise to the literary genre of *disputatio*, or debate. The Humanism, on the other hand, prefers rhetoric to dialectic, and exalts the rhetorical syllogism (see for example the work of Lorenzo Valla, who recuperates Cicero's project) as it can be applied to practical situations. The rejuvenation of rhetoric was most importantly due to the discovery, in 1416, of a complete copy of Quintilian's *Institutio Oratoria*, of which only abridged versions had been known, and, five years later, of Cicero's *De Oratore*. In the following century, Pierre de la Ramée notably promoted a restriction of the idea of rhetoric, by claiming that *inventio* and *dispositio* are in fact part of logic or dialectic, while the sole realm of rhetoric is the combination of *elocutio* and *pronuntiatio*, a position greatly emphasised by Descartes' rationalism.

With very few exceptions, first and foremost Giambattista Vico's studies, in fierce polemic with Descartes, and in which rhetoric is seen as a human exigence for explaining knowledge, in the Age of Enlightenment, and indeed until now, rhetoric is synonym of "composition". It becomes basically the school of style and the study of literary genres, where scholars dedicated themselves to the classification of the most fine grained aspects of discourse. Such "cosmetic" emphasis has contributed to the discredit of rhetoric characterising the nineteenth century, so much so that the very term "rhetoric" has come to signify artificious, futile, insincere, decadent.

2.1.2 The New Rhetoric

If the misfortune of rhetoric was due to its being deprived by all its features but *elocutio*, its renaissance had to be due to bringing it back to its original conception of a theory of persuasive discourse.

In a project concerned with giving back to rhetoric its deserved place in the study of argumentation, the Polish-Belgian philosopher Chaim Perelman, together with the social scientist and philosopher Lucie Olbrechts-Tyteca, published, in 1958, their theory, which they called the *New Rhetoric*, explicitly referring to the Aristotelian definition (Perelman & Olbrechts-Tyteca 1958). The New Rhetoric produced an awesome impact on the theory of argumentation. The two philosophers aimed at identifying "discoursive techniques allowing us to induce or to increase the mind's adherence to the thesis presented for its assent" (NR[2] p. 4). Their point of departure was the observation that, in every day life, people rarely produce just a perfect logical proof of their claims, as it is "illusory" that facts speak for themselves, even when presenting scientific theories. Arguments rarely rely on knowledge, but rather they are justifications of points of view by appeal to the values and opinions of those to whom they are addressed. The notion of *audience* is clearly essential in this context: the same argument can produce very different results when addressed to different people. The aim of the orator will then be to identify the characteristics of the audience to appeal to, in order to be more effective, and it is this aspect that most closely adheres to the Aristotelian concept of rhetoric. The two philosophers' study, rather than establishing an a priori "logic of value judgements", went, in a Fregean style, from examples to generalisation, by collecting and classifying arguments that are successful in practice.

The New Rhetoric more or less directly inspired various movements, like Informal Logic in North America, or Formal Dialectics in the Netherlands, as well as many philosophers, especially in Italy, such as Norberto Bobbio, who translated it into Italian, or Umberto Eco, with his semiotics of conversational interactions. Indeed, in Italian tradition, argumentation has always been mainly rhetorical: so much

[2] We will refer, throughout the thesis, to the English translation of the treatise (Perelman & Olbrechts-Tyteca 1969), which we will abbreviate in citations as NR.

so that, for instance, Lo Cascio (1991), when describing “the three main argumentation genres”, refers in fact to the three *rhetorical* argumentation genres identified by Aristotle (*iudiciale, deliberativum, demonstrativum*). Perhaps because fundamentally close to this tradition, we took inspiration from this theory for our model, as we will explain later on.

2.1.3 Modern Developments

In mentioning, albeit very briefly, current trends of rhetoric in the philosophy of argument, we cannot ignore a work that, together with that by Perelman and Olbrechts-Tyteca, gave a tremendous contribution to shape the modern field, that is the work by a British-American philosopher, Stephen Toulmin. *The Uses of Argument* (Toulmin 1958) shook the old fashioned way to approach the study of arguments, which was mainly based on a, sometimes forced, similarity with logic demonstrations, by claiming that arguments cannot be seen as just chain of premises leading to a conclusion, but have a much more complex architecture. His notorious schema (see Fig. 1) describes the argument structure as comprising a Claim which is supported by a Datum via a rule, or Warrant, in turn having a Backing. The support can be more or less strong, as expressed by the Qualifier. The Claim holds unless there is an exception, a Rebuttal. Toulmin’s classical example of how the schema works is:

- Harry was born in Bermuda (*Datum*).
- A person born in Bermuda is a British subject (*Warrant*),
- on account of the following statutes... (*Backing*).
- So, presumably (*Qualifi er),*
- Harry is a British subject (*Claim*),
- unless both his parents were aliens, or he has become a naturalised American, ... (*Rebuttal*).

Despite various criticisms (see van Eemeren *et al.* (1996) for a discussion), Toulmin’s schema is still one of the milestones of the philosophy of argument, and its simplicity, yet comprehensiveness, has managed to break away from the community of philosophers, to be adopted, among others, by scientists in artificial intelligence.

Not that Toulmin’s almost aseptic structure for arguments represents an attempt to disregard rhetoric: on the contrary, with his idea of a complex, beyond logical characterisation of arguments, Toulmin actively tries to stress the rhetorical aspects of the argumentation process, as opposed to the demonstrative

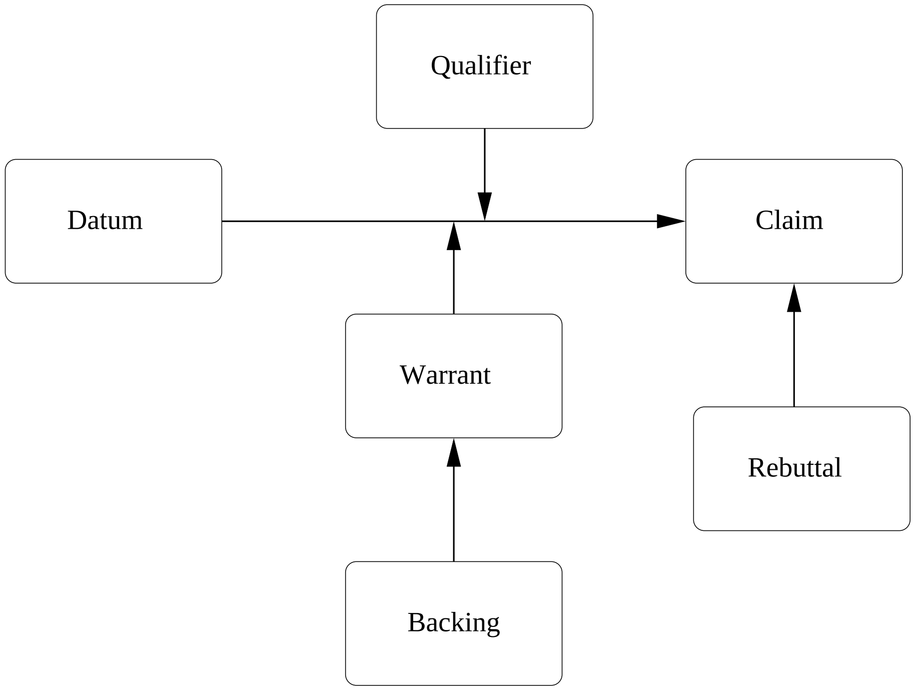


Figure 1: Toulmin's argument structure

ones. Also, his theory of rationality is fundamentally rhetoric, maintaining that new ideas have to be argued over by their creator in order to be adopted by the community.

The work by Toulmin was, as van Eemeren *et al.* (1996) put it, one of the causes for the rejuvenating interest in argumentation theory in the second half of the twentieth century, and was, as admitted by Johnson & Blair (1994b), one of the inspirations for the already mentioned movement called *Informal Logic*, began in North America in the early seventies.

Informal logicians set themselves three main tasks: to ascertain which texts and discourses are actually "arguments", to analyse those arguments by identifying their subparts, and to evaluate them, particularly by pinpointing fallacious ones. The emphasis here is on the analysis of the *structure* of good arguments, in an attempt to capture, for argumentation, a formal aspect which could affiliate its study with the study of other, typically formal sciences, such as logic or mathematics. They do not want to be alternative to logicians, however, but rather to try and use logic tools to analyse everyday arguments. Inspired, together with Toulmin, by the work of Hamblin (1971), as well as classic argumentation theorists, researchers in informal logic emphasise both the study of fallacies (Woods & Walton 1982; Groarke, Tindale, & Fisher 1997, among others), and the analysis of conversational argumentation (Walton 1989; van Eemeren & Grootensdorst 1992, among others), two fundamentally rhetorical aspects of argumentation. Even the work on fallacies is, in fact, not aimed at ostracising them, as happened in the past, but rather at identifying when they are admissible and how they should be structured. The assumption is that fallacies are both part of human argumentation and, in many cases, perfectly accepted by the majority of people as arguments. The perspectives of informal logicians are diverse, including computational ones: "in the future, informal logic may be linked to formal logic by attempts to convert its insight into formal analogues which can be used to construct computer models of natural language reasoning, in research in artificial intelligence, and so on" (Groarke 1997).

An interesting application, with some adaptations, of Toulmin's categories in rhetoric, is, in Italy, the *argumentative grammar* by Lo Cascio (1991), a work that, by defining associative rules for argumentative acts, is naturally applicable, and indeed has been applied, to the analysis of discourse organisation structure, an important rhetoric task.

Another important development is the Dutch *Pragmadialectics* (van Eemeren & Grootensdorst 1992), which, on the basis of Austin, Searle, Grice and Lorenzen, provides a comprehensive analysis of argumentative "natural" dialogues. The pragmadialecticians combine the study on the formalism to represent data, from modern logic, and empirical observations, from descriptive linguistics, for the analysis of argumentative dialogues, modelled by dialectics, seen as sets of linguistic acts, or moves. The model includes ten fundamental principles ruling "correct" argumentative dialogues, whose violations define ten fundamental fallacies. While descriptively interesting, however, the model does not provide the functional dimension that would permit generating rhetorical discourse from a given context.

As rhetoric is also about linguistic style, modern rhetorical developments include literary movements, concentrated on a stylistic perspective rather than argumentative/reasoning ones, a perspective which is perhaps summarised in the work of the so called μ *Group*, where, among other things, it is provided a fine re-organisation of the uncountable figures of speech coming from classic rhetoric (μ Groupe 1976).

Among recent movements in rhetorics, it is important to mention the one concerned with a "rhetoric of inquiry", that is the way in which rhetoric must be used in science communities in order to establish new theories. The movement argues that scientists must recognise that preconceptions and passion have so important a role in scientific inquiry that the progress of science cannot be made without a rhetorical debate (Fahnestock 1999; Pera 1991).

Finally, moving towards an artificial intelligence, or better a cognitive science perspective, it should also be mentioned the project of Cognitive Rhetoric (Turner 1991), advocating the introduction of cognitive modelling to the study of language and literature, and inspired by Lakoff & Johnson (1980) and Fauconnier (1985).

2.2 The New Rhetoric: a Point of Departure for our Framework

The principal aim of the thesis is to build a complete computational framework for rhetorical argumentation. We have already mentioned that, in this endeavour, our intention was to inspire ourselves by a well established theory of argumentation, as we firmly believe that the artificial intelligence researcher should not try to reinvent the wheel, but rather to draw upon what is already available in different disciplines. In particular, we were interested in theories that, by virtue of their rhetorical nature, could give hints on the functionality, rather than the mere formality, of the rhetorical argument. In other words, we were not so much interested in theories analysing rhetorical arguments as such, but in those providing suggestions on how to generate rhetorical arguments, accounting for the various circumstances in which an argument could be used. Also, we were mainly interested in a more cognitive, high level aspects of the rhetorical process, rather than a low level, strictly linguistic one: therefore work on, for example, figures of speech was felt to be part of a subsequent development of this thesis work, and was set aside for the time being.

As we anticipated earlier in this Chapter, we found a suitable theory in the work by Chaim Perelman and Lucie Olbrechts-Tyteca, the New Rhetoric. The emphasis on the audience perception of the argument, and the way to classify argument schemes in terms of an ontological account of the objects of argumentation, make the New Rhetoric extremely interesting from the point of view of computational modelling. The theory has inspired the development of our framework in many ways, as it will become clear as the thesis unfolds. In the rest of this Chapter we will concentrate on the model of rhetorical arguments, but in following Chapters we also show how the New Rhetoric helped us in the tasks of dialogue structuring, knowledge and belief modelling, together with providing heuristics for the actual implementation of the argumentative model. We will start, however, with a description of Perelman and Olbrechts-Tyteca's work in somewhat more detail, for the reader to be able to understand the subsequent parts of the thesis.

2.2.1 Argumentation Is Based on Premises

Any argumentation theory, and the New Rhetoric is no exception, is based on the analysis of what can serve as premises in argumentation, that is what the speaker can assume adherence to by the audience in order to build up a stronger argument. The New Rhetoric's notion of these "objects of adherence" takes into account not only facts, but also the importance that the audience attaches to facts. Moreover, in the rhetorical spirit of argumentation, premises should comprise also less strict beliefs and feelings of the audience. With this in mind, Perelman and Olbrechts-Tyteca provide the following classification of premises[3]:

Premises relating to the "real" are the premises that more closely relate to analytical argumentation. These consist of:

- *Facts*: statements about reality which can be assumed to require no further justification, accepted by both the audience and the speaker. In Perelman and Olbrechts-Tyteca's view, however, it is not possible to classify data as facts once and for all, as there is no statement on which one can postulate uncontroverted agreement. A fact is therefore what can be claimed to be valid in front of an ideal "universal audience", bearing in mind that the status of fact can be lost, either because doubts may have been risen within the particular audience or because the current audience has expanded with new members.
- *Truths*: more complex systems of connections between facts for example scientific theories, or philosophic or religious conceptions, which are supposed to be unarguable in the audience's point of view.
- *Presumptions*: statements that the audience is likely to accept without a proof, but to which adherence may need to be reinforced. In Perelman and Olbrechts-Tyteca's words, *presumptions are connected with what is normal and likely* (NR § 17).

Premises relating to the "preferable" are peculiar to rhetorical argumentation, as they are associated with the specific viewpoint of the particular audience one is addressing. They consist of:

- *Values*: statements related to the preference for one thing as opposed to another;
- *Hierarchies*: the ways in which an audience arranges its values;
- *Loci*: very general statements which serve to justify values or hierarchies, for instance the *Locus of Quantity* (a thing is worth more than another for quantitative reasons) justifies values such as *Duration* or *Stability*, in contrast with the *Locus of Quality* which justifies values such as *Beauty* or *Goodness*. Perelman and Olbrechts-Tyteca do not want to provide an exhaustive list of *loci*, but they classify them under few very general headings: quantity, quality, order, the existence, the essence and the person. This classification lies in the importance of considerations relating to these categories in the actual practice of argumentation.

[3]The classifi cation will be elaborated and analysed in more detail in Ch. 4.

2.2.2 Argumentation Consists of Schemata

The force of an argument lies not only on the acceptability of its premises, but also on how the arguer presents premises to the audience. An important aspect of rhetorical argumentation is its communicative dimension: an argument's strength varies with its exposition. In consideration of this, the New Rhetoric is a collection of a number of *schemata*, that is ways to arrange premises and claims that are successfully used by people in ordinary discourse. This catalogue of techniques is ordered with respect to two wide classes:

Argumentation by association is applied when two elements that are separate in the mind of the addressee are put together by the arguer by using a particular relation. Different links give rise to different styles of schema:

1. *Quasi-logical argumentation*, using links that "give the impression" that the argument presented has a logical connection between its sub-parts. This is made by using an exposition structure which resembles a logical or a mathematical proof, such as: *You said that you only buy cheap things, apples are cheap, so you should buy apples!* It should be stressed here that the logical connection may be, and indeed in most cases is, only apparent, justifying the *quasi* connotation. In a perfectly logical argument, the consequence *must* follows from the premises (e.g. *All cats are mammals, lions are cats, therefore lions are mammals*), whereas in quasi-logical examples the conclusion does not necessarily hold, as it may depend on other, non considered circumstances. According to Perelman and Olbrechts-Tyteca, using elements of ordinary language in a logical relation makes the argument seem stronger, as the soundness of logical and mathematical thinking is well acknowledged.

2. *Argumentation based on the structure of the reality*, using links about the the reality that the audience perceives as real, such as cause/effect, object/attribute; group/constituents; means/end etc. The arguer, by expressing the relation, hopes to pass the audience's acceptance of one statement to the other, for example *Eating fruit helps slimming* encourages the audience to apply their (positive) view of slimming to that of eating fruit.

3. *Argumentation establishing the structure of the reality*, using *new* (for the audience) links between claims, such as by giving an example (*Fruit is very sweet: consider pineapples*), or by appealing to a model, especially to promote an action (*Healthy people have fruit for breakfast*). As in the previous class, positive or negative value is passed from one element to the other because of the association.

Argumentation by dissociation, on the contrary, introduces a division into a concept that the audience considered as a whole. This is not the same as opposing to association, or refusing to link two separate elements: in dissociation the audience is presented with just one element, which is then decomposed into two, “new” separate elements. For example: *You said that people who are concerned about diet are self-centred, but I prefer to consider them just responsible persons*. In this case a concept (*responsible*) is differentiated from another the audience considered as a whole (*self-centred*). Dissociation is usually done by introducing new terminology, so it is a “creative” process.

Examples of schemata are given in Fig. 1. The classification, proposed in Perelman and Olbrechts-Tyteca’s original essay as a collection of practical examples, together with philosophical considerations, was subsequently elaborated by Warnick & Kline (1992), with the aim of abstracting a collection of coding guidelines which could identify “as precisely as possible” the attributes of each schema, in order to make its recognition (and reproduction) easier.

2.3 Rhetorical Argumentation and Computation

Before proceeding with the description of our framework, we feel it is important to have a look at how the concept of argument has been used in computation, or more specifically in artificial intelligence. Very diverse are the applications, among which the most natural is perhaps to the field of Artificial Intelligence and Law (Bench-Capon 1997; Prakken 1997; Jones & Sergot 1993). The use of argumentation and negotiation has also become quite widespread in the model of artificial agents, to represent how they communicate, and possibly come to an agreement for the execution of either a common or a competitive task (Parsons & Jennings 1996; Karacapilidis & Papadias 2001; Sillince 1994; Kraus, Sycara, & Evenchik 1998). Argumentation has been used as a way to represent uncertain and defeasible reasoning, where the acceptance or otherwise of a proposition would entail more a weighing of arguments in favour and against, than a logical or probabilistic process of going from some premises to the conclusion (Amgoud & Cayrol 1998; Dung 1995; Krause *et al.* 1995; Vreeswijk 1993). Some other applications concern the use of arguments for collaborative learning (Lajoie *et al.* 1995; Pilkington *et al.* 1992; Ravenscroft 2000), or to improve expert system explanations (André & Rist 2000), in software engineering (Carbogim, Robertson, & Lee 2000), and so on.

Only a handful of researchers, however, have been undertaking works that focus on the rhetorical aspect of arguments, or on the structure of rhetorical discourse, and we will concentrate on these in this Section.

Argumentative Schemata	**Description**	**Example**
Quasi-Logical		
- Incompatibility	X and Y cannot coexist	You can either go out with your friends or see me tonight.
- Definition	Identity or Description of X	To call this a *warm* day it would have to be at least 25 degrees!
- Tautology	X = X	Boys will be boys.
- Reciprocity	X and Y should be given the same treatment	We all love those who love us.
- Transitivity	a R b, b R c, therefore a R c	Our friends' friends are our friends.
- Part/Whole	the properties of a set apply to its members	Tell me what sort of friends you have and I will tell you who you are.
- Whole/Part	the properties of a set member apply to the set	His behaviour disgraced all his family.
- Partition	one of the partitions must apply	If you don't like vegetables and pulses, you must get your Vitamin C from fruit!
- By the probable	X, though good, is very unlikely to happen	Don't stop here: this car park is always full at this time.
Based on the Structure of Reality		
- By Effects	X is OK, because of Y	He stole because he was hungry.
- Pragmatic	Y is a consequence of act X	Eating less fat helps slimming.
- Waste	opportunities and efforts should not be wasted	You have started this job, now you have to finish it.
- Of the Easy	X is an easy way to do Y	Eating fruit can simply be done by having a glass of orange juice.
- Persons and Acts	X is Y because he did Z	He is so mean: he didn't tip the waitress.
Establishing the Structure of Reality		
- Example	establishing a rule	Women are great interviewers: think of Oriana Fallaci.
- Illustration	reinforcing/explaining a rule	You should help your friends: when one of them is in trouble, stand with him.
- (Anti) Model	X is a person (not) to imitate	Italians know everything about healthy eating.
Dissociation	Term I vs Term II	You talk of law, I talk of justice.

Table 1: Examples of New Rhetoric's schemata

2.3.1 Discourse Processing

In talking about computational perspectives, and in consideration of the communicative dimension of rhetorical argumentation, we have first and foremost to refer to the field of computational linguistics, and in particular to those researchers who concentrate on the study of discourse structure.

In the field, three main theories emerge of organisation of discourse: the schemata organisation (McKeown 1985), the approach by Grosz & Sidner (1986) and Rhetorical Structure Theory (RST) (Mann & Thompson 1988).

McKeown (1985) defines schemata as “a representation of standard pattern of discourse structure which efficiently encodes the set of communicative techniques that a speaker can use for a particular discourse purpose” (p. 20). Apart from the computational slant, the main difference in the schema philosophy, when compared with Perelman and Olbrechts-Tyteca’s approach, is that McKeown’s schema comprises a whole set rhetorical predicates, “compiled” in a routine that can be used on the basis of a very general goal (e.g. “compare and contrast objects X and Y”). This, underlined by Moore & Paris (1992), makes all rhetorical processes transparent to the discourse generator, and therefore impossible to modify when the context is changed. This lack of explicitness makes, in fact, these schemata unsuitable for rhetorical argumentation, where one should be sensible to any change in the audience.

The approach of Grosz & Sidner (1986) is generally an analytical one, aimed at evaluating whether a given discourse, typically a dialogue, is coherent. The evaluation is based on three aspects of discourse: intentional structure, attentional structure, and linguistic structure. The approach postulates that, behind each discourse segment, identified by means of the linguistic structure, one should be able to see an intentional purpose, which relates to those of other discourse segments in one of two ways: satisfaction precedence (one must be satisfied before the other) or dominance (one is a sub-purpose of the other). The whole intentional structure cannot however be successful if one does not account for the attentional state of the hearer, that represents what is relevant at each stage. Perhaps because of the genericity of the relations described, this approach is mainly used for analysis rather than production of discourse, a task for which the most widely used approach is the RST.

The thesis of RST is that a text is coherent if and only if it can be recursively decomposed into pairs of non overlapping, adjacent spans, which are related one to the other by one of a set of rhetorical relations. A text can therefore be mapped unto a tree, whose root, with the rhetorical relation associated to it, represents its main purpose. The rhetorical dimension of the theory lies exactly in the definition of the relations: Mann & Thompson (1988) list a set of relations, found on the basis of empirical observations, for each of which they postulate a set of constraints that have to hold for the relation to be successful, summarising what the speaker and the hearer have to believe about the world. The two components of a rhetorical relation typically have an asymmetric role: one (the nucleus) is more fundamental than the other (the satellite) to the success of the relation itself. The set of relations is open, and has in fact being augmented by many authors (e.g. Hovy (1993) incremented the original collection into a taxonomy of about 120 relations), but it is not meant to cover all types of discourse: it is not possible, for instance, to represent "enveloping" structures like opening and closing (although Marcu (1997) defines *ad hoc* relations to capture them), "parallel" structures (e.g. comparisons), laws or poetry. Nevertheless, the theory is widely used for the analysis of hierarchically organised text, and has also had some variations able to account, for instance, for dialogues (Stent & Allen 2000). Moreover, the hierarchical hypothesis makes RST particularly suitable for automated generation of text by means of an AI planner (Hovy 1993; Moore & Paris 1993; Marcu 1996a; de Rosis, Grasso, & Berry 1997).

As pointed out by Reed & Long (1998), RST is however unsuitable for generating arguments, because of a series of problems. First, as also identified by Moore & Pollack (1992), RST does not adequately handle intentions, an important shortcoming for rhetorical argumentation, although this problem has been partially solved by Marcu (2000a) by coupling RST with intentions *à la* Grosz and Sidner. Second, RST considers as exceptional the cases in which nuclearity breaks down to give place to parallel structures, structures which are the rule rather than the exception in argumentation. Finally, RST does not account for argumentative relations as such, nor for high level organisations such as Modus Ponens, or, most importantly, for structured combinations of higher level units.

Computational Approaches

One of the earliest work in the computational study of argumentative discourse is perhaps the one by Cohen (1987), with an emphasis more on argument understanding than generation. The approach stressed the importance of a theory of coherence structure, as well as a theory of the relationships among speaker and hearer's beliefs, that should be coupled with the analysis of linguistic clues in order to be able to understanding arguments.

The research by Maybury (1993) focuses on the application of argumentation to a natural language generation system, representing argument as a series of communicative acts that have the purpose of performing some communicative goal. The aim of a communicative act is to affect the knowledge, the beliefs or the goals of the addressee. Different "argumentative" communicative act are identified, not surprisingly, in an Aristotelian fashion: "deductive arguments" are used to affect the beliefs of the addressee by providing a "proof" of a claim (such as the various kind of syllogism); "inductive arguments" are used to convince of some claim by providing some examples or evidences (such as illustrations, analogy, etc.);

whereas "persuading arguments" are used to affect the goals of the addressee (such as motivation, indicating the desirable consequences of an action, how the action is part of a goal, etc). The communicative acts are used in a hierarchical planner *à la* Sacerdoti (1977). Here, "plan operators" define the constraints and the preconditions that must hold before a communicative act applies, its intended effects on the addressee's cognitive state, represented by means of intentional operators expressing knowledge and desires, and how the act can be decomposed into sub-acts. However, in Maybury's work argumentation theory is used simply to define new communicative acts for generating a monologue, so no measure is possible to calculate whether the act was effective. Nor is there an accurate description of the addressee's cognitive state, necessary to trigger the appropriate act.

The focus of the system by Reed *et al.* (1996) is on the representation of *logical arguments*. A classification of beliefs, which we find fundamentally rhetorical, distinguishes among *factual beliefs*, testable or definitional, *opinions* based on moral and aesthetic judgement, and therefore personal and unprovable, and *cultural beliefs*, based upon "sociocultural maxims" as defined by the authors, but which can easily relate to the classical concept of *Loci*, or *Topoi*. The knowledge of the system consists of a set of *events* which can be related by means of a relationship of *support*: all the events, or perceived beliefs, which are supported by held events are themselves held. Events with no supporting events are called *grounded*: these can be due to external sources, internal motivations and value-systems (another rhetorical concept) or a priori beliefs about the ontology of the construction of the beliefs, which are necessary because both agents must agree on the description of the knowledge. An argument consists of one or more premises, which can themselves be sub-arguments, and exactly one conclusion. Premises can be linked to the conclusion by standard logical relations, e.g. rules of classical logic, by inductively reasoned implications, such as inductive generalisation, causal implication or analogy, or by rhetorical fallacies, which are however not fully investigated by the authors. An attempt is also made to structure the argumentative discourse by specifying the position that the statements comprising the argument hold in the sentence: "Conclusion-First", when the premises are examples, or when the conclusion is deliberately provocative; "Conclusion-Last", used for longer, more complex or less convincing arguments; or "Conclusion-Sandwich", typically used when the speaker has completed a Conclusion-Last argument which has not been accepted and therefore further support is required. The arguments are generated with an abstraction based planner (Fox & Long 1995), where the operators implement rules of logical inference, such as Modus Ponens or Modus Tollens, plus a few refutation rules (Reed & Long 1998). The approach is therefore dialectical, rather than rhetorical, in its nature, more focused on the way claims support each other than on the communicative, audience oriented dimension of the argument proposed. The work is however complete in its treatment, for instance, of ordering (Reed & Long 1997) and saliency

(Reed 1999).

Another important work to mention is the one by Elhadad (1995), important because it, like ours, relies on an established theory, namely Anscombre & Ducrot's "Argumentation Within Language" (AWL) (Anscombre & Ducrot 1983). AWL is, in fact, a linguistic theory, describing the ways in which sentences are expressed to support a set of conclusions rather than the objects used as premises. A fundamental notion is the "sentence argumentation orientation" which gives an indication of what conclusions can be supported by a sentence. Two sentences are said to have the same orientation if they lead to similar conclusions. This is independent from the content of the sentence itself. So for instance the sentence *Tom is as tall as Mary* can have the same orientation both of *Tom is tall* and of *Tom is short*, depending on which sentence follows describing the height of Mary. Elhadad showed how AWL could be used in several phases of the text generation process, from content determination and organisation, to lexical choice. With the aim of advising students about the courses to follow, the device of *topos* is used to express rules to perform *evaluations*. An evaluation is simply an answer to the question *How P is X?*, where P is a scale and X is an entity, and originates rules of the form *the more/less X is P, the more/less Y is Q*. A topos is then formed by four primitive relations:

1. an evaluation of the entity X on the scale P (for example a class on the scale of difficulty);
2. an evaluation of the entity Y on the scale Q (for example a student on the scale of desiring to take a course);
3. the expression of a gradual relation between P and Q (*the harder the class, the less desire*);
4. the expression of a topical relation between X and Y.

The topoi are used to select information from the knowledge base which could lead to the conclusion desired (to follow/not to follow a course). They are used in process of content organisation, interpreted as rhetorical relations, in that they explain the connection between two text spans. Topoi are also used in the lexical choice of determiners (*many, few* in the phrase *the AI course has many/few assignments*), adjectives (*diffi cult, interesting*), verbs (*require, enjoy*) and connectives (*but, therefore*).

de Rosis, Grasso, & Berry (1997) compared computer produced and human produced texts in a drug-prescription domain. Although the computer generated explanations reproduce the overall characteristics of the human texts, several elements, among which most importantly the argumentative techniques, contribute to increase the *argumentative strength* of the human messages, and are not reproduced in the artificial ones. The authors claim that an ideal solution to this problem would only be possible with a more fine grained representation of goals and beliefs of Speaker and Hearer, together with more sophisticated text planning techniques. However, a more practical strategy is proposed, employing a simple top-down planner to collect and order the information content of the explanation, and subsequently refining the plan by applying some rhetorical and argumentative strategies. No definition of effectiveness of an argumentation strategy in terms of changes in the beliefs of the user is given, as the argumentative rules are the mere application of empirical observations.

A major emphasis on argumentative text can be found in the work by Carenini & Moore (2001), who concentrates on "evaluative arguments", that is arguments advising on the appropriateness of a particular object or course of action. Arguments are tailored to users' values and preferences to the objects of discussion, so that a numerical measure can be obtained of how valuable an entity is to the user, a measure based on utility theory. Such a measure helps shape the produced piece of advice, by mentioning or otherwise features of the entity, changing the order of the features mentioned, or changing the emphasis of the evaluation (e.g. "good" as opposed to "excellent"). The actual generation of the piece of advice uses a pipelined architecture, quite a standard mechanism in natural language generation (Reiter & Dale 2000), including a text planner, aimed at organising the general structure of the text, a microplanner, shaping the text at a sentence level, and a sentence realiser, aimed at actually producing the sentence in a language. While the major emphasis is put on audience's modelling, the work does not however address the question of using rhetorical devices/schemata in the argument generation.

2.3.2 Generation of Arguments

We list here works that, despite a smaller emphasis on discourse production, are addressing the issue of producing persuasive (to an audience) arguments, and therefore can be included in the rhetorical category.

PAULINE (Hovy 1988) is one of the earliest systems generating natural language with a concern on *pragmatic* goals. These goals are meant to take into account the hearer's characteristics, the conversational setting, and the speaker and hearer's "interpersonal" goals. The characteristics of the hearer include, most notably, opinions, as opposed to knowledge, and emotional state, both towards the topic of discussion and towards the speaker. *Rhetorical goals*, as they are called in the system, are mainly used at the surface generation level, to help choose among different realisation of the same sentence (e.g. humorous, formal, etc.). They are, therefore, more style features than argumentative features. Higher level goals, like the strategies for topic collection when supporting the speaker's opinion, are simply based on general considerations coming from the analysis of various arguments in leaflets, newspapers, etc.

PERSUADER (Sycara 1990) is a model of persuasive argumentation, in a scenario representing the negotiation process in a settlement. Here assumptions are made which are similar to ours: the argument generation process is always guided by argumentation goals, aiming at changing the importance the *persuadee* attaches to concepts, and argumentation strategies are used to achieve them. The theoretical basis on which the strategies are defined, however, is not made explicit. An order is defined among a set of arguments in the labour mediation domain, according to their presumed strength, for instance a threat or a promise is meant to be the strongest argument, whilst an appeal to universal principles is described as the weakest one in this domain. The choice of the argument depends on the current goal of the persuader. This goal can be one of the following: (i) changing the importance of a goal/issue, (ii) changing the perception of an issue's value, or (iii) aiming at goal abandonment on the part of the persuadee via threats or promises. The latter goal is in our view a variation of (i) and (ii), as threats are promises are in fact meant to change the persuadee's priorities with respect to the issue in consideration[4].

Zukerman, Korb, & McConachy (1996) designed a system specifically aimed at analysing and composing argument. The system, given a proposition and a user model, is able to produce an argument supporting the proposition on the basis of the user model characteristics, in order to bring the user to an established degree of belief in the proposition. Especially interesting is the authors' definition of *nice* arguments, that is arguments that achieve a compromise between what is justifiable and what persuades, preferring what is believable to what is correct. The system can also analyse arguments proposed by the user, and agree with them or counter-argue. The architecture of the system is composed by four modules: (1) an argument strategist, (2) an argument generator, (3) an argument analyser and (4) a user interface. The argument generator takes as input the proposition to be argued for, the degree of belief to be achieved and a system *attitude*, which determine the degree of fairness of the system (it can go from "orthodox", which allows only normative inferences, to "licentious", which allows any action to increase the effectiveness of the argument). The output of this stage is an *argument graph* in a Bayesian network (Pearl 1991), which assigns to each inference a strength value, based on the conditional probability to achieve the given degree of belief. The analyser evaluates the argument graphs, guided by several parameters, such as the argument's normative strength (whether it is sound, or it is an inductive inference, and so forth) and its effectiveness (determined by the expected effect on the belief model of the user). The argument strategist select the approach for presentation of the argument, with the principal aim to make it as concise as possible; it selects the presentation method (for example, whether to present it completely or selecting branches of the graph; which style is the most appropriate, e.g. concessive or hypothetical, and so on); and it also selects the strategy to respond to the user's argument. Despite the attention to persuasiveness rather than soundness, however, the only concession granted to these *nice* arguments is their having some steps missing to their logic chains. Values and persuasiveness techniques are not considered here, so these arguments cannot be defined as "rhetoric".

Finally, it is worth mentioning the work by Saeedi & Sillince (1999), which implemented "rhetorical"

[4]See also Castelfranchi (2000) for a treatment on agent's internal and external conflicts.

rules, such as fairness, reciprocity and deterrent, to simulate argumentation and debate. The model is based on Toulmin's schema, and is meant to provide a framework to facilitate collaborative decision making. The rules have, however, a high degree of domain dependence, and are not defined in a way that encourage discrimination among them from case to case.

2.3.3 Representing Knowledge as Argument

A final aspect in the model of rhetorical arguments that is worth mentioning in our survey concerns the representation of knowledge. We will talk more extensively in Ch. 6 about knowledge representation, but here we would like to mention those systems which adopt a rhetorical view in modelling their knowledge structure, by explicitly representing preferences and values.

One of the earliest work on this respect is the SYNVIEW system (Lowe 1985), which proposes a method for collecting and synthesising large quantity of information and presenting them to a user in a manageable structure. Of particular interest from our point of view is the way SYNVIEW combines multiple viewpoints in a unique structure. The user is presented with an overview of topics related to a lead topic, ranked in order of their "importance" for a general understanding of the lead topic. A user can then suggest modification or addition at many levels. At the simplest level, a user can modify the attribution of the importance score. Or, the user can suggest a different verbalisation of the topic, or even add new subtopics. If a topic is a statement of a fact, this can open a structured debate. In a debate, the subtopics are further statements in support or against the topic of the debate. The explicit representation of arguments of a debate follows Toulmin's general structure, but is simpler: at the first level an argument is represented by a ranked list of topics for and against it, and there is no explicit consideration of the warrant.

Optimist (Clark 1990) was designed with the aim of helping geologists in oil explorations. The hypothesis of using a process of argumentation for this purpose is that "argumentation" among experts is useful to pool knowledge together and enforce consistency, and that, as experts frequently disagree, it is useful to collect evidence from similar cases which form a focus for a discussion. An argument is defined as an inference (or chain of inferences) whose validity is open to question. To reason about arguments, some meta-rules can be used, which are represented as meta-level knowledge. In order to represent different opinions, several knowledge-bases, labelled with the owner's name, are necessary. To keep the knowledge base manageable, a *skeleton + certainty sets* representation is used: a single rule-base is developed and then the body of the rules are separated from the certainty measures attached to them. In this way, several certainty values can be stored under different names.

More recently, Karacapilidis (1996) suggested an argumentation based framework, inspired by models coming from Game Theory and Operations Research. The basic elements of the framework are the *Positions, Issues and Arguments* of different agents during an interaction. Positions are any data the agents assert (declarations, justifications etc). Issues are the decisions to be made or the goals to be achieved. Arguments are assertions about the positions regarding their attributes that can be used for or against them. Comparison of alternative positions is made according to the value of *importance* for the agents. Preference relations are introduced between intervals of importance of arguments. As explained in (Karacapilidis, Trousse, & Papadias 1997), broader *discussion domains* are associated with every viewpoint of a case. The association proceeds “top down” from the more general topics to their subtopics.

A different use of viewpoints is the one by McCoy (1989), where perspectives are used to dynamically highlight selected attributes of a knowledge base. Similarly, Cawsey (1991) and Jameson *et al.* (1994) use a notion of perspective in order to provide “biased” information, where the bias consists in focusing the user’s attention on aspects of the knowledge base which are thought to be of particular interest, diverting attention from other ones the speaker wants to hide or underplay. In all these cases, therefore, a perspective is more a lens hiding portions of the knowledge base, rather than a rhetorical *locus* device, to attribute evaluations to concepts.

We will see in the next Section how an explicit representation of perspectives is fundamental in building a framework for rhetorical argumentation.

2.4 A Model of Rhetorical Argumentation

In this Section we present a characterisation of rhetorical argumentation based on the New Rhetoric. We should perhaps stress immediately that the formalisation, albeit given in logical (or quasi-logical?) terms, is by no means advocating a logical approach to rhetoric, as this would miss the point of a rhetorical argument altogether. The effort is only aimed at providing a well-defined semantics for our concepts. In fact, the logicalisation of rhetorical argumentation is only part of the model: in this Section, after providing a definition for rhetorical argument in Sect. 2.4.1, we will add to this definition its rhetorical dimension by coupling it with the concept of a rhetorical schema, in Sect. 2.4.2. The notion of schema will of course draw from Perelman and Olbrechts-Tyteca’s definition, but will also be inspired by theories on discourse structure, as rhetorical argumentation cannot ignore language and discourse as its prime means for being conveyed.

2.4.1 A Formal Model of Rhetorical Argument

Before giving our model for a rhetorical argument we need to introduce the first class objects of our language, and in particular, after a general hypothesis on the content of a domain ontology, we introduce a notion which is peculiar to rhetorical argumentation, and to Perelman and Olbrechts-Tyteca's New Rhetoric, that is the notion of *value*. This notion, together with the one of *perspective* will serve to build up the definition of a rhetorical argument. We will then conclude with an example of what it means to argue rhetorically, showing the advantages and the idiosyncrasies of our model.

Objects of Discourse

We will assume that the argumentation process can take place on a certain set of objects of discourse, forming an ontology. We will see in Ch. 6 a more formal definition of the ontology we need, together with the explanation of the very idea of having an ontology, but now it will suffice to say that we need a set of concepts, C_o, of an ontology O, and a set of relationships among these concepts, R_o, again in the ontology O. We think of a relationship, in this context, as any of the possible ways in which two concepts may be linked in the ontology: one might be the generalisation of the other, they may both contribute to define a third object, they may be an action and one of its roles, they may have identical values for one of the attributes, and so on and so forth. We express the fact that there exists a relationship r_k among the objects c_i and c_j in an ontology O by writing that:

$$r_k(c_i, c_j)$$

with

$$r_k \in R_o; c_i, c_j \in C_o$$

For the sake of simplicity in the notation, we will consider, without loss of generality, only binary relations, but the definitions below can be naturally extended to consider n-ary relations.

Values and Perspectives

We have seen above that a rhetorical argument is, by and large, a means to *pass value* from one object of discourse to another: a good value if arguing in favour or a bad one if arguing against. The notion of value is therefore crucial to our formalisation. In order to argue rhetorically, we need to express in some way that "X has good value" or "X has bad value", where X is a concept, as one of our primitive notions. In fact, we want also to be able to express "how" X has good or bad values, that is we want to specify a *perspective* from which this happens. A perspective is a way to express what Perelman and Olbrechts-Tyteca, and indeed classical philosophers, call a *locus* of argumentation, indicating the fact that there are some *fi elds*, or points of view, that help the process of evaluation of a given concept. As an example, we would not want merely to say that "eating apples is good", but rather that "eating apples is good from the health perspective", therefore capturing the fact that, for the same agent, eating apples can also be bad from another perspective (for instance because they do not taste good).

Definition 1 (Evaluation) *We say that there exists an* **evaluation** *of a concept c in the set of concepts C_o of an ontology O from a certain perspective p, from a set P_o again in the ontology O, if there exists a mapping E of the pair (c, p) into the set V of values. Assuming that V is a set consisting of two elements representing "good" and "bad", we write this as:*

$$E(c, p)$$

where E is the mapping:

$$E : C_o \times P_o \rightarrow V$$

and:

$$V = \{good, bad\}$$

Again, for simplicity here we will assume a dichotomy of values, but the framework is obviously extensible to a finer grained definition of values.

Rhetorical Argument

If rhetorical argumentation aims at reaching an evaluation of an object or state of affairs, then a rhetorical argument is a way to "pass value" from one topic to another, in the same way as a deductive argument "passes truth" from one proposition to another. So, as a deductive argument expresses something like: "if X is true, and Y derives from X, then Y is true", a rhetorical argument expresses something like "if X has value, and Y *is related to* X, than Y has value". Unlike deductive argumentation, the way in which the argument is presented to the audience forms an important aspect of the argument itself. Therefore, in order to be able to control that the value transfer across objects makes sense, we not only impose that a relationship should exist between the concepts involved, but also that the value transfer should take place according to a pre-defined schema.

Definition 2 (Rhetorical Argument) *We defi ne a rhetorical argument as the act of putting forward the evaluation of a concept, on the basis of a relationship existing between this concept and another concept, and by means of a rhetorical schema.*

If we call S the set of the available rhetorical schemata, and E' the set of all the evaluations, that is:

$$E' = C_o \times P_o \times V$$

then we define a rhetorical argument A_R as a function:

$$A_R : E' \times R_o \times S \rightarrow E'$$

In other words, if we have a concept $c_i \in C_o$ and an evaluation of such concept $E(c_i, p_i)$ from a given perspective $p_i \in P_o$, we can put forward a rhetorical argument in favour or against a second concept $c_j \in C_o$, if a relationship exists between the two concepts $r_k(c_i, c_j), r_k \in R_o$, by means of a rhetorical schema $s_l \in S$. The argument will result in the evaluation of the second concept (not necessarily with the same orientation) from the same perspective $p_i \in P_o$:

$$A_R(E(c_i, p_i), r_k(c_i, c_j), s_l) \rightarrow E(c_j, p_i)$$

The sign of the evaluation (good or bad) will depend on the particular relationship exploited, as well as on the use of the schema. Therefore, we will have orientation preserving relationships among concepts, let us call them R^+, and orientation changing relationships, let us call them R^-. Supposing for the moment a neutral impact of the schema s_l, we have therefore that:

$$r_k \in R^+; A_R(E(c_i, p_i), r_k(c_i, c_j), s_l) \rightarrow E(c_j, p_i) = E(c_i, p_i)$$

and:

$$r_k \in R^-; A_R(E(c_i, p_i), r_k(c_i, c_j), s_l) \rightarrow E(c_j, p_i) = E(c_i, p_i)^{-1}$$

where

$$R^+ \cup R^- = R_o$$

and $E(c, p)^{-1}$ is meant to be the complementary evaluation of $E(c, p)$, that is, in the case of a dichotomy of values, like Good and Bad:

$$E(c, p)^{-1} = V - \{E(c, p)\}$$

Rhetorical Argumentation

Rhetorical argumentation can typically be more complex than one single evaluation. Rhetorical arguments can be chained, by exploiting relationships among the concepts involved. Therefore, if a rhetorical argument can be put forward on a concept c_1 on the basis of another concept c_2, and another rhetorical argument can be put forward on a concept c_3 on the basis of c_2, we can combine the two arguments to generate an evaluation of c_3 on the basis of c_1:

$$A_R(E(c_1,p),r_1(c_1,c_2),s_1) \oplus A_R(E(c_2,p),r_2(c_2,c_3),s_2) \rightarrow E(c_3,p)$$

The argumentation above basically involves the concatenation the two schemata to create a chain of reasoning, on the bases of the two relationships identified in the ontology of concepts.

In general, we can produce rhetorical argumentation in favour or against a concept $c_n \in C_o$ from a certain perspective $p_n \in P_o$ if we can construct a chain of rhetorical arguments:

$$E(c_{n+1},p) = \bigoplus_{i=1}^{n} A_R(E(c_i,p),r_i(c_i,c_{i+1}),s_i)$$

where

$$c_1,\ldots,c_{n+1} \in C_o; p \in P_o; s_1 \ldots s_n \in S; r_o \ldots r_n \in R_o$$

Rhetorical Reasoning: a Simplified Example

It is perhaps evident that the rhetorical reasoning produced with the mechanism described above will not necessarily always produce "valid" arguments, where valid in this context means not fallacious. In fact, a somewhat arbitrary chain of arguments, whose only restriction lies in the identified ontological links among their main objects, may, in extreme cases, produce fallacious conclusions.

An example will illustrate below both the main idea of arguing rhetorically, and the extreme cases that this kind of reasoning may generate.

Let us assume we have a, purposely very unsophisticated, and hopefully self-explanatory, knowledge base expressing concepts relating to dietary attitudes, like the one in Fig. 2. Here we wish to make the least possible assumptions about the knowledge representation: we only assume that there are concepts (represented as boxes in the figure) and there are labelled relationships among these concepts (represented as arrows connecting two boxes with a label on them). We maintain here our limitation to binary relationships only. The picture, therefore, represents, among other things, that "antioxidants prevent cancer", that "eating fruit can never be done at pubs", "eating fat prevents slimming", and so on.

Now, let us assume we can associate evaluations to concepts on the basis of a set of given perspectives, for instance let us assume that one can say that "cancer" and "stroke" are "bad" from a "health" perspective, and that "good self image" is good from a "social life" perspective, while the concept of "pub" is "good" from the "social life" perspective.

We need also to enumerate the relationships that are orientation preserving or otherwise, and we can say that "Achieve", "Favour", "MoreGeneral" are in R^+, while the remaining relationships are in R^-.

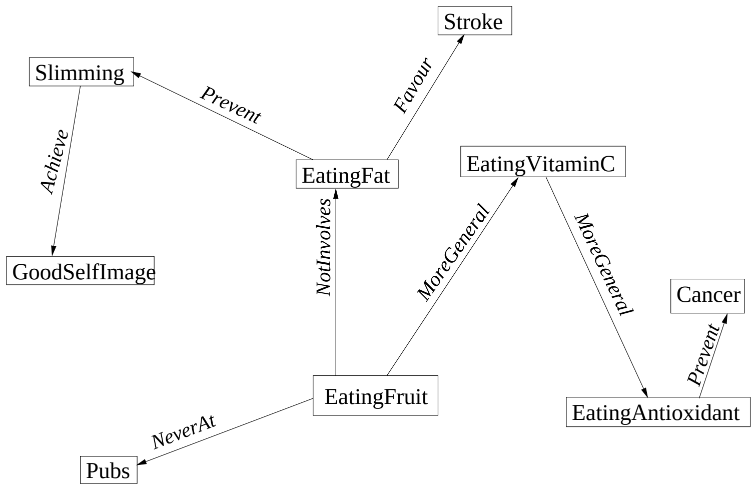


Figure 2: A (rough) example of a knowledge base

Now we can create rhetorical arguments, that add new evaluations, on the basis of what we have. For instance from:

$$E_{\mathit{Stroke},\mathit{health}} = \{\mathit{bad}\}$$

we can create the argument (let us ignore for the moment the problem of choosing an argumentative schema):

$$A_R(E_{\mathit{Stroke},\mathit{health}}, \mathit{Favour}(\mathit{EatingFat}, \mathit{Stroke}), s_l) \rightarrow E_{\mathit{EatingFat},\mathit{health}}$$

In particular, given that $Favour \in R^+$, we can say, assuming that the schema has a null impact on the argument, that:

$$E_{\mathit{EatingFat},\mathit{health}} = \{\mathit{bad}\}$$

By concatenation of rhetorical arguments, we can create evaluations of any object in the knowledge base, by propagating those of the neighbours. Fig. 3, for example, shows four possible rhetorical argumentations for the concept EatingFruit. We can say in fact that:

1. Eating fruit is good from the health perspective because Eating Fruit is a way for (is more general than) eating Vitamin C, which is a way for eating antioxidants, which prevents cancer, which is bad from the health perspective.

2. Eating fruit is good from the health perspective because Eating Fruit does not involve eating fat, which favours stroke, which is bad from the health perspective.

3. Eating fruit is good from the social life perspective because Eating Fruit does not involve eating fat, which prevents slimming, which helps achieving a good self image, which is good from the

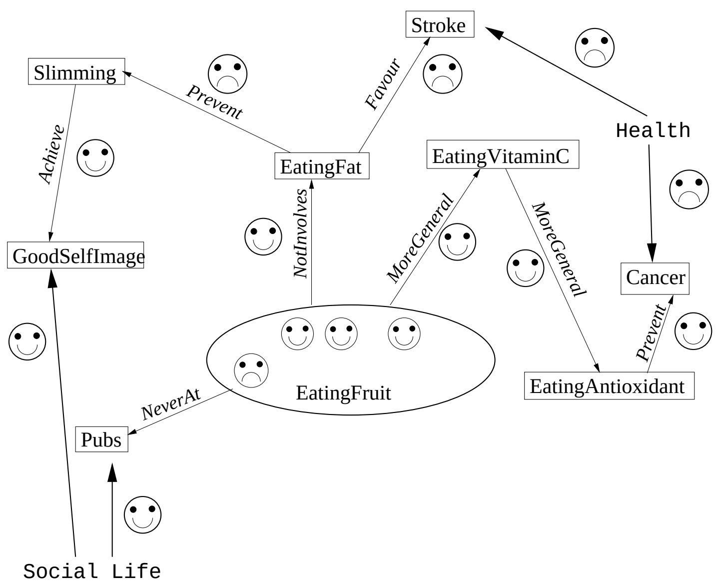


Figure 3: Rhetorical arguments derivable from the knowledge base in Fig.2

social life perspective.

4. Eating fruit is bad from the social life perspective because it cannot be done at pubs, which are good environments from the social life perspective.

The example shows two important things. First of all, that it is perfectly possible for a concept to have two opposite evaluations, even from the same perspective. In the knowledge base shown, we have that eating fruit is both good and bad from the social life perspective, depending on the line of reasoning used. We regard this as a positive effect of our model: people can have contradictory opinions about one subject, depending on the point of view, and in many cases decisions are more a weighing of positive and negative views than a deductive judgement.

The second consequence of our formalisation is that the longer the argumentation chain, the weaker the argument itself tends to be. With this model, a sort of “sophistic arguments” can be obtained which are unlikely to convince anyone. For instance one can say, in the example, that pubs are bad from the health perspective, as you cannot eat fruit there. And while we could imagine someone loosely agreeing with this, it is definitely more difficult to maintain that stroke is good from the social life perspective, on the grounds that it is prevented by eating fruit, which is something that cannot be done at pubs. While we concede that these arguments would be considered as fallacies by most argumentation theorists, we are also convinced that they should not be banned altogether: even if fallacious arguments, they are arguments nevertheless, and we rather have a formalisation in which they are included. In implementing the formalisation, heuristics will ensure these arguments are not preferred, or are not preferred too often.

As a matter of fact, we believe it is interesting to account for these arguments as well, in the spirit of producing arguments which are similar to those people use (and these are definitely used, even if just in extreme situations).

2.4.2 Rhetorical Schemata

In our vision of rhetorical argumentation, while it is important to be able to reason about values by referring to the arrangement of objects in the knowledge base, major emphasis should be put on the way arguments are presented to the audience. This is rendered in the formalisation above by not allowing a rhetorical argument to be used without explicitly associating it to a schema. This means that, even if a relationship exists in the knowledge base that would allow one to pass value from one object to the other, this cannot be done as such, but if and only if a schema can be found that exploits the relationship, and presents it to the audience in a given format.

The intended meaning of the schema is therefore to try and capture the restrictions that, from case to case, have to apply for a given relationship to be used legally. By definition of rhetorical argumentation, these restrictions cannot be fixed once and for all, but vary from context to context, and more importantly, from audience to audience. The recent developments of argumentation theory, and in particular of informal logic, show plenty of examples of patterns of reasoning that used to be considered fallacious in the past, but are recognised now as admissible in some circumstances. One of such examples is the so called "slippery slope" argument, one famous instance of which is S. John Chrysostom's argument against laughter. The slippery slope argument makes an extensive use of the cause/effect relationship, by showing how one event or situation can trigger a chain circumstances culminating in a final event or situation, which has to be clearly perceived as good or bad. A value can then be attributed to the first event/situation in the chain on the basis of this final effect. A few interesting examples of this and other in principle fallacious arguments, are given by Groarke, Tindale, & Fisher (1997), together with an evaluation on their quality, or "admissibility". The definition of the circumstances that make these arguments schemata good or bad is somewhat blurred, but the important point here is that such circumstances do exist, and can be defined in some way. What is clear is that such definition cannot be done *a priori* and for each type of context, but has to be established whenever the context changes. We could easily assume that in some special circumstances, for some communities, or even just for the sake of a polemic, for instance in an argumentation game, slippery slope arguments will always be acceptable.

It is beyond the scope of this exercise to establish the set of constraints that need to apply in the schema definition for each relationship, and in this sense our definition of rhetorical schemata differs perhaps from what Perelman and Olbrechts-Tyteca intended, and certainly differs from other approaches, such as the Rhetorical Structure Theory (Mann & Thompson 1988), to the definition of discourse schemata. Our sole claim is that it is important that a definition of a rhetorical schema should express the schema's admissibility, together with the definition of the relationship that the schema is meant to formalise, and we propose here an ontological discussion of this concept of admissibility, and what it should entail.

Admissibility Parameters for Rhetorical Argumentation

Literature on critical thinking and informal logic (see Groarke, Tindale, & Fisher (1997) or Walton (1989) for example) mentions three important notions in the evaluation of a good argument. We have already introduced the concept of *acceptability* of an argument, as opposed to its validity or soundness: an argument is acceptable if the audience is ready to commit to it. Acceptability is a subjective property: the arguer does not appeal to incontrovertible truths, but rather to what the audience believes to be true, or what a "general", or, in Perelman and Olbrechts-Tyteca terms, "universal" audience would believe to be true. Conditions on the acceptability of a schema have therefore to be based on the beliefs of the participants to the argumentation process, both the arguer and the audience.

But acceptability alone does not suffice to make a good arguments: informal logicians also mention the concept of *relevance* (Groarke, Tindale, & Fisher 1997; Tindale 1994; Woods 1994), stressing that, because of relevance problems, arguments which could otherwise be acceptable have to be rejected. The notion of relevance, in fact, gives rise to the definition of some fallacies, for example Aristotle's *ignoratio elenchi*, or ignoring the question, where an argument is produced to prove a conclusion different from the one under dispute. The problem or relevance is therefore not so much the one of producing arguments which will not be believed by the audience, or which do not exploit the audience's perceptions. Nor it is a problem of producing sound arguments: irrelevant arguments can be sound, and unsound arguments can be relevant.

Finally, good arguments should propose sufficient elements for the audience to be enabled to reach an evaluation (Groarke, Tindale, & Fisher 1997): by adhering to the principle of *suffi ciency*, an argument should put forward a not biased and balanced point of view.

Acceptability, Relevance and Sufficiency, the "RSA-triangle" as called by Johnson & Blair (1994a) are the three criteria for a good argument: a complete model of a schema for rhetorical argumentation should therefore comprise all these elements.

Definition 3 (Rhetorical Schema) *We define a schema for rhetorical argumentation as a 6-tuple, as follows:*

$$R_S = \langle N, C, O_c, A_c, R_c, S_c \rangle$$

where:

- N *is the name of the schema,*
- C *is the claim the schema supports,*
- O_c *are the ontological constraints the schema is based on,*
- A_c *are the acceptability constraints,*
- R_c *are the relevance constraints, and*
- S_c *are the sufficiency constraints.*

Examples of schemata are given in what follows, but first let us discuss the rationale behind these admissibility constraints.

The separation of the admissibility constraints in different categories is not solely due to clarity, but has also an operational meaning. The order in which the constraints are defined in the schema represents also a ranking of importance of the desiderata for the successfulness of the schema itself. In a sense, we may say that it should be more important, or more fundamental, to satisfy constraints in the first slots than it is to satisfy those in the last slots. Or, in other words, that arguments which do not satisfy the last slot are of poorer quality, but argument which do not satisfy the first slot cannot exist.

Before examining the constraints in turn, we would like to stress that we are defining desiderata, rather than obligations, and that one could imagine, and we will do, a context in which, for various reasons, some or all of these constraints could be relaxed. So we assume that "good" arguments are those for which the complete set of constraints holds. Lack of knowledge about what the audience believes, or a deceptive attitude on the arguer's side, may lead to the relaxation of some of the constraints, and the consequent reduction of the argument's quality.

Ontological Constraints

The ontological constraints represent the characterisation, in the language of the particular knowledge base in use, of the r_k element in the tuple defining rhetorical arguments (Def. 2), that is they spell out how the two concepts that are used in the rhetorical argument should relate to one another in the ontology. They should represent, in an ideal scenario, the condition *sine qua non* for the application of the schema: if the knowledge model that the arguer is using defines no such relationship between the two concepts, the schema itself cannot be applied. In particular, imposing the satisfaction of the ontological constraint means that the arguer cannot create relationships *ex nihilo*, while its relaxation means that deception is allowed. It is however important to mention that the ontological constraints, as we envisage them, do not express absolute truths, but are only a projection of the arguer's view on reality. The arguer is in a sense a cognitive agent relying entirely on its knowledge base: we do not assume a situation in which a third, omniscient party can judge the veracity of what the arguing parties say. An honest or deceptive behaviour, therefore, will not necessarily entail a true or false argument as such.

Acceptability Constraints

Second in order of importance, the acceptability constraints refer to the audience's set of beliefs. We have maintained in several occasions, together with rhetoricians, that rhetorical argumentation is based on what the audience perceives as acceptable, therefore not taking this into account, albeit leading to perhaps valid arguments, is prone to failure: valuing a red Ferrari on the basis of its colour, while knowing that the audience dislikes red, would just be missing the point altogether. Still, these constraints are not strictly necessary for the rhetorical schema to be applied: the evaluation of the Ferrari can make sense, and can even be accepted to the audience, who might believe that the arguer, in turn, likes red, or that red is a commonly appreciated colour for cars. In fact there will be even cases in which this slot is empty, for example, when arguing by "establishing the structure of reality", the arguer may communicate concepts which do not need to be necessarily known to the audience. Therefore, relaxing the acceptability constraints would in general simply involve ignoring the audience's belief, but would not in principle lead necessarily to false or unsuccessful arguments as such.

In modelling the audience's beliefs, all the considerations typical of agent cognitive modelling apply, as we will see in Ch. 4, but here we want to stress on the distinction made by Perelman and Olbrechts-Tyteca between "facts" and "presumptions". In the hypothesised absence of an objective, omniscient source, we interpret it as the distinction between what the audience has explicitly committed to, and what the arguer can only presume is believed. In checking the A_c constraints, the speaker should therefore give different weight to their lack of satisfiability depending on how they have been acquired. This accounts for what, in Perelman and Olbrechts-Tyteca's view, the speaker does when assessing tokens of adherence to the objects of discussion: it may be done explicitly (like the dialectical use of question and answer in Socratic dialogues) or may be assumed. The rules for making these assumptions are of course not clearly cut: if the audience has in one occasion claimed to like red, can one assume that the Ferrari's colour will receive a good evaluation? It is not reasonable to think that one would go through every possible instances of red objects before expressing an inclination towards red. The argument pro Ferrari could then receive a lukewarm acceptance. But, the assumptions on the audience's likes and dislikes can also be made with a malicious aim: a previously expressed propensity for red could be exploited by the arguer to use an audience's bad evaluation of red cars as a way to undermine the audience's reliability when expressing their own preferences.

We will explore the problem of representing belief nestings in more detail later on in Chapter 4, but for the time being we assume to express the acceptability constraints by means of a meta-predicate, β, which stands for “the audience believes”, and has as argument an expression of an ontological constraint.

Relevance Constraints

We define the relevance constraints as the representation of which premises have to be communicated for the audience to act towards the acceptance (or rejection) of the conclusion.

There are many theories and interpretations of relevance in artificial intelligence, in philosophy and linguistics/pragmatics (Dascal 1979; Grice 1975; Greiner & Subramanian 1995; Walton 1982), with various levels of formality, but we prefer to appeal to a more informational notion, as advocated by the theory of Sperber & Wilson (1995). That is, we think of these constraints, together with, for example, Reed (1999) or Walker (1996) as representing the “effort” put by the audience in reaching the conclusion.

In their approach, Sperber and Wilson see human communication as a way to produce some “effect” in the cognitive state of the interlocutor, a process that requires a mental effort on the interlocutor’s side. The authors then define a notion of relevance[5] stating that:

- the greater the cognitive **effect** produced by a piece of information, the **more** relevant that piece of information is;
- the greater the **effort** involved in processing a piece of information, the **less** relevant that piece of information is.

Paraphrasing Sperber and Wilson, we define relevance constraints as the beliefs that the audience should not only possess, in order to process the implication, but also have “accessibility” to, that can be retrieved without great effort. In a discourse modelling jargon, we can summarise this by saying that relevance constraints represent the beliefs that have to be in the “focus of attention” (Grosz & Sidner 1986), that is the collection of the entities that are salient to what is being said. In the example of the Ferrari dealer above, if the arguer says “if you like cherries you should like Ferrari too”[6], the assumption is that the redness of cherries and Ferrari is both known and immediately evident to the audience, as lack of knowledge will fail to produce the desired cognitive effect, and lack of availability will increase the audience’s effort to process this message. If this cannot be assumed, the good arguer should first put the

[5] Interestingly, the authors used this notion of relevance in explaining core rhetorical phenomena, such as tropes and figures of speech (Sperber & Wilson 1990).

[6] The reader who finds this argument not particularly appropriate could try and substitute cherries with lipsticks and have a look at a random set of TV adverts.

claim in focus, perhaps by saying "Ferrari cars have the colour of cherries", which in turn assumes that the audience knows which colour this is.

The relevance constraints may be misused just as the ontological constraints, leading to misleading arguments, especially in situations in which the argument suggests that a piece of data is common knowledge. This happens, for instance, with *Modus Brevis*, syllogistic arguments where the middle term is omitted (*All men are mortal, so Socrates is mortal*) on the basis that is already evident to the audience, so much so that mentioning it would be so redundant to irritate. If the omitted piece is not common knowledge, or, even worse, it is false, we range from cases of ineffective arguments to cases of more or less blatant deception.

As far as our formalism is concerned, while acknowledging that a greater sophistication might be desirable (for instance, we do not specify here the rules to manipulate the focus), we represent the relevance constraints as another meta-predicate, which we denote Φ, whose meaning is "the audience has in focus", and whose argument is an expression of an ontological constraint.

Sufficiency Constraints

Sufficiency constraints are defined in the informal logic literature as the representation of all the premises that are needed in order for the audience to establish that the conclusion is more likely than not (Groarke, Tindale, & Fisher 1997). They are meant to measure the speaker's "fairness" in putting forward the argument: in support of the point made, the speaker should have a well balanced list of issues that lead to the proposed conclusion. The argument pro Ferrari, in other words, cannot, and should not be based only on its being red. In non deductive argumentation, it is not always possible or feasible to give precise rules to establish whether sufficient premises have been presented (Groarke, Tindale, & Fisher 1997), and this slot is likely to vary from argument schema to argument schema, and, within the same schema, with different ontologies of concepts. A precise logical characterisation of *all* is needed to satisfy this constraint is not feasible here, nor it is in our intentions. We adopted a circumscription approach by encapsulating in the notion of sufficiency Perelman and Olbrechts-Tyteca's list of counterarguments to each schema, that is the expression of the set of circumstances that can weaken the argument proposed. However, we do this from a positive perspective, by listing instead the beliefs that complete the line of reasoning of the schema, and that can be provided if needed in support to the main point, but that will weaken the argument if they cannot be taken to hold.

As for the other constraints, ignoring the sufficiency constraints, may lead not only to weak arguments, but also to fallacious ones, and can be thought as a form of deception, however in good faith it might be[7].

2.4.3 Examples

We will present in this Section three examples of rhetorical schemata, formalised according to the framework described so far. Other schemata are included in Appendix A. The schemata presented here constitute one example from each of the three main Perelman and Olbrechts-Tyteca's categories: quasi-logical, based on the structure of reality and establishing the structure of reality. The fourth category, dissociation schemata, is not treated in depth in the present thesis, and only one example of such schemata is given, in the Appendix. Indeed, the dissociation class represents a big challenge for the modeller, and can be thought of a subject for a thesis in its own right.

In all the examples we will not commit ourselves to a particular knowledge representation, but we will assume that the system can rely on a knowledge representation in which one can represent a Concept (or Class, or Entity), c, for which one can enumerate a certain set of Attributes (or Predicates, or Slots, or Roles[8]), $c.a_1 \dots c.a_n$, and a set of Relations, $r_1 \dots r_k$ that relate it with other Concepts in the knowledge representation.

We indicate the semantics and the value of an attribute of a concept c respectively as $\mathcal{S}(c.a_j)$ and $\mathcal{V}(c.a_j), \forall j$.

In what follows we will use the following notation:

- *variables* will be denoted by lower case words: *x, y, concept, act, perspective, val*;
- *constants* will be denoted by capitalised words: *Cookies, Good, Health*;
- *predicates* will be denoted by upper case words: *LEADSTO, GENERALISE*;
- for economy of notation, we will also abbreviate $E(c, p)$ as $E_{c,p}$.

Quasi-Logical Argumentation: the RECIPROCITY SCHEMA

In the New Rhetoric, the argument of reciprocity is based on the concept of symmetry. The aim is to propose that two situations should be given the same evaluation on the grounds of some symmetrical property that they share (NR § 53). As in other quasi-logical arguments, such symmetry does not necessarily have to be equivalent to the formal property of relations (that is the fact that if $a\mathcal{R}b$ then $b\mathcal{R}a$,

[7] A common example could be the one of a doctor who avoids mentioning to patients the side effects of a particular drug, a much studied situation in patient education (De Carolis *et al.* 1996).

[8] While we are aware of the ontological differences between these notions we can assume without loss of generality that they can be treated similarly in this context, so we do not differentiate among them explicitly.

where $\mathcal{R}$ is a relation), but it rather expresses a more or less objective interchangeability of two states of affairs. This can happen for example because of some complementarity (e.g. between the acts of accusing and condemning) or because they can be thought of as one the inverse of the other (e.g. borrower and lender) and so on.

An example or reciprocity argument can be: *You said that you like cookies. But apples are crunchy like cookies, so you should like apples too!* The verbalisation does not have necessarily to be as above, and in fact the argument will typically be put forward in a more concise way. Let us try and understand how.

Claim. The schema aims at passing good or bad value on to one concept for a certain perspective:

$$E_{c,p} = val$$

In the example this would instantiate to passing good value on to the concept of apples (or eating apples), from a certain perspective, p:

$$E_{Apples,p} = Good$$

Ontological constraints. The O_c element of the 6-tuple for the reciprocity schema is based on the analysis of the attributes that two concepts, or two instances of a concept, share. As this is a quasi-logical schema, we almost never will pay attention to the actual meaning of the relationships: it will suffice that a relationship exists with some properties. A relation of reciprocity can be defined to hold between two concepts when for each of them an attribute can be identified having same semantics and same value. That is, if for two concepts,

$$c_1, c_2 \in O_c$$

we have that

$$\exists\, k, j : \mathcal{S}(c_1.a_k) = \mathcal{S}(c_2.a_j)$$

and

$$\mathcal{V}(c_1.a_k) = \mathcal{V}(c_2.a_j)$$

then an argument of reciprocity may claim that the two concepts should have the same evaluation on the basis of this symmetry.

The specific argument depends of course on the particular content of the chosen knowledge base.

In the example, we can assume we have a knowledge base in which both cookies and apples have an attribute “texture” (we assume they have the same name, but this is not necessary) with same semantics (assuming values in the same domain), and same value “crunchy”:

$$O_c : \{\mathcal{S}(\text{Apples.Texture}) = \mathcal{S}(\text{Cookies.Texture}); \mathcal{V}(\text{Apples.Texture}) = \mathcal{V}(\text{Cookies.Texture}) = \text{Crunchy}\}$$

Acceptability Constraints. In order for the argument to be successful, the audience should share the view that the “reciprocal” concept used in the argument has the desired evaluation, from a certain perspective:

$$\beta(E_{c_2,p} = val)$$

In the example, we need the audience to like cookies, for instance from a “Taste” point of view:

$$A_c : \beta(E_{Cookies,Taste} = Good)$$

Note that the argument does not ask that the audience should like cookies *because they are crunchy*: this would entail a level of knowledge of the reality that is not requested by quasi-logical argumentation, as it is by the other two categories. In order to express a quasi-logical argument one only need to create a “skeleton” that aims at making two concepts comparable (NR § 45).

Relevance Constraints. The conclusion the argument proposes will be reached with little effort by the audience if all of the elements involved are in the focus of attention:

$$\Phi(\mathcal{S}(c_1.a_k) = \mathcal{S}(c_2.a_j)); \Phi(\mathcal{V}(c_1.a_k) = \mathcal{V}(c_2.a_k)); \Phi(E_{c_2,p} = val)$$

In the example, the crunchiness of apples and cookies, and the tastiness of cookies should be in the focus of attention:

$$R_c : \Phi(\text{Apples.Texture} = \text{Crunchy}); \Phi(\text{Cookies.Texture} = \text{Crunchy}); \Phi(E_{\textit{Cookies, Taste}} = \textit{Good})$$

If this can be assumed, one can safely use a *Modus Brevis* form. If this cannot be assumed, that is the argument is put forward as *if you like cookies, you should like apples*, the audience will be left bemused. This, however, does not necessarily mean the audience will not be convinced, as it happens when the constraint is relaxed maliciously, for instance when the arguer can assume to have a strong influence (as it is in a situation of peer pressure).

Sufficiency Constraints. These constraints should enable the audience to reinforce the symmetry between the two concepts, as opposed to it being limited to one attribute only. An exhaustive list of what should be included here is perhaps not feasible, and it definitely depends on the underlying ontology, but we can assume that an appropriate constraint could require that other pairs of attributes of the two concepts in question can be identified sharing the same semantics and the same value. Of course, the more attributes with equal semantics the two concepts have, the more they can be compared to each other, and the more pairs of comparable attributes have equal values, the stronger the reciprocity relation will be. Measures of similarity among concepts can be established in the ontology to have a sounder definition of the reciprocity relation, but for the moment we can assume a sufficiency constraint asking that:

$$\textit{Most } j,k : \mathcal{S}(c_1.a_j) = \mathcal{S}(c_2.a_k) \rightarrow \mathcal{V}(c.a_j) = \mathcal{V}(c_2.a_k)$$

in the example:

$$S_c : \textit{Most } j,k : \mathcal{S}(\text{Cookies}.a_j) = \mathcal{S}(\text{Apples}.a_k) \rightarrow \mathcal{V}(\text{Cookies}.a_j) = \mathcal{V}(\text{Apples}.a_k)$$

where *Most* should be read as “the more attributes there are satisfying the condition, the better the argument”.

As a remark, we should underline again that, as the schema is quasi-logical, no emphasis is put on the set of attributes, or roles, that are more meaningful to the success of the schema. In the example, an attribute referring to the “edibility” of both cookies and apples is clearly crucial: a reciprocity argument based on the crunchiness of apples as opposed to, say, gravel will definitely miss the point altogether.

The schema for the reciprocity argument is summarised in Table 2.

Argumentation Based on Reality: the PRAGMATIC SCHEMA

The pragmatic argument aims at evaluating an act or an event in terms of its positive or negative consequences (NR § 62). The consequences can be observed, foreseen or even purely hypothetical.

The pragmatic argument belongs to the class of arguments related to the structure of reality, so its characterisation would heavily depend on the particular concept of reality that the arguing agents have.

An example of pragmatic schema could be: *since eating apples will help you slimming, and you would like to slim, then you should eat apples.*

N	RECIPROCITY
C	$E_{concept,perspective} = val$
O_c	$\exists\ concept_2 \in O, k, j$:
	$\mathcal{S}(concept.a_k) = \mathcal{S}(concept_2.a_j)$,
	$\mathcal{V}(concept.a_k) = \mathcal{V}(concept_2.a_j)$
A_c	$\beta(E_{concept_2,perspective} = val)$
R_c	$\Phi(\mathcal{S}(concept.a_k) = \mathcal{S}(concept_2.a_j))$;
	$\Phi(\mathcal{V}(concept.a_k) = \mathcal{V}(concept_2.a_j))$;
	$\Phi(E_{concept_2,perspective} = val))$
S_c	*Most* j, k :
	$\mathcal{S}(concept.a_j) = \mathcal{S}(concept_2.a_k) \rightarrow \mathcal{V}(concept.a_j) = \mathcal{V}(concept_2.a_k)$

Table 2: Reciprocity Schema

Claim. A typical claim of a pragmatic argument is to pass value to an action, from a given perspective:

$$E_{a,p} = val$$

In the example:

$$E_{Apples,p} = Good$$

Ontological constraints. Pragmatic argumentation can only be applied to an action, and the action must have a recognised consequence. We once again do not want to enter into the details of how this is represented, or what type of ontological concept a "consequence" is (it might be another action, a state of affairs, an event etc.), but we presume the ontology can represent the notion of "leading to" a consequence as a relationship between two concepts:

$$ACTION(a); \exists\, x \in C : LEADSTO(a, x)$$

In the example we may instantiate as:

$$O_c : ACTION(Apples); LEADSTO(Apples, Slimming)$$

Where Apples stands for eating apples here.

Acceptability constraints. As Perelman and Olbrechts-Tyteca point out, pragmatic arguments can only be successful if there is agreement on the value of the consequences, so for the argument to be accepted the audience should evaluate positively (or negatively if arguing against) the consequence of the action from one perspective:

$$\beta(E_{x,p} = val)$$

In the example, the audience may perhaps think that slimming is good from the social life perspective:

$$A_c : \beta(E_{Slimming,Social} = Good)$$

Note that we do not assume that the audience is already aware of the fact that Slimming is a consequence of eating Apples: this can in fact be learnt for the first time with the argumentative act, as shown in the relevance constraints below. On the contrary, if the audience does not value positively the consequence of slimming beforehand, the arguer cannot be confident in the success of the schema.

Relevance constraints. It should be evident to the audience that the action leads to that positive (negative) consequence, so the consequence relation, and the evaluation of the consequences should be in focus:

$$\Phi(LEADSTO(a, x)); \Phi(E_{x,p} = val)$$

In the example:

$$R_c : \Phi(LEADSTO(Apple, Slimming)); \Phi(E_{Slimming,Social} = Good)$$

If this holds, the argument can be put forward in a short form *you should eat apples as you would like to slim*. This, as in the previous example, can also be an instance of misleading argument (if apples in fact do not help slimming).

Sufficiency constraints. Perelman and Olbrechts-Tyteca claim that in order for the transfer of value to take place, an effort should be made to show that the action is necessary and sufficient for the consequence to happen. The satisfaction, and indeed the characterisation, of this constraint would therefore depend on whether the knowledge representation can express temporal modalities, like eventualities or possibilities. We will not, however, ascribe a temporal dimension to the knowledge base, again for the sake of simplicity, but it should be clear that such extension can be naturally obtained.

Another problem in pragmatic argumentation may be the imputation of the consequence to a too distant cause (NR §62). A closeness notion could be included into the semantics of the *LEADSTO* predicate to account for this, but again we do not consider this here.

N	PRAGMATIC
C	$E_{act,perspective} = val$
O_c	$ACTION(act); \exists\, x \in C : LEADSTO(act, x)$
A_c	$\beta(E_{x,perspective} = val)$
R_c	$\Phi(LEADSTO(act, x)); \Phi(E_{x,perspective} = val)$
S_c	$\nexists\, y : LEADSTO(act, y), E_{y,perspective} = val^{-1}$

Table 3: Pragmatic Schema

A third cause of weakness for the pragmatic argument might be in the action having other unfortunate consequences, that is negative consequences if arguing in favour of the action, or positive ones if arguing against. As it would be unreasonable to define this property in absolute terms, we might limit ourselves to make sure the action has no unfortunate consequences only from the perspective chosen for the argument:

$$\nexists\, y : LEADSTO(a, y), E_{y,p} = E_{x,p}^{-1}$$

In the example:

$$\nexists\, y : LEADSTO(Apple, y), E_{y,Social} = Bad$$

The schema for the pragmatic argument is summarised in Table 3.

Argumentation Establishing Reality: the MODEL SCHEMA

The model schema is one of the arguments establishing the structure of reality, that is those that, rather than abiding to more or less agreed knowledge, aim to establish new rules, by resorting to particular cases or by analogy. In particular, arguing by model aims at encouraging imitation of a behaviour (NR § 80). The idea is that if a model can be identified, adherence to the model's behaviour guarantees the value of this behaviour. Similarly, if an "anti-model" is identified, one can discourage imitation of a behaviour.

An example of argument by model can be: *it is good practice to use olive oil for cooking: all TV chefs do so*.

Once again, let us go through all the slots of our schema.

Claim. As for the pragmatic schema, the arguer is interested in creating an evaluation for a given action:

$$E_{a,p} = val$$

In the example:

$$E_{OliveOil,p} = Good$$

where OliveOil is used for "using olive oil for cooking" (again, we are not interested in the way this is going to be represented in the knowledge base).

Ontological constraints. In this argument, as in the previous example, we cannot ignore the semantics of the knowledge representation, therefore a precise definition of the constraints can only be made once the application is clear. In any case, the ontological constraints will aim at being sure that the concept to be evaluated is in fact an action, that we can identify a person, or a category of person, or an abstract "perfect being" that could have that action as a habit. We can summarise them as follows:

$$ACTION(a); \exists\, person : HABIT(act, person)$$

in the example:

$$O_c : ACTION(OliveOil); HABIT(OliveOil, TvChefs)$$

Acceptability constraints. The audience should share the view that the individual chosen for imitation is in fact a model, from some perspective:

$$\beta(MODEL(person, p))$$

in the example we might assume that the TV chefs can be models from the professionalism point of view:

$$\beta(MODEL(TvChefs, Professionalism))$$

As in the previous schema, we do not assume that the audience is already aware that the chosen model does have that particular habit.

Relevance constraints. It should be evident to the audience that the individual chosen for imitation is in fact a model:

$$\Phi(MODEL(person, p)$$

in the example:

$$\Phi(MODEL(TvChefs, Professionalism)$$

The relaxation of this constraint may lead to the false impression that somebody is a model and

N	MODEL
C	$E_{act,perspective} = val$
O_c	$ACTION(act); \exists\, person : HABIT(act, person)$
A_c	$\beta(MODEL(person, perspective))$
R_c	$\Phi(MODEL(person, perspective))$
S_c	$\nexists\, act_2 : HABIT(act_2, person), E_{act_2,perspective} = val^{-1}$

Table 4: Model Schema

has to be imitated (what, in fact, most TV adverts are inclined to do).

Sufficiency constraints. As Perelman and Olbrechts-Tyteca point out, an argument by model is weakened if the model "exhibits undesirable characteristics" (NR § 81). Once again we cannot assume in general the model is a perfect being, but we can reasonably ask the arguer to make sure that other "relevant" habits of the model, where relevant means here that can be evaluated from the perspective chosen, are not counterproductive:

$$\nexists\, act_2 : HABIT(act_2, person), E_{act_2,p} = val^{-1}$$

in the example:

$$\nexists\, act_2 : HABIT(act_2, TvChefs), E_{act_2,Professionalism} = Bad$$

The schema for the argument by model is summarised in Table 4.

2.4.4 Schemata and Their Use in Generating Rhetorical Arguments

Our main objective in the study of rhetorical argumentation is to draw some insights leading to the generation of arguments, rather than their analysis. Our work, therefore, has not as its main focus typical pragmatics issues (Lascarides & Asher 1993), although we hope our exercise might be useful in this respect as well. We envisage our definition of the rhetorical argument schema, perhaps along the tradition of informal logic and critical thinking teaching (Groarke, Tindale, & Fisher 1997), as the formalisation of what the speaker has to take into account when producing a "good" argument. So one might see our collection of constraints also as a collection of tokens that need to be added or otherwise to the presentation of the argument in order for it to be successful. For instance, in the reciprocity example shown in the previous Section, one might associate with each constraint a piece of discourse to be put forward, that can be eliminated if the constraints can be assumed to hold by the speaker. A complete exposition of the argument may therefore go along the line of:

(i) *I think you should like (the taste of) apples (Claim)*

(ii) *as I think they have the same texture of cookies (O_c)*

(iii) *and I believe you like cookies (A_c)*

(iv) *and I think apples and cookies are comparable because... (S_c)*

In the exposition, (ii) and (iii) can be eliminated if the speaker can assume that they are not only believed by the hearer, but also in the focus of attention (R_c). Seeing this from another point of view, we can say that the speaker may eliminate, or partially eliminate (ii) and (iii) if it can be assumed that the hearer is able to make the relevant associations with the previous discourse utterances by means of suitable rhetorical relations, so that the current utterance can be understood (Lascarides & Asher 1999). For instance, if the interlocutor has just mentioned how tasty crunchy cookies are, the speaker could simply say *Then you should like apples, they are crunchy too!*, with the second part of the sentence that can be further dropped if the crunchiness of apples can be assumed as a belief the interlocutor adheres to (and has in focus). Finally, the speaker can choose whether to add (iv) to the argument from the start, or wait until challenged by the hearer.

2.5 Discussion

We have analysed the notion of rhetorical argument, both from the philosophical and the computational perspective. We have been mainly interested in the rhetorical structure of arguments, and as a consequence, in the structure of rhetorical discourse. From this point of view, only a few researchers in AI have been undertaking works that are relevant to the one presented here. They can be divided into those mainly concentrating on the representation of multiple viewpoints, perhaps ranked in order of their importance to the audience, such as Clark & Wilkes-Gibbs (1986), Fox & Parsons (1997), Karacapilidis (1996), Lowe (1985), Zukerman, Korb, & McConachy (1996), and those focusing more on the structure of argumentative discourse, such as Elhadad (1995), Hovy (1990), Maybury (1993), Reed (1999).

The work presented here is perhaps more akin to the latter, although research in discourse analysis usually appeals to theories from linguistics rather than philosophy, most often the already mentioned Rhetorical Structure Theory (Mann & Thompson 1988). It should be evident that our formalisation can be comfortably used in a system for generating argumentative discourse based on planning, such as the ones by Hovy (1990) or Maybury (1993). While not suggesting an approach to discourse generation different from RST (which, despite criticisms (Moore & Pollack 1992) is still most widely used), this work sets itself apart from the issue "RST or not RST", in order to advocate for a more principled collection of the knowledge necessary to make any discourse strategy, hence also RST, work better. For example, RST has a perspective on discourse based on relations among text spans, which allows one to build a tree to represent the whole discourse. This concept is very powerful, and easily formalisable in computational terms (Marcu 2000b), as a tree is typically the output of a hierarchical planning algorithm, as we have mentioned before. The semantics of the relations, however, is not very formal in RST: relations are defined in terms of constraints, or assumptions, that can be made on the goals of the speaker, which are probably more useful to the analysis of a piece of text than to its generation. The same, long debate on deciding how many relations exist, and whether they can be classified, is seen from the analysis point of view, and generally approached on the basis of surface characteristics of the text, such as cue phrases (Knott & Sanders 1998). Using RST, or any other discourse structure theory, for the generation of text, requires precisely a better definition of the grounds on which the text should be generated, and therefore the objectives of the speakers and the intended effect on the hearer need to be better formalised. Moreover, researchers in discourse processing typically concentrate on discourse in general, which happen to be argumentative in some cases. For such cases standard argumentation schemata might be embedded in the system (Dalianis & Johannesson 1998; Reed 1999), but there are very few discourse processing systems that base themselves on one of the frameworks among the plethora that philosophical theories on rhetoric provide (an already mentioned notable exception is the work by Elhadad (1995)). Our work constitutes an attempt in that direction, and is in line with a series of recent efforts to put together two communities, the AI/NLP and philosophy of argument ones, that we believe have much to gain from cross-fertilisation (Reed & Norman 2000; Carenini, Grasso, & Reed 2002; Reed, Grasso, & Carenini 2003).

2.6 Conclusions

In his position paper presented at the Symposium on Argument and Computation, Crosswhite (2000) pictures three possible avenues for research should an AI scholar wishes to undertake the task of creating a computational model of rhetorical argumentation. The first is the exploitation of the argumentative schemata, of which literature in rhetoric provides a rich repository. The second is the exploitation of the figures of speech, and the ways they influence argumentation. The third is the explicit representation of the audience. In this Chapter we have showed how we have addressed the first issue. In Chapter 4 and 6 we will address, from different perspectives, the third one. As to the second issue, we believe this relates, speaking in natural language processing terms, to the surface representation of the argument, that is its verbalisation in a chosen, human language. We leave this issue outside the scope of our work, although it definitely constitutes its natural extension.

Chapter 3

Argumentation as a Dialogic Activity

"Well, goodness me! I can't see anythin' ter be glad about gettin' a pair of crutches when you wanted a doll!"
Pollyanna clapped her hands. "There is–there is," she crowed. [...]
"Well, then, suppose YOU tell ME," almost snapped Nancy.
"Goosey! Why, just be glad because you don't–NEED–'EM!"

- Eleanor H. Porter, *Pollyanna*

In "Fundamentals of Argumentation Theory" (van Eemeren *et al.* 1996), an important, recent handbook, which in a way testifies to the current period of good fortune which Argumentation Theory is enjoying, argumentation is defined as:

> *a verbal and social activity of reason aimed at increasing (or decreasing) the acceptability of a controversial standpoint for the listener, by putting forward a constellation of proposition intended to justify (or refute) the standpoint before a rational judge.*

The definition emphasises, among others, the social aspect of the activity of arguing: apart from few exceptions, argumentation is a process involving at least two rational agents[1]. This is especially true for rhetorical argumentation, whose communicative character has been remarked several times (see Crosswhite (2000), for example). A complete model of rhetorical argumentation cannot therefore ignore such dialogic dimension. The present Chapter will examine this aspect of the argumentation activity, providing an overview of different kinds of approach to the issue, and concluding by describing our own.

[1] "An agent is said to be rational if it chooses to perform actions that are in its own best interests, given the beliefs it has about the world" (Wooldridge 2000, p. 1).

3.1 Models of Argumentative Dialogue

In a brief survey on theoretical models of argumentative dialogue, we start by introducing those approaches that more directly addressed the problem of the structure of argumentative discourse, therefore more closely related with computational linguistics. We then include approaches concerned with the *descriptive* study of dialogue, and conclude with those concerned with the *formal* one, in a distinction outlined by Hamblin (1970).

3.1.1 Discourse Analysis and Dialogue Games

The term "dialogue game"[2] was used in the 70s by Levin, Mann and Moore (LMM) (Levin & Moore 1977; Mann, Moore, & Levin 1977; Mann 1988), as an extension to Wittgenstein's (1953) "language games", to describe structures capturing knowledge about how language is used in order to achieve some given goals. LMM's dialogue games seek to formalise regularities in natural language interaction, with an approach very similar to the one that, a few years later, the same Mann, together with different colleagues, used to elaborate the Rhetorical Structure Theory (see Sect. 2.3.1, page 18). The purpose of LMM's study of dialogue games is in fact similar to the one of RST: giving a set of guidelines to establish whether a random sequence of speech acts performed by two dialoguing agents is *coherent*. The proposed dialogue game theory aims to represent an extension to speech act theory (Searle 1969), seeking to handle a group of sentences, called "Indirect speech acts" (Searle 1975), that were "not handled satisfactorily by any of the current theories of the direct comprehension of language" (Levin & Moore 1977).

The dialogue game structure proposed is a triple (Mann 1988):

$$\langle IP, G_R, CC \rangle$$

indicating the Illocutionary Point IP of the game, that is the goal of the player who have started the game, the Goals of the Responder G_R, and a set of Conventional Conditions CC that have to hold for the game to exist, or in other words the set of presumptions that the two players can assume are valid (see Table 5). Games are not necessarily competitive, consciously pursued, or zero-sum. A dialogue is said *motivationally coherent* if each bid in the dialogue "serves" to the goals of the game. The theory identifies several types of game, whose description is given in terms of the triple $\langle IP, G_R, CC \rangle$, though the list is not assumed to be exhaustive. Examples from Mann (1988) are in Table 6.

Similar in spirit, although conducted independently, the work of Lauri Carlson (Carlson 1983; 1984)

[2] A fairly comprehensive survey on dialogue games is the one by McBurney (2002).

Parameter	Description
roles	initiator (I), responder (R)
illocutionary point	goal of the initiator in starting the game
goals-of-R	non-empty set of goals of the responder during the game
conventional conditions	set of state descriptions, including at least the following: 1. I is pursuing the illocutionary point as goal. 2. I believes that achieving the illocutionary point is feasible. 3. I believes that R's achieving the goals-of-R are feasible. 4. I has the right to achieve the illocutionary point. 5. I has the right to use the dialogue game. 6. The illocutionary point has not already been achieved. 7. R is willing to pursue the goals-of-R.

Table 5: Parameters of Mann's Dialogue Game.

seeks to describe a theoretical construct, called again "dialogue game", able to characterise well formed discourse. Carlson's games are "cooperative activities of information exchange" (Carlson 1983, p.XVIII), and are restricted to question-answer games: the players, through a series of declarative and interrogative sentences, try to achieve a common understanding on an answer to some question. Carlson's work is admittedly inspired by the work of Jaakko Hintikka (1984), who outlined a logical model of question-answer dialogues. Both models deal with informative aims, and both of them restrict the players to a set of possible moves, such as *deductive, interrogative, assertoric, definitory*, and so on. In Carlson's games, there is also an explicit representation of the *aims* and *attitudes* of the players, which are included in the description of the admissible moves. Hintikka's games are collaborative to the extreme: each player has the goal of reaching the closure of his set of belief, so, even when one of the two players reaches this goal, the game goes on until this becomes true for the other player as well. Carlson's game can also be competitive debates: players increase their payoff whenever they succeed in replacing one of the opponent's assertions with one of theirs. Carlson's model is, however, much richer, as it comprises also the notions of propositional attitudes, presuppositions and style, and in this sense it is much closer to works in discourse analysis than to those related to dialogic logic.

3.1.2 Philosophical Approaches

The approach to the philosophy of dialogue which is perhaps the best known to researcher in artificial intelligence is the one of Grice (1975). On the basis of a very strong, and perhaps not completely realistic assumption of total cooperativity among the dialogue participants, Grice proposes a set of four *maxims*

Game	Illocutionary Point	Goals-of-R	Conventional Conditions
information seeking	I knows ϕ	I knows ϕ	R knows ϕ
information offering	R knows ϕ	R knows ϕ	I knows ϕ ; R's knowledge and ϕ are consistent
information probing	I knows whether R knows ϕ	R informs I of R's knowledge of ϕ	I knows ϕ
helping	I is able to perform ψ	I is able to perform ψ	R is able to cause I to be able to perform ψ; I has the right to perform ψ
dispute	R believes ϕ	R justifies that I might not believe ϕ	I believes ϕ; R does not believe ϕ
permission seeking	I knows that R approves that I performs ψ	R chooses whether to approve that I performs ψ, and I knows this choice	I wants to perform ψ; I does not have the right to perform ψ without R's permisson
action seeking	R causes ψ to be performed	R causes ψ to be performed	R might not cause ψ to be performed in the normal cause of events

Table 6: Examples of Mann's Dialogue Game.

that should dictate a well conducted dialogue:

1. Maxim of Quantity: make your contribution as informative as required, and do not make your contribution more informative than is required.

2. Maxim of Quality: make your contribution one that is true

3. Maxim of Relation: be relevant

4. Maxim of Manner: be clear

A variety of work in Natural Language Processing has been grounded on these maxims, especially related to the explanatory ability of artificial systems (Cawsey 1992; Maybury 1992; de Rosis *et al.* 1995), or to the interpretation of utterances in a dialogue (Walker & Moore 1997; Walker 1992), although the maxims themselves are far from being precisely formalisable.

From this point of view, more insight can be gained from an important argumentation theory, called *Pragmadialectics* and developed in Amsterdam (van Eemeren & Grootensdorst 1984), aimed at combining linguistic studies of the language, whereby discourse is seen as a collection of speech acts (pragmatics), with the study of critical dialogue, whereby such speech acts are in fact a structured way to solve disputes (dialectics). By combining the normative and the descriptive dimensions of argumentative discourse, critical discussion is defined by pragmadialecticians by means of a set of rules, a sort of "code of conduct" two rational agents must adhere to. The rules, shown in Table 7, form a "heuristic, analytical and critical framework [...] for dealing with argumentative discourse" (van Eemeren *et al.* 1996, Ch.10), which consists of speech acts belonging to five categories:

	PRAGMADIALECTIC RULES	EXAMPLES OF VIOLATIONS
1	Parties must not prevent each other from advancing standpoints or from casting doubt on standpoints.	Banning standpoints, pressurising, personal attacks, threatening...
2	A party that advances a standpoint is obliged to defend it if asked by the other party to do so.	Presenting standpoints as self evident, personal guarantees, ...
3	A party's attack on a standpoint must relate to the standpoint that has indeed been advanced by the other party.	Distorting standpoints, taking utterances out of context, ...
4	A party may defend a standpoint only by advancing argumentation relating to that standpoint.	Using irrelevant points, playing with emotions...
5	A party may not disown a premise that has been left implicit by that party or falsely present something as a premise that has been left unexpressed by the other party.	Denying or magnifying an unexpressed premise.
6	A party may not falsely present a premise as an accepted starting point nor deny a premise representing an accepted starting point.	Presenting premises as self-evident, circular arguments,...
7	A party may not regard a standpoint as conclusively defended if the defense does not take place by means of an appropriate argumentation scheme that is correctly applied.	Fallacious argumentation, as populistic fallacy, slippery slope, false analogy, hasty generalisation...
8	A party may only use arguments in its argumentation that are logically valid or capable of being validated by making explicit one or more unexpressed premises.	Affirming the consequent, denying the antecedent...
9	A failed defense of a standpoint must result in the party that put forward the standpoint retracting it and a conclusive defense of the standpoint must result in the other party retracting its doubt about the standpoint.	Concluding a standpoint is true/false merely because it has/has not sucesfully defended.
10	A party must not use formulations that are insufficiently clear or confusingly ambiguous and a party must interpret the other party's formulations as carefully and accurately as possible.	Using implicitness, unfamiliarity, vagueness, ambiguity, uncleareness, indefiniteness...

Table 7: Pragmadialectic rules for critical discussion

1. *assertives*, used for expressing a standpoint, supposing, denying, conceding, and so on.
2. *directives*, used for making requests, questions, challenging to defend standpoints, and so on.
3. *commissives*, used to accept or otherwise a standpoint, the challenge to defend a standpoint, a rule of discussion, the start of a new discussion, and so on.
4. *expressives*, used to manifest an agent's feelings, although they "play no direct part in a critical discussion" (van Eemeren *et al.* 1996, Ch.10).
5. *declaratives*, used to make definitions, to exemplify, to explain, to clarify interpretation, and so on.

Argumentative discourse is analysed by pragmadialecticians with the aim of identifying the elements that are crucial in a dispute, and "reconstructed", by excluding irrelevant speech acts, in order to produce an idealised dialogue that shows how the dispute was solved. The reconstruction is done by re-elaborating the argumentative discourse in four "stages": (1) deletion of irrelevant elements, (2) addition of relevant elements which have been left implicit, like unexpressed premises, (3) permutation of the discourse

Type of Dialogue	Initial Situation	Participant's goal	Goal of Dialogue
Persuasion	Conflict of opinions	Persuade other party	Resolve or clarify issue
Inquiry	Need to have proof	Find and verify evidence	Prove/Disprove hypothesis
Negotiation	Conflicts of Interests	Get what you most want	Reasonable settlement that both can live with
Information-seeking	Need information	Acquire/Give information	Exchange information
Deliberation	Dilemma or Practical choice	Co-ordinate goals and actions	Decide best available course of action
Eristic	Personal conflicts	Verbally hit out at opponent	Reveal deeper basis of conflict

Table 8: Walton and Krabbe dialogue types

elements, rearranging them so that the dispute is made clearer, and (4) substitution of ambiguous or vague standpoints with standard constructions. The rules as they stand presume complete fairness in the two party dialogue: the very simple fact that the arguing agent must be able to prove a point prior to putting it forward excludes the use of evaluative arguments, as they cannot be "proved" as such, as we observed in Ch. 2. The whole theory is in fact based on the concept of logical (deductive?) validity, although in a recent paper van Eemeren & Houtlosser (2000) add a rhetorical dimension to the pragmadialectical dialogue, by arguing that the speakers do make rhetorical choices whenever (1) they select the topic they feel most comfortable with, (2) they choose the perspective most agreeable to the audience and (3) they present their move in the most effective wording.

Another influential work on dialogical argumentation is due to Walton and Krabbe, whose notion of dialogue is that of a "conventionalized, purposive joint activity between two [...] speech partners" (Walton 1998, p. 29). A dialogue is then a normative framework, where the two parties exchange arguments in a turn-taking fashion in order to achieve a collective goal. Such collective goal determines the type of dialogue the parties are conducting: resolve a dispute, prove a hypothesis, exchange information, and so on. In addition to the collective goal, each participant has individual goals, and the nature of their collective and individual goals determine the participant's obligations in the dialogue. Shifts are permitted from one dialogue type to another, although illicit shifts are associated with fallacies, where a shift might be done deceptively, for instance presenting irrelevant arguments in order to confuse the opponent. In the theory, it is interesting how the evaluation of an argument can only happen in the context of the dialogue situation, and therefore looking at the goals of the dialogue, the commitments of the participants, the burdens of proof, and so on. The characteristics of the dialogue types, as defined in Walton & Krabbe (1995) (see Table 8) have recently become very popular in the multi-agent systems community, as briefly outlined in Sect. 3.2.

3.1.3 Formal Approaches

Formal approaches to the study of argumentative dialogue are mainly based on the models of dialectic argument due to philosophers Charles Hamblin (Hamblin 1970; 1971) and Jim MacKenzie (Mackenzie 1979; 1990). These models exploit once again the game metaphor to help understanding the mechanisms of the dialogue: the two partners are *players*, each of them having one or more *goals*, which define their role in the dialogue, there is a fixed set of *moves* that are allowed at each stage of the game, and a *payoff* can be defined, if a clear winning situation can be identified. As opposed to the models in the previous Section, however, these models are less focused on naturally occuring dialogues, and more on creating a logic framework.

Formal Dialectics

Hamblin first introduced the term *formal dialectic*, defined as "the setting up of simple systems of precise, but not necessarily realistic rules, and the plotting of the properties of dialogues that might be played out in accordance with them" (Hamblin 1970, p.256). Despite the terminology, however, Hamblin was not interested in the logical properties of his systems, locating himself, in a sense, as a bridge between discourse based approaches and strict *logic* approaches. Inspired by Hamblin's work, a plethora of frameworks have been proposed for formal dialectic. A remarkable example is the work by Kamlah & Lorenzen (1984), who introduces a logic where the usual logical connectives are defined dialogically, rather than using truth values. The logic, therefore, states in which way players can attack/defend each other's statements by means of dialogue game rules, which also define a winning condition. Grounded in this work are, *inter alia*, those by Barth & Krabbe (1982), or van Eemeren & Grootensdorst (1992), who identified norms that must hold in critical discussion to resolve a difference in opinion.

A different aim is the one of Loui (1993), who investigates whether formal dialectics can constitute a basis for a form of non-monotonic reasoning which takes into account computational limits. Interest is focused on *disputations*, where parties can cooperate on some aspects (e.g. to reach a reasoned opinion, without unfair playing) and can be non-cooperative on other ones (e.g. what such opinion should be). The model is based on the concept of *locution*, where a *move* is a sequence of consecutive locutions put forward by a player. The notion of locution also embeds the explicit expression of the *resources* consumed by the player in putting forward the locution. This way, protocols may be devised which in some way impose constraints on such resources, for example by timing responses, limiting the number of locutions per turn, limiting the number of previous locutions that can be used, and so on. Interestingly enough, Loui defines a *dialectical* protocol as a protocol in which no limitation is given to the number of previous locutions to use.

More recently, Vreeswijk (1997) proposed an *abstract argumentation system*, defined as a triple $\langle \mathcal{L}, R, \leq \rangle$, composed by a language, $\mathcal{L}$, a set of rules of inference, R, and an ordering relation, $\leq$, which is reflexive and transitive. The language has no particular constraints but the fact that it must contain a distinguished element ($\perp$). Elements of the language are called *sentences*. The rules determine which inferences are possible, and can be *strict* or *defeasible*. A chain of rules forms an argument. Each argument is characterized by: *premises*, a *conclusion*, *sentences* (or proposition), *assumptions*, *subarguments*, *top arguments*, a *length* and a *size*. The ordering relation can be constructed according to diverse parameters, such as the number of defeasible steps, the strength of the weakest link, the specificity of the argument etc. Crucial notions is that of *defeat*, which is defined in terms of *undermining* a concept and *incompatibility*. Three operators are defined for this purpose: (1) enablement (a form of defeat that takes into account the effects of defeaters on (set of) sub-arguments), (2) inductive warrant (involving iterated defeat) and (3) warrant (that is a relation between competing arguments). Strong points of the system are its capability of dealing with the length of arguments as distinguished from their force, on the hypothesis that it is not always true that the longer the argument is, the stronger it is; and its capability of handling situations in which it is not clear which argument should win. The system has been recently extended by Baroni, Giacomin, & Guida (2000), who also integrate it with a model for belief revision and defeasible reasoning.

Formal Dialogue Games

Mackenzie (1990) defines a dialogue as a sequence of *locution acts*, in each of which a *participant* utters a sentence or *locution*. A *System of dialogue* is defined by:

1. **Language:** set of *statements* (or declarative sentences) together with locutions produced from statements. For instance: Assertions ("P"), Questions ("Is it the case that P?"), Withdrawals ("I'm not sure that P"), Resolution demands, ("Resolve whether P"), Challenges ("How is it known that P?").

2. Syntactically recognizable **properties** of (or relation between) locutions. For instance: an answer to a question of a statement can only be the statement itself, its withdrawal or its denial.

3. A function which specifies the **commitment store** of each participant at each stage of the dialogue, together with rules to change this store, for instance: uttering a statement adds that statement both to the speaker's and to the hearer's commitment store.

4. **Rules of dialogues** which enable to decide, given a legal dialogue on length n whether the addition of a certain event at n results in a legal dialogue of length $n+1$. For instance: after A asks a question, B must say something which is an answer to that question.

Mackenzie formalized 4 dialogue systems, which differ according to the previous mentioned characteristics, including a system which provides a framework in which relations such as *misleading* or *puzzling* can be formalized. Even in these cases, however, there is no appeal to notions such as *intentions* or *conversational implicature* (“psychologism should have no part in logic”).

3.2 Computational Approaches

Theoretical work on argumentative dialogue has found a good number of applications in computation, mainly distributed among three main streams of research: multi-agent systems, computer supported collaborative learning/work and legal informatics. This Section surveys the most significant works in the respective streams.

3.2.1 Negotiation among Autonomous Agents

Perhaps the most natural environment for the implementation of dialogic argumentation systems comes from the field of Multi-Agent Systems (MAS). In the model proposed by Wooldridge & Jennings (1995), a MAS is a society of “agents” which can autonomously interact with other agents (or with humans), by taking proactive decisions, based on their set of goals and beliefs, and cannot rely upon their interlocutor’s knowledge, nor, in extreme cases, even their *bona fi de*. From this conceptualisation, to the need for the agents to negotiate, and argue, in order to achieve their goals, the step is short indeed.

Jennings, Parsons and Sierra (Parsons & Jennings 1996; Parsons 1996; Parsons, Sierra, & Jennings 1998) studied precisely how negotiation by argumentation can occur among autonomous agents. In their model, agents can elaborate logical arguments for and against propositions of interest, and, to better mirror human argumentation, arguments can be weighted by examining the support of steps. The reasoning mechanism is extended to allow arguments to indicate both support and doubt in propositions. The system is grounded in the *Beliefs, Desires and Intentions* (BDI) agent model (Rao & Georgeff 1991)[3]. Agents choose among several arguments, to put forward in favour or against a given proposition, on the basis of their weight and their structure, expressing characteristics like defeasability, non-triviality and tautology. A negotiation protocol is defined (Parsons, Sierra, & Jennings 1998), which has the role of the set of “rules of dialogue” of formal dialogue game systems. The protocols define four main dialogue directives: proposal, acceptance of a proposal, critique of a proposal and withdrawal of a proposal. Proposals are generated by constructing arguments (proofs) for an agent’s intentions, and the critiques of the proposals are in fact counterarguments attacking or undercutting the proposal’s proof.

[3]See Ch. 4 for more detail on the BDI model.

Reed (1998a) further explores the notion of a negotiating agent by proposing a framework encompassing the variety of dialogue types in the typology of Walton & Krabbe (1995) (Table 8), some of which are not always recognised in MAS research. The notions of "belief", "contract" and "plan" are made explicit in the agent communication for each dialogue type, by proposing the definition of *dialogue frame* as a tuple:

$$F = \langle\langle t, \Delta\rangle \in D, \tau \in \Delta, \{u^0_{x_0 \to y_0}, \ldots, u^n_{x_n \to y_n}\}\rangle$$

where t is the type of dialogue frame and τ is the topic of the dialogue frame (the remaining parts of the definition refer to the various utterances between interlocutors x and y). Therefore, a dialogue frame, in addition to being of a particular type, has to focus on a particular topic, that will be different from type to type (a persuasion frame will focus on a belief, a negotiation on a contract, a deliberation on a plan, etc.). The same points of departures are used by Parsons, Wooldridge, & Amgoud (2002), where protocols are defined for information-seeking, inquiry and persuasion dialogues.

Another, more recent model which is inspired by Walton & Krabbe's work is the *Agent Dialogue Framework* by McBurney (2002). The framework is defined as a tuple:

$$(\mathcal{A}, \mathcal{L}, \Pi_{Atom}, \Pi_{Control}, \Pi)$$

where $\mathcal{A}$ is the community of agents, $\mathcal{L}$ is the language in use, Π_{Atom} is the set of atomic dialogue types, again in the Walton & Krabbe (1995) philosophy, and $\Pi_{Control}$ are the "control dialogues", that is the dialogues "about dialogues", which intend to formalise the two participants' agreement on starting/ending a dialogue, changing dialogue type, and so on. Dialogues can be combined according to formally defined operators, like Iteration, Sequence, Embedding, where Π in the tuple represents the closure of the set $\Pi_{Atom} \cup \Pi_{Control}$ under these operators.

Finally, Kraus, Sycara, & Evenchik (1998) define a logical model of argumentation among agents based on a representation of the agents' belief, desires and intentions/goals. The model is sensible to a characterisation of agents based on these properties, so one can define, for example, omniscient agents, cooperative agents, knowledgeable agents, and so on, each defined by a set of axioms on the modal operators representing their mental state. Axioms are also defined for the argumentation process, distinguishing among various argument "types", which, unlike those of Walton & Krabbe (1995), are in fact "strategies" to win the dispute: threats, promise of future reward, appeal to prevailing practice, and so on.

3.2.2 Argumentation Based Collaborative Activities

A second, main area of application of argumentative dialogues aims at supporting situations where a collaborative activity takes place between two parties, one of which might be "virtual". A large part of these applications involve educational environments: supporting collaborative learning is seen as a very promising area to test principles and systems. A good number of other applications involve supporting collaborative decision making.

Pilkington *et al.* (1992) describe a tutoring system that teaches students how to debate. It engages into a conversation upon controversial issues such as abortion or capital punishment, by seeking the view of the student on the subject and adopting the diametrically opposite position. The system has the characteristics of a Dialogue Game *à la* Carlson (1983): there are constraints about the roles of the participants, such as what the topic of a phrase must relate to, the participant must behave *"fairly and logically"*, and the game consists of the ordered moves of the participants. The game has the structure proposed by Mackenzie (1990). There are five fixed move types:

1. statements,
2. withdrawals,
3. questions,
4. challenges and
5. resolution demands.

Commitment is regulated by fixed rules, while dialogue rules regulate the two agent turns, and rules of strategy and of focus are used by the system to decide what to do next. The interaction has the form of a mixed-initiative dialogue, in which either party can challenge or rebut the arguments of the other. There is no fixed end point of the interaction. The characterisation of dialogue moves is computationally very simple: one move corresponds to just one "locution", formed by a statement and an operator (e.g. assert, request, withdraw. . .).

Baker (1999) proposes a framework for *constructive argumentation*, that is argumentation where new knowledge or meanings are elaborated jointly by the two interlocutors, in a goal oriented cooperative activity. The author observes that research on argumentation in collaborative learning is often based on models of argumentation designed for the analysis of argumentative text, rather than interactions. Baker argues that a suitable model should instead account for different dimensions of the argumentative activity: dialectical, rhetorical, epistemological, conceptual and interactive. From a dialectical dimension, the one of interest in this Chapter[4], Baker's definition of an argumentative interaction sees the presence of three phases: an opening one, where the two participants agree that there is in fact a conflict with respect to differences of positions or attitudes. An argumentation one, where each participant must perform at least one communicative act in defence of their position, and/or attacking the partner's one. Finally, a closing one, where the conflict is settled, a phase that may be implicit, for example when the two parties tacitly move onto another topic of discussion. Argumentation is therefore once again compared to a dialectical game, a "verbal jousting tournament", that begins with a conflict, $\langle A, B, Con, T \rangle$ between participants A and B on a thesis T, where A and B agree on a set of preliminary concessions Con. The participants then must perform a series of moves, or communicative acts, in a turn taking fashion, where the moves are governed by rules that specify which are the admissible ones, what counts as an attack or defence, how to decide who wins the game.

Burton & Brna (1996) are interested in the characteristics of collaborative dialogue, as opposed to general dialogue, and in particular to the modelling of the dialogue, rather than its analysis. To simulate collaborative dialogue, attention is paid not only onto the "content" of the dialogue, taking care of participants goals and intentions in terms of the BDI paradigm, but also onto the "form" of the dialogue. To this aim, the authors draw upon the definitions of dialogue game of both Levin & Moore and Mackenzie, but integrate all the game types proposed by the two theoretical frameworks into a single, large dialogue game, where all moves are interconnected and players can "change game" easily. This would model situations where, for instance, doubts on the opponent's reply in the middle of an inquiry game in fact start an argumentation game, to discuss the statement. In order to keep track of the different games, a tree structure is maintained, where each branch represents a parallel game. The attention is only on the surface form of the game, so no effort is made to evaluate the relevance of a move in terms of its content: to deal with focus shift, the assumption is made that the most recent game is also the most relevant in each moment, and the game itself is designed so that moves that can cause confusion are not permitted (Burton, Brna, & Pilkington 2000).

Maudet & Moore (1999) prototyped a system to train students to exercise their critical skills by engaging in debates on controversial issue, like capital punishment. The framework proposes a dialogue game based on the DC system by Mackenzie (1979), with five move types:

1. statements,

[4] As to the other dimensions, the rhetorical one for Baker refers to the specific epistemic effects on the participants: their points of view, and their attitudes towards them (knowing, believing, suspecting, seeing as plausible, and so on). The epistemological one refers to the nature of the knowledge involved, both ontologically and specifically related to the domain. The conceptual dimension is concerned with the the way in which knowledge is represented, in the universe of reference. Finally, the interactive dimension refers to the processes by which knowledge or concepts are "transformed, reformulated, refined, elaborated or negotiated" in communicative interaction.

2. withdrawals,
3. questions,
4. challenges and
5. resolution demands.

Rules regulate commitment store amendments and general dialogue conduct. The computer plays one of the debate parties, and moves according to strategies that dictate whether to change focus or not, and whether to attack the other party's position rather than defend its own. Like Burton & Brna (1996), the authors opt for a general structure of dialogue game, rather than a specific one, that can accomodate different dialogue types, both symmetrical (like debate) and asymmetrical (like the classical student-tutor one).

Ravenscroft (2000) implemented a system for collaborative argumentation in virtual learning environments, based again on dialogue games. The object of learning is an underlying ontology of concepts, upon which the sytem can use common sense reasoning rules. Students are challenged by the system to explore these concepts, and a dialogue game takes place between the system as a virtual tutor and a tutee, where the system can use argumentative tactics, like challenge-construct, probe-question, persuade, acquire-construct, assume, summarise.

In decision support systems, Fox & Parsons (1998) argue against the classical decision theory view that privileges quantitative approaches, and present a qualitative theory based on argumentation, that is meant to combine the "naturalness" of dialectic argument and the soundness of deductive ones. A labelled deductive logic is used for reasoning in this framework. The application is a system to support general practitioners in drug prescriptions.

Karacapilidis & Papadias (2001) investigated avenues to support collaborative decision making by means of argumentative discourse among the decision makers. The argumentation framework resulting from the investigation, HERMES, is based on four argumentative elements: alternatives, positions, constraints and preference relations. A discussion forum can be created where "issues", corresponding to decisions to be made or goals to be achieved, can be discussed. For each issue, there can exist a set of alternatives, each of which can be supported or objected to by participants by means of positions. Constraints and preferences are used to evaluate issues and propositions. Argumentation happens by means of "discourse acts", that can open issues, add new alternatives, insert new positions in favour or against one alternative. The implemented tool allows the users to keep track of this chain of reasoning in a tree-like structure, where each node can be added by any of the participants. The dialogue is therefore asynchronous: there is no specific notion of "turn taking", as participants can add discourse acts at any time, while the appropriateness of one act at any time is controlled by the system.

3.2.3 Computational Dialectics

The third, important area of application or argumentative dialogues is "Computational Dialectics", a term allegedly due to T. Gordon and R.P. Loui to denote a relatively novel branch of artificial intelligence born in the middle nineties, as a cross fertilisation of philosophy, artificial intelligence and legal theory, aimed at formalising the notion of disputes[5].

In the field, a good amount of research is aimed at representing legal knowledge and the way facts and rules can be interconnected to form arguments, mostly based on Toulmin's model of argument structure, as in the works by Bench-Capon (1997), Clark (1990) or Marshall (1989) for example. Another good portion of scholars working in legal informatics is interested in the dialectical aspect of legal practice: it is recognised that, because of the very nature of the law, abundant with many defeasible principles, the problem of justifying legal reasoning is crucial (Hage 2000). Works in legal informatics are mainly inspired by the work of Barth & Krabbe and Lorenzen, with a specific stress on the defeasibility aspects of legal argumentation, and its sensibility to the process used to put forward arguments.

Several dialogue game structures have been proposed specifically tailored to this domain, after earlier works, like the one by Ashley (1991). For example, in DiaLaw (Lodder & Herczog 1995) a dialogue game is proposed with four dialogue moves (claim, question, accept, and withdraw) that can be used by two opponents to reach an agreement on a dispute. Another influential work is the Pleading Game (Gordon 1995), a battle of arguments where the two parties have to abide by rules such as "no party can contradict himself", or "supporting argument may only be put forward if the issue has been denied by the opponent". Other dialogical models of the law are due, *inter alia*, to Bench-Capon (1997), Freeman & Farley (1996), Hage (1997), Loui (1994), Prakken (1997), Sartor (1994), and, most interestingly, to Feteris (2000), who adopts a pragmadialectical view to the study of legal discussion, in the style of van Eemeren & Grootensdorst (1984) already seen above.

An important part of computational dialectics, although less strictly interested in the dynamics of dialogue, is dedicated to the study of *argumentation frameworks* (AFs), aimed at the definition of a formal logic for defeasible argumentation. The definition of an AF is due to Dung (1995), in terms of a couple:

$$AF = \langle AR, Attacks \rangle$$

[5] The discipline has been alimented by several workshops and symposia, among others at AAAI 1994, FAPR 1996, ECAI 2000.

where AR is a set of arguments, and *Attacks* is a relation defined in $AR \times AR$ so that for any two arguments $a_i, a_j \in AR$, it is true that *Attacks*(a_i, a_j) if the argument a_i attacks, or undermines the argument a_j. Notions like the "acceptability" of an argument, "conflict free" set of arguments, "admissible" set of arguments, and so on are also defined with the aim of establishing what can be safely deduced from the initial AR set, which is supposed to be not necessarily consistent. Extensions and elaborations on these concepts are due, among others, to Amgoud & Cayrol (1998), Kowalski & Toni (1996), Vreeswijk (1997).

3.3 A Dialogue Model for Rhetorical Argumentation

In this Section we present our proposal for a dialogue structure for rhetorical argumentation discussion. We have seen in the previous Sections that various authors, while defining the dialogue models or games in different ways, all agree on the main notions that have to be identified:

1. the two participants,

2. the types of admissible moves,

3. the sequence of moves,

4. the goal of the dialogue, as a whole.

We present our proposal for each of these notions in what follows.

3.3.1 The Participants

A rhetorical dialogue happens among two participants[6]. We will call them Speaker (S) and Hearer (H), following a classical notation used in dialogue modelling, but this by no means implies that the ability put utter sentences is unilateral. The two participants maintain a representation of their own mental state, of their own commitment store, of their interlocutor's commitment store, as well as a (subjective) representation of the mental state of their interlocutor. No hypothesis is made on the participants level of honesty and fairness in the dialogue conduct. We leave until Chapter 4 a more detailed analysis of the mental models of the two participants, and will concentrate in what follows on the characteristics of the dialogue only.

[6]We will limit ourselves to two-party dialogues, although the formalism can naturally be extended to larger communities.

3.3.2 The Moves

We start by giving our definition of a dialogue move:

Definition 4 (Dialogue Move) *We define a dialogue move as a 6-tuple:*

$$\langle i, S_i, H_i, T_i, R_i, C_i \rangle$$

where:

- i *is a unique identifier for the move in the dialogue. It will typically be a progressive number;*
- S_i *is the author of the move, the currently speaking agent;*
- H_i *is the agent the move is addressed to. While obvious in a two-party dialogue, the addressee must be specified in a community of more than two agents;*
- T_i *is one of a set of pre-defined moves Types, which will be detailed below;*
- R_i *is a (possibly empty) reference to a past move in the dialogue; to which the current move is in some way replying, or addressing to, and*
- C_i *is the content of the move, the actual utterance that* S_i *is putting forward.*

Fixing a pre-established set of move types is common in artificial dialogue systems, mainly with the aim to simplify the process of clarification of the intention of the move itself. While acknowledging this makes dialogues less natural, we too believe that the loss of generality is worth the effort, and we adopt the same principle here. As to how many and which types of moves should be needed, researchers tend to have diverse views, going to as many as 31 different locutions in McBurney (2002). We are for a more parsimonious approach, in the spirit that the effort of distinguishing between moves of different types is only justified by a correspondent relief in the effort of understanding the intention of the move. We distinguish among four main move types, some of them comprising a number of subtypes.

Move Types

1. **Assertions:**

   In this type we include all general assertive utterances an agent can perform, that is all the moves in which the agent makes a point. By varying the parameters of the move, an assertion can be distinguished in:

   (a) *Claims*, that is statements which, although might address previous points in the conversation, are put forward with the aim of making a new point. The speaking agent puts forward a claim in move i if the R_i element of the move is empty, that is:

   $$\langle i, S_i, H_i, Assert, \emptyset, C_i \rangle$$

(b) *Acknowledgements*, that is statements which accept another agent's point, by in a sense "re-stating" is. The speaking agent puts forward an acknowledgement in move i:

$$\langle i, S_i, H_i, Assert, k, C_i \rangle$$

if the move refers to a move $k < i$:

$$\langle k, S_k, H_k, Assert, R_k, C_k \rangle$$

and it is true that: $H_i = S_k$, and $C_i = C_k$[7].

(c) *Replies*, that is statements which reply to questions previously posed. The speaking agent replies to a question in move i:

$$\langle i, S_i, H_i, Assert, k, C_i \rangle$$

if the move refers to a move $k < i$:

$$\langle k, S_k, H_k, Query, R_k, C_k \rangle$$

and it is true that: $H_i = S_k$.

(d) *Disputations*, that is statements which explicitly disagree with previously made points. The speaking agent contests a point in move i:

$$\langle i, S_i, H_i, Assert, k, C_i \rangle$$

if the move refers to a move $k < i$:

$$\langle k, S_k, H_k, Assert, R_k, C_k \rangle$$

and it is true that: $H_i = S_k$, and $C_i = \neg C_k$.

[7]Note that the two contents will not necessarily be identical, as different mental modalities can be used. Here we therefore assume that the inner content of the mental modalities is the same, or is the opposite in the case of negations.

(e) *Withdrawals*, that is statements which explicitly deny one's own previously made points. The speaking agent puts forward a withdrawal in move i:

$$\langle i, S_i, H_i, Assert, k, C_i \rangle$$

is the move refers to a move $k < i$:

$$\langle k, S_k, H_k, Assert, R_k, C_k \rangle$$

and it is true that: $S_i = S_k$, and $C_i = \neg C_k$.

2. **Queries:**

In this type we include one agent's queries to which the interlocutor is supposed to reply with an assertion. These comprise:

(a) *Open Questions*: these are requests for information, on items the querying agent does not suppose previous knowledge. They have the form:

$$\langle i, S_i, H_i, Query, \emptyset, C_i \rangle$$

where the content of the move contains uninstantiated variables: $C_i = \lambda(x, y, z, \ldots)$.
A reply to this questions would provide an instantiatiation of the variables of interest. Note that the open questions do not refer to any previous move in the dialogue.

(b) *Requests for Argument*: these are requests for an argument in support of the claim expressed in the content of the move. They have the form:

$$\langle i, S_i, H_i, Query, \emptyset, C_i \rangle$$

where C_i is a completely instantiated clause. This move type represent utterances like *"Query(Fruit is healthy)"*, which basically asks why is it that *Fruit is healthy*. Note also that the utterance *Is fruit healthy?* is captured by YN queries (see below). A request for an argument is made with respect to one generic claim, and does not refer to any previous move in the dialogue, as opposed to what happens with the:

(c) *Challenges of Claim*: these are requests to provide an argument supporting a claim previously made by another agent. They have the form:

$$\langle i, S_i, H_i, Query, k, \emptyset \rangle$$

where $k < i$ refers to a move:

$$\langle k, S_k, H_k, Assert, R_k, C_k \rangle$$

in which $H_i = S_k$. Note that no hypothesis is made on the Assert sub-type of the referred move.

3. **YN queries:**

   We have included in a class of their own closed questions, that is questions whose only reply can be either Yes (True) or No (False). They have the form:

$$\langle i, S_i, H_i, YNQuery, \emptyset, C_i \rangle$$

   where C_i does not contain any uninstantiated variable.

4. **YN:**

   Finally, we include the moves corresponding to the answer to a YN-question. They have the form:

$$\langle i, S_i, H_i, YN, k, C_i \rangle$$

   where $k < i$ refers to a move:

$$\langle k, S_k, H_k, YNQuery, \emptyset, C_k \rangle$$

   in which $H_i = S_k$ and either $C_i = C_k$ (Yes) or $C_i = \neg C_K$ (No).

The reason why the yes/no queries are treated differently from other queries will be clear when defining our dialogue games.

The move set is inspired by the theoretical work discussed at the beginning of this Chapter, but has also a more practical rationale. The move set exploits the combination of properties that can possibly be associated with the move components, as shown in Table 9. All the combinations that are not shown in the table (for instance of an Assert move that refers to a YN move) are invalid.

Move Type	Referred Move Type	Referred Move Speaker	Move Content	*Resulting Dialogue act*
Assert	∅	NA	any	*Claim*
Assert	Assert	Other	same than the referred move	*Acknowledgement*
Assert	Assert	Other	opposite to the referred move	*Disputation*
Assert	Assert	Self	opposite to the referred move	*Withdrawal*
Assert	Query	Other	any	*Reply*
Query	∅	NA	with variables	*Open question*
Query	∅	NA	no variables	*Request for argument*
Query	Assert	Other	no variables	*Challenge of a claim*
YNquery	∅	NA	no variables	*Closed (YN) question*
YN	YNQuery	Other	no variables	*YN reply*

Table 9: Dialogue Move Types

Commitment Rules

Once the dialogue moves have been described, rules must be defined that guide the way the agents "react" to each move. This typically involves the modification of what various authors call the "commitment store" of the participants. This modification basically is described by a set or rules that govern how the two players' "mental state" is changed after each move. A more specific description of the notion of mental state is left until Chapter 4; here we will simplify the notion by saying that all agent's beliefs, desires etc. are contained in this set.

Definition 5 (Commitment Store) *We defi ne the commitment store of an agent/player A as a triple:*

$$\langle S_A, O_A, Q_A \rangle$$

where:

- S_A *is a representation of A's mental state;*
- O_A *is the set:*

$$O_A = \{O_A^1, O_A^2, \dots O_A^{N-1}\}$$

  where N is the number of dialogue participants, and each O_A^i represents the mental state that agent A attributes to agent i. In case of two-player games, $O_A = \{O_A^1\}$.
- Q_A *are the requests made to A by other agents which are still open. These are the questions other agents have posed to A and to which A still have to reply.*

We make the simplifying hypothesis that the players do not need to acknowledge each move, so the appropriate update of the commitment stores is done without confirmation. We define the rules that manipulate these sets, for each of the possible move types, as follows:

1. **Assert**

- Following the assertion:

$$\langle i, S_i, H_i, Assert, \emptyset, C_i \rangle$$

C_i is entered in the components S_{S_i} and $O^{S_i}_{H_i}$ of the commitment store of S and H respectively.

- Following an assertion:

$$\langle i, S_i, H_i, Assert, k, C_i \rangle$$

in reply to a previous question, $k, k < i$:

$$\langle k, S_k, H_k, Query, \emptyset, C_k \rangle$$

then

  - C_i is entered in the components S_{S_i} and $O^{S_i}_{H_i}$ of the commitment store of S and H respectively, and
  - i is deleted from the component Q_{S_i} of S's commitment store.

- Following an assertion:

$$\langle i, S_i, H_i, Assert, k, C_i \rangle$$

in reply to a previous question, $k, k < i$:

$$\langle k, S_k, H_k, Query, \emptyset, l \rangle$$

where $l < k$ is the move:

$$\langle l, S_l, H_l, Assert, R_l, C_l \rangle$$

(that is k is a challenge of the claim expressed in l), then

  - C_i is entered in the components S_{S_i} and $O^{S_i}_{H_i}$ of the commitment store of S and H respectively, and
  - the justification: $C_i \rightarrow C_l$ is entered in the components S_{S_i} and $O^{S_i}_{H_i}$ of the commitment store of S and H respectively.

- **Query** or **YN Query**

Following a question:

$$\langle i, S_i, H_i, Query/YNQuery, R_i, C_i \rangle$$

then i is added to the component Q_{S_i} of S's commitment store. No hypothesis is made on the knowledge or otherwise of C_i by the S_i agent.

- **YN**

  Following a YN answer:

  $$\langle i, S_i, H_i, YN, k, C_i \rangle$$

  to a previous query $k, k < i$:

  $$\langle k, S_k, H_k, YNQuery, \emptyset, C_k \rangle$$

  then:

  – C_i is entered in the components S_{S_i} and $O_{H_i}^{S_i}$ of the commitment store of S and H respectively, and

  – i is deleted from the component Q_{S_i} of S's commitment store.

We will elaborate on the way claims are "entered in the commitment store", in Ch. 4 and 7. We would like, however, to anticipate here that, as opposed to many dialogue game systems, we defined no explicit way to deal with the "retraction" of a move. An agent may, nevertheless, assert P at some point, and then perform a withdrawal of $\neg P$ at a later moment in the conversation. In this case we would enter $\neg P$ in the commitment store of the agent, but would not delete P as a consequence. Instead, an appropriate mechanism would be used, to ensure that the commitment stores do not become inconsistent. This, we believe, better represents natural dialogues: one cannot assume that the retraction of a claim would just erase the fact that the claim has ever been made from other agents' mind. In other words, the situation in which an agent first asserts P, and then negates it, is different, we believe, from the situation in which the agent asserts $\neg P$ in the first place, and this should be captured by the system.

3.3.3 Dialogue Rules

In Walton's definition, a dialogue is a "goal-directed conventional framework in which two speech partners reason together in an orderly way" (Walton 1998, p.3). The hypothesis is the one of cooperative argumentation: even if in conflict with each other, the two partners will behave fairly in their alternation of moves. Dialogue rules are therefore needed to make the "orderly way" explicit. The metaphor of the "game" helps here once again: if each dialogue turn is seen as a "move", then one can define, for each move type, which other move types are allowed to follow, either by the same speaking agent or by the other agent in reply. Also, one should define the modalities by which turn passing is performed: how many moves (and their sequencing) one agent is allowed to make before passing the turn, and in which order the agents can enter into the dialogue: shall they all in turn say something? Shall they ask for permission to speak (especially for many party dialogues)? And so on. The explicit definition of these constraints create a sort of “board” on which the game can be played.

Dialogue Game Boards

We define a game board as a state transition diagram[8]. In the diagram, we associate a state with the situation in which the dialogue is after an agent has performed a move of a given move type. As we consider a general situation in which agents can or can not be allowed to perform any move type, we will duplicate states representing move types for each agent who is allowed to perform them. We will therefore have at most as many states as the number of players multiplied by the number of move types. In the case of two-player games, we will have a number of states equal at most to double the number of move types.

If the game is in one particular state, a player can move to reach another state if:

- an arc connects these two states, and
- the player is allowed to traverse such arc.

The permission to traverse an arcs is given by a label on the arc, which can be one of the following:

1. an arc is labelled K if it can be traversed by the current speaking agent, should the agent decide to keep the turn (if there are no arcs of this sort departing from the current state, the current speaking agent must pass the turn);
2. an arc is labelled G if it can be traversed by an agent different from the current speaking one, should the agent take the turn in the dialogue;
3. an arc is labelled C if it can be traversed by an agent different from the current speaking one, should the agent take the turn in the dialogue, but the arc represents a move that changes the flow of the conversation.

The notion of changing flow moves is intended to capture that, despite the fact that there exists an internal “coherency” of the game if appropriate sequences of moves are executed, one can imagine a situation where such coherency can be disrupted, for a reason. A typical example concerns the position of questions: after a question, an appropriate (coherent) subsequent move is a reply to the question (type

[8]Similar constraints on turn sequencing, whether represented as a network or as a table, are widely used in conversational systems (e.g., see Frohlich & Luff (1990) who propose a general table of turn sequence constraints based on human conversation).

G arc), but the interlocutor might need some clarifications in order to reply, so another question can be accepted (type C arc). The appropriateness of a move is not defined absolutely, but depends on the type of interaction that the game board is modelling, as will be clear later on. Some of the states are labelled as **start states**, and are the only ones allowed to start the conversation.

More formally:

Definition 6 (Game Board) *Let $P = \{p_1 \dots p_N\}$ be the number of players in the game and $M = \{m_1 \dots m_L\}$ the move types allowed. A game board G is a set of game states g_i:*

$$G = \bigcup_{i=1}^{N \times L} g_i$$

with

$$g_i = \langle p_i, m_i, K_i, G_i, C_i \rangle$$

where $p_i \in P$ is the speaking player, $m_i \in M$ is the move type associated with the state (the type of the move just performed by p_i), and $K_i, G_i, C_i \subseteq M$ are respectively the move types the speaking player can perform to keep the turn, and those of the regular and changing flow moves that the player whose turn is next can perform if taking the turn (for games with more than two players we assume a fi xed turn taking sequence).

For completeness, a dummy g_0 is defined, where p_0 is the player that conventionally starts the conversation, $m_0 = NIL$, K_0 are the start states for the first speaking player, and $G_0 = C_0 = \emptyset$.

Note that we assume that $G_i \cup C_i \subseteq M$, so we envisage cases in which a move is forbidden altogether to the agent taking the turn. However, in no case it is allowed that $G_i \cup C_i \cup K_i = \emptyset$.

Two states $g_i = \langle p_i, m_i, K_i, G_i, C_i \rangle$ and $g_j = \langle p_j, m_j, K_j, G_j, C_j \rangle$ are connected (g_j is reachable from g_i) if and only if $m_j \in K_i \cup G_i \cup C_i$.

If $g_i = \langle p_i, m_i, K_i, G_i, C_i \rangle$ and $g_j = \langle p_j, m_j, K_j, G_j, C_j \rangle$ are connected, then $p_i = p_j$ if and only if $m_j \in K_i$.

Graphical Notation

Figure 4 shows the graphical notation we use for two-player game boards: solid boxes and arrows are associated with Player 1, while dashed ones are associated with Player 2. From each state, we indicate:

1. type K arcs as arcs with the same line type as the state they depart from (e.g. solid arcs departing from solid boxes);

2. type G arcs as arcs with the opposite line type as the state they depart from (e.g. dashed arcs departing from solid boxes);

3. type C arcs as bold arcs.

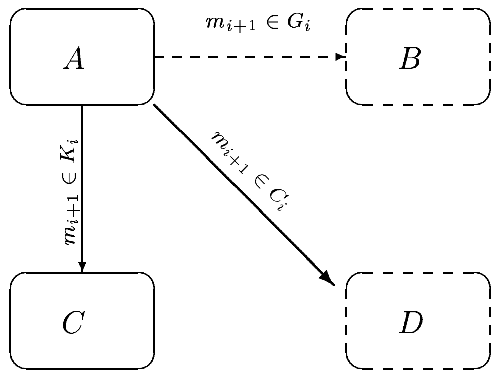


Figure 4: Graphical Notation for Dialogue Game Boards

Therefore, Fig. 4 represents a portion of a dialogue game board such that if Player 1 is in A, after having executed m_i (the current game state is therefore $\langle p_i, m_i, K_i, G_i, C_i \rangle$, with $p_i =$ Player 1) then:

1. Player 1 is allowed to execute m_{i+1}, and keep the turn, if and only if move m_{i+1} is in the set K_i, moving then to state B on the board. The new game state $\langle p_{i+1}, m_{i+1}, K_{i+1}, G_{i+1}, C_{i+1} \rangle$, will have $p_{i+1} =$ Player 1.

2. Player 2 is allowed to execute m_{i+1}, and take the turn, if and only if move m_{i+1} is in the set G_i, moving then to state C. This move will preserve the coherency of the game. The new game state $\langle p_{i+1}, m_{i+1}, K_{i+1}, G_{i+1}, C_{i+1} \rangle$, will have $p_{i+1} =$ Player 2.

3. Player 2 is allowed to execute a non-coherent move m_{i+1}, and take the turn, if and only if move m_{i+1} is in the set C_i, moving then to state D. The new game state $\langle p_{i+1}, m_{i+1}, K_{i+1}, G_{i+1}, C_{i+1} \rangle$, will have $p_{i+1} =$ Player 2.

We also eliminate the need for a specific start state, by shadowing the boxes representing the moves that are allowed to the first speaker to start the conversation.

Which Specific Game Board?

Most commonly (Carletta 1992; Kowtko, Isard, & Doherty 1992; Pilkington *et al.* 1992), when modelling various types of interaction, defined on the basis of the main goal of the dialogue partners in each interaction, different boards are defined, in order to be able to play different games accordingly (e.g. Inform game, Negotiation game and so on). Usually, this dialogue model would impose the constraint that a game has to be completed before moving to another one. Or, it might be allowed to have a nesting of games, subjected to explicit agreement by all the parties, and games are played in a sort of stack fashion, where a game must be completed before resuming the immediately previous one. A typical example of the latter situation is the need for clarification sub-dialogues, for instance on the meaning of a concept, before being able to proceed with, say, a negotiation. Such models assume that the games have so many different characteristics that they can hardly be intermingled. Agents should therefore "declare" the game they are willing to play beforehand, and wait for their opponents' agreement before the game can start.

While understandable, in terms of simplicity of implementation, from the point of view of modelling artificial agents, this approach considerably limits the flexibility and variety of dialogue types, as it does not permit agents' idiosyncratic behaviour, but forces agents to follow the main script at each stage of the interaction. For instance, even when allowing the main dialogue to be interrupted, to start a clarification dialogue, the players need to conclude the subdialogue before going back to the main one. We follow Burton & Brna (1996) in taking a different approach, and we disregard, at a modelling level, the different types of interaction. The aim is to model the whole dialogue as a single, large game, where all the possible theoretical games are interconnected, and the players can at no cost pass from one game to another. However, we keep track that there has been a change in the game normal flow, by means of the concept of less appropriate, or less coherent moves: an agent changes game if it executes a C-type move (it traverses a bold arc, in the graphical notation). In addition, in the model we consider, as part of the board game, only high level move types (Assertion, Queries, YNQueries, and YN) and we treat all move sub-types of the same type as equivalent. This leads to a high level of simplification in the dialogue management, and more closely mimes natural communication among agents, at the same time accounting for unconventional situations: it is easier to model, for instance, the behaviour of a non cooperative agent which avoids answering a question. As a shortcoming, again similarly to Burton & Brna (1996), by using only a coarse grained definition of the moves for the game board, we lose the opportunity to enforce a greater control on the actual semantics and pragmatics of the moves. For the concepts of "allowed", "changing game" and, by exclusion, "forbidden" moves permit defining the concept of "structural" coherence of the dialogue, a notion that expresses the fact that the dialogue is coherent from a synctactical point of view, that is the moves are correctly ordered, but it does not capture whether the dialogue actually "makes sense". For the same reason, we do not represent in our dialogues the fact that the speaker can change the flow of the game: this has only sense if we consider the meaning of what is said (e.g. a change of topic move performed by the same speaker), but not if we stop at the synctactic level. However, we do maintain that one should disregard the semantics of the moves altogether: indeed, this is simply controlled at a different "level of abstraction", that is the rhetorical argumentation level we have described in Ch. 2. We will see in Ch. 5 how the two levels are combined to capture all the features of the dialogue.

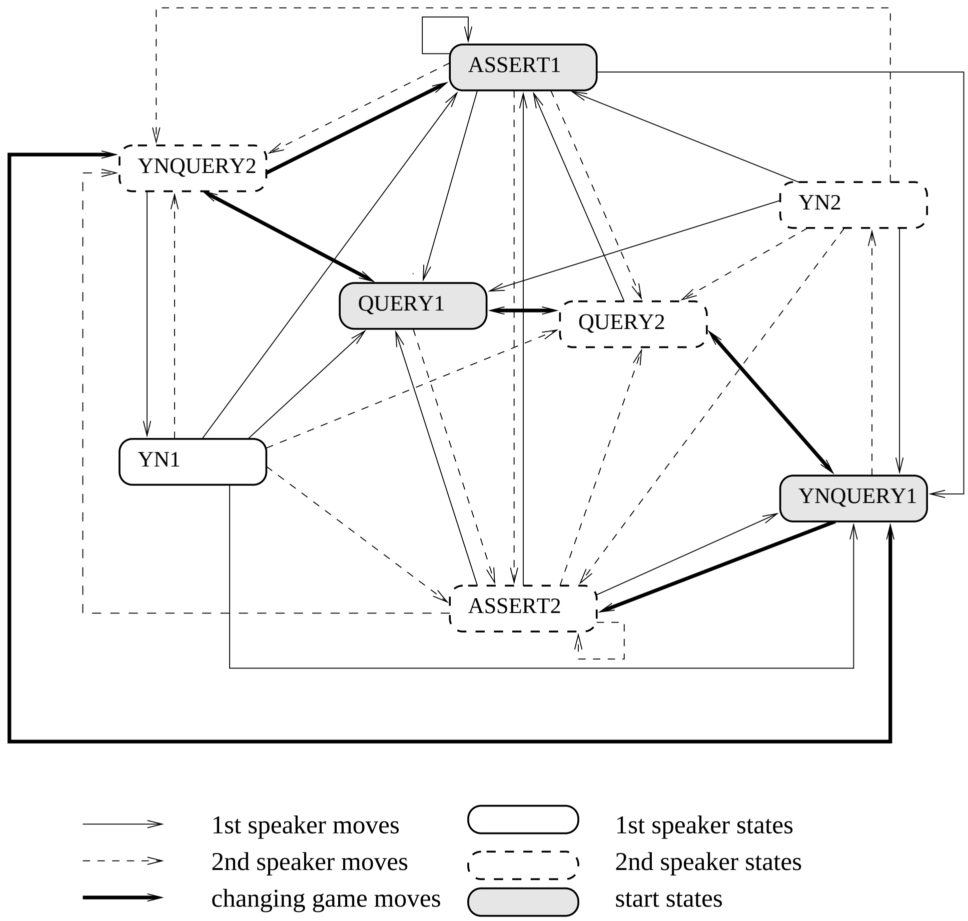


Figure 5: A dialogue game

As to which specific game is preferable, we have no specific preferences: the definition allows for a great variety of games, by opportunistically changing the content of the admissible move sets. We present here what we think is the most generic one, that is the one that imposes less constraints on the two player. The game is shown in Fig. 5, where we follow the notation introduced above whereby states associated with the first player/speaker are identified with a solid box, those associated with the second speaker with a dashed box. The shadowed boxes in the figure represent valid starting states for the first speaker. In the figure, the actual move performed by the agents is associated with the state rather than with the arc leading to it. So a “Y/N” state corresponds to the state which the system is in after one of the agents performs a “yes” or a “no” speech act, and in particular a solid “Y/N” state is reached when the first speaker has performed as “yes/no” speech act, and a dashed one is reached when the same is done by the second speaker. Similarly, when speaking of, for example, an “assert move”, we intend the crossing of the arc which, from the current state, leads to one of the “assert” states, depending on the speaker. This convention permits a simplified encoding of the state diagram, avoiding superfluous labels on the arcs.

$$
\begin{aligned}
g_0 &= \langle \text{Player1}, NIL, \{\text{Assert}, \text{Query}, \text{YNQuery}\}, \emptyset, \emptyset \rangle \\
g_1 &= \langle \text{Player1}, \text{Assert}, \{\text{Assert}, \text{Query}, \text{YNQuery}\}, \{\text{Assert}, \text{Query}, \text{YNQuery}\}, \emptyset \rangle \\
g_2 &= \langle \text{Player2}, \text{Assert}, \{\text{Assert}, \text{Query}, \text{YNQuery}\}, \{\text{Assert}, \text{Query}, \text{YNQuery}\}, \emptyset \rangle \\
g_3 &= \langle \text{Player1}, \text{Query}, \emptyset, \{\text{Assert}\}, \{\text{Query}, \text{YNQuery}\} \rangle \\
g_4 &= \langle \text{Player2}, \text{Query}, \emptyset, \{\text{Assert}\}, \{\text{Query}, \text{YNQuery}\} \rangle \\
g_5 &= \langle \text{Player1}, \text{YNQuery}, \emptyset, \{\text{YN}\}, \{\text{Assert}, \text{Query}, \text{YNQuery}\} \rangle \\
g_6 &= \langle \text{Player2}, \text{YNQuery}, \emptyset, \{\text{YN}\}, \{\text{Assert}, \text{Query}, \text{YNQuery}\} \rangle \\
g_7 &= \langle \text{Player1}, \text{YN}, \{\text{Assert}, \text{Query}, \text{YNQuery}\}, \{\text{Assert}, \text{Query}, \text{YNQuery}\}, \emptyset \rangle \\
g_8 &= \langle \text{Player2}, \text{YN}, \{\text{Assert}, \text{Query}, \text{YNQuery}\}, \{\text{Assert}, \text{Query}, \text{YNQuery}\}, \emptyset \rangle
\end{aligned}
$$

Table 10: Definition of the game board for the game in Fig. 5

The game leaves a good amount of freedom to the dialogue participants as to which move to perform next. The players are allowed to avoid answering questions (bold arcs after a query state), and more than one move can be made by the same participant before passing its turn. The only restrictions are that:

- players are constrained to pose only one question per turn;
- the question must be the very last move of a turn: after a question the current speaker must pass the turn
- YNreplies cannot be executed unless a YNquery has been posed.

Finally, as can be noticed from Fig. 5, the game is perfectly symmetrical with respect to the two participants. A formal definition of the game board is given in Table 10.

3.3.4 Goals of the Dialogue

After defining the participants, the move types and the rules that govern the rule use, to complete the description of our dialogue model we need to define the goals of the dialogue. The goal, as we have seen previously, is a way to describe the dialogue type. For instance, in Table 8, Walton & Krabbe's classification of dialogue types (persuasion, inquiry, etc.) is determined by the goals of the participant (persuade, prove hypothesis, etc.). Rhetorical dialogues, if thinking in terms of the classification in Table 8, are *persuasion* dialogues: the two parties start with a difference in opinion, and the goal of the dialogue is to clarify this difference, coming to an agreement, while the goal of each participant is to persuade the opponent that their point of view is the one that counts. In particular, rhetorical dialogues are centred around a difference in the evaluation of a particular issue: one party may think a state of affairs is valuable and worth achieving, while the other one thinks the opposite. In terms of the definitions given in Chapter 2, we can say that the rhetorical dialogue starts with the contrasting evaluations:

$$E(c, p_1) = good \text{ and } E(c, p_2) = bad$$

of a concept c which are held by the two dialogue partners, where not necessarily $p_1 = p_2$, and has the aim of coming to an agreed evaluation of c, perhaps from an altogether different perspective:

$$E(c, p_3) = e$$

where the goal of each participant is that the agreed evaluation e coincides with their own starting one.

3.4 Discussion

In this Chapter we concentrated on the dialogical dimension of rhetorical argumentation, and we have analysed various approaches to the issue, both theoretical and computational. We have proposed a dialogical framework, in the spirit of a dialogue game, for the model of rhetorical dialogues. The game is meant to capture only the structure of dialogical argumentation, and not its “content”, and in this we are supported by the view, among other, of Burton & Brna (1996). Therefore, our dialogue games as such do not model, for instance, the type of rules that are more proper of computational dialectic frameworks, on why and when one of the two partners can put forward one argument. Our structure does not seek to determine whether the argument proposed is “successful”, nor implements the notion of “burden of proof” which is typical of pragmatic and informal logic approaches (Walton 1996).

However, when seen in the context of our complete framework, one can easily realise that such notions are in fact captured by the model of rhetorical argument given in Chapter 2: the very concept of constraints to the application of argumentative schemata is meant to represent the restrictions that have to guide the dialogic partner in the way arguments are put forward. In this way, we separate the epistemic knowledge related to the argumentation process and the domain of application, from the mechanisms of the dialogue itself, which is then seen only as a way to orderly arrange the utterances of the two partners, thus providing a framework of expectations to support comprehension and maintain focus. We will see in Chapter 7 how, in the architecture of a rhetorical agent we propose, the decisions on what to say and why are internal to the agent, who “plans” to achieve communicative goals, and are separated from the decisions on turn taking, which are in some way imposed by the dialogue game board. Separation also means independence of modelling, and a higher level of interoperability, for one can easily re-use the schemata architecture in order to produce monologues rather than dialogues, as well as re-using the dialogue architecture for interaction among less discourse-oriented agents, as often happens in Multi Agent Systems frameworks (McBurney & Parsons 2001).

3.5 Conclusions

The term dialectic has had various and diverse interpretations over time, and the very definition given by Aristotle has been object of controversy. As noted, among others, by Cattani (1990), the Greek term that Aristotle uses at the beginning of his *Rhetoric*, to contrapose it with dialectic, gives rise to ambiguous interpretations: for Aristotle rhetoric is *antistrophos* of dialectic, an adjective which may mean both "analogous" (the term used, among others, in the Italian translation) and "counterpart" (the term used in the English one). The Aristotelian definition of dialectic refers to the "art of discussion", where dialectic reasoning is based on premises which are accepted by a general audience, but still, as opposed to demonstrative reasoning, not objectively true, although at the same time not evidently deceptive.

Cattani (1990) maintains that even Perelman and Olbrechts-Tyteca's New Rhetoric is in fact based on Aristotelian dialectics, but the ambiguity of the term dialectic has led the authors to use the term rhetoric instead. In fact, the very idea of providing a classification of "moves", as we have seen in Sec. 2.2, is dialectical in its essence, although the importance given to the audience, as well as the stress given to premises containing value judgements, is inherently rhetorical.

We feel we agree with Cattani (1990) that rhetoric is *also* dialectic, in the sense that persuasion also involves comparing and contrasting different premises and thesis, and with this in mind we have called our prototypical model "dialectic" (see Ch. 7).

Chapter 4

The Arguing Partners

> You act, and you know why you act, but you don't know why you know that you know what you do.
>
> - Umberto Eco, *The Name of the Rose*

Perelman and Olbrechts-Tyteca considered it essential for the arguer to have an accurate image of the audience in mind:

> "In argumentation, the important thing is not knowing what the speaker regards as true or important, but knowing the views of those he is addressing" (NR § 5).

This Chapter goes therefore back to the task, left suspended in Ch. 3, of modelling the dialogue partners in terms of their mental attitudes. It especially explores the problem of mental models where an agent needs to represent and reason about other interlocutors' beliefs, in addition to one's own. It analyses findings of different schools of thought about reasoning with and on beliefs. It then presents a simple but efficacious belief model that allows us to represent an agent's various mental attitudes.

4.1 Mental States

The question of what mental phenomena are, how they can be explained, and hopefully modelled, has troubled philosophers of mind for many decades[1]. Since Descartes' intuition that mind and body are two different essences, philosophers have tried to explain how these two essences can interact and influence

[1]In this report, we do not intend to give an account of the fundamental issues of the philosophy of mind, as we would barely manage to scrap the surface of theories whose roots can be found in works by the like of Aristotle (about 350 BC 1968), Carnap (1947), Dennett (1978), Descartes (1931), Russell (1921), or Wittgenstein (1953). For the interested reader, a good introduction is the one by Rosenthal (1991).

each other[2]. One position which is bound very tightly with mainstream artificial intelligence is "functionalism". The main claim of functionalists is that one way to understand how the mind works is to concentrate on the "function", or causal role, of mental states. In a more radical version of the theory, "computational functionalism", Putnam (1975) argued that a state of the mind can be associated with a state of a Turing Machine, where beliefs, desires and the like are inputs to "algorithms" which output actions[3]. The standard enunciation of the model is due to Fodor (1975), who also argued that a state of the mind should rather be associated with an "operation" of the machine, its "computational" state. Therefore, a mental state depends on the process the machine is computing, rather than its internal state. In a further variation of functionalism, the mind is thought to work as a system of cognitive functions, each of which can be reduced to less complex functions, until very simple tasks are reached. This notion is very similar to the "society of minds", advocated by Minsky (1985). Whichever the interpretation, if the functionalists' claim is accepted, the problem is shifted towards the one of providing a more precise characterisation of these "mental states".

A fundamental characteristic of mental states, as opposed to physical states, is, as suggested by Franz Brentano (1874 1973), that they can "contain" other states, or other objects, to which they refer. This characteristic is called by Brentano, inspired by the Scholastic movement of the Middle Ages, "intentionality", from the Latin sense of *intentio* as "reference", "relation". Thus the mental state expressed by the sentence "Floriana believes the thesis is complete" intentionally refers to the state "the thesis is complete". An interesting property of intentionality, Brentano observed, is that of *intentional inexistence*, that is the fact that the reference may be to an object which is not taken to be real, or to exist (Brentano talked in fact of *quasi-relations*, as one of the relation's terms may not exist). The state "Floriana believes the thesis is complete" does not say anything about the actual existence of a thesis: it may exist or it may well not. Intentional states show two other interesting properties (R.M. Chisholm 1957): they do not imply the truth or falsity of referred states (the thesis may or may not be complete), and they are sensitive to the way in which the reference to the given state is expressed (even if the thesis will be stored as Report No. XYZ, it would be inappropriate to say "Floriana believes Report No. XYZ is complete").

A more precise account of what a theory of mental states should present is due to Fodor (1978), who describes five properties that such theory has to account for. The first reiterates that a mental state should intuitively represent a relation, between an agent and "something else". Therefore the sentence "Floriana believes the thesis is complete" represents the fact that agent "Floriana" is in the "belief"-type relation with the token "the thesis is complete", while "Floriana intends to finish off the thesis" represents

[2]This "dualistic" intuition has been opposed, most notoriously, by Damasio (1994).
[3]Among the arguments against this position, a recent one is given by Eliasmith (2002).

an "intend"-type relation between "Floriana" and the token "finish off the thesis". Second, the theory should explain the isomorphism that appears to exist between what can be the object of a mental state and what can be the object of a speech act. It cannot be just accidental, Fodor notes, that sentences like "Floriana says the thesis is complete" and "Floriana promises to finish off the thesis" are acceptable whenever the sentences in the examples above are. Third, the theory should account for the "Fregean opacity condition", that is the fact that the truth value of the sentences expressing a mental state cannot be a function of the truth value of the state referred to by the mental state: the inappropriateness of the sentence "Floriana believes Report No. XYZ is complete", even if Report No. XYZ *is* the thesis, should be captured by the theory. Fourth, the objects of a mental state should have a logical form. This is named by Fodor the "Aristotle's Condition". As humans do relate mental attitudes on the basis of their content (what is believed, wished, etc), the theory should allow referring to the form of this content. It should recognise, for instance, conjunctions, causal chains, and so on, between what can be believed. This may seem to contradict the fourth, Fregean condition: it seems pointless to think of the logical form of the belief content, if one cannot rely on truth values to reason about it in logical terms. But Fodor's point is rather that we have to be able to reason about mental states in some way: we need to represent the fact that some beliefs interact with other beliefs, and may contribute to create desires, which can be transformed in intentions, and so on. Finally, the theory should find confirmation in empirical findings on how mental processes work, in line with the modern philosophy of mind, that moves away from mere conjectures and approaches problems with a more scientific attitude.

Fodor's (and Chisholm's) conclusion on how a theory of mental states should look like is that one should use a special sort of "intentional language" to manipulate their description, a language which is not similar to any human language, as several paradoxes would result, but which is an *internal* "Language of Thought", or "Mentalese" (Fodor 1975). Indeed, the notion of internal language seems to entangle more and more mental attitudes and communication language, that is what can be believed and what can be said, to the extent that for some scholars, such as Davidson (1975), an agent possesses mental attitudes if and only if that agent has the ability to communicate.

4.1.1 Propositional Attitudes

If the hypothesis of a language of thought is accepted, deciding how to express sentences in this language is the next step. One way to define a language talking about mental states is in terms of *propositional attitudes* (PA), a term introduced by Russell (1912)[4]. These are "sentences" like:

$$\text{Agent Verb}_{\text{A}} \text{ “that C”}$$

[4]Not all scholars subscribe this view, see for instance the argument of Ben-Yami (1997).

where Verb_A is an "attitude" verb, like believing, hoping, wanting, saying, and so on, and "that C" is a proposition, usually named that-clause, or *t-clause*. For example:

$$\underbrace{\text{Floriana}}_{\text{Agent}}\ \underbrace{\text{believes}}_{\text{Verb}_\text{A}}\ \underbrace{\text{that the thesis is complete}}_{\text{t}-\text{clause}}.$$

PAs are therefore sentences that are intentional, or are used intentionally, in Brentano's sense, as they are an expression of an attitude relation between an Agent and a *t-clause*.

Richard (1990) observes that there are two ways to look at the nature of these *t-clauses*. They can either be structureless propositions, or have internal structure. In the first case, they can be identified with the semantics of a logical proposition, for instance with a set of "possible worlds" (Kripke 1963). As the structure of the *t-clause* is not of interest, all the structures which correspond to the same world are in this case identical. Therefore, the following three *t-clauses*, although having different structure, represent the same world:

1. "Either Floriana does not complete the thesis, or she is happy",
2. "It is not possible that Floriana completes the thesis and is unhappy",
3. "If Floriana completes the thesis she is not unhappy"

This view is typically identified as the Russellian one.

A second account of the nature of the *t-clause* holds that their structure does matter. Thus, even if two *t-clauses* are logically equivalent, if their structure differs, then they are not naming the same thing. The three different versions of the beliefs on Floriana's happiness, or lack of it, are three different beliefs. The belief ascription in this view, which is typically associated with Frege, depends strictly on the way in which the believer represents the belief.

If this view, that Richard calls "sententialism", is accepted as a more psychologically plausible view, the next step is to understand what the structure of *t-clauses* is. Richard reminds us that authors like Carnap and Lewis are inclined to closely tie the structure of the *t-clause* with the structure of the natural language sentence that expresses it. This view trivially goes around problems like referential opacity. However, as Richard observes, it does not account for some obvious exceptions that concern either particular structures, like "and" clauses (cases in which "A and B" is taken to be different from "B and A" seem rare), or particular attitudes (for instance, "saying" and "believing" seem to react differently to structure changes). Richard's proposal is a "psychological" sententialism, in which what the agent is committed to by the PA is a sort of equivalence class of sentences, rather than a single sentence. The way in which the equivalence class is built is not, however, fixed, as in the case of the possible worlds solution, but will depend on some characteristics of the agent itself. This would take the best of the two worlds, the Russellian and the Fregean views, and would better represent the agent's specific expertise: for a good logician agent, the three sentences in the example above will be the same sentence, while for a less skillful one they will be different. In other words, Richard distinguishes between "what someone believes [...] from how she believes it" (p. 121), the semantic from the pragmatic account of the way beliefs (or other attitudes' ascription) are formed. McKay (2000) reports of a number of objections to "mixed" theories such as Richard's, for instance by Schiffer, on the grounds that in these theories it is difficult to explicitly deal with the way in which the believer represents concepts.

A more precise account of how this might work is given by Barwise & Perry (1983), in terms of their semantic theory of attitudes, although they deliberately avoid talking about *propositional* attitudes[5]. The authors maintain instead that attitude ascriptions are relations between an agent and some *situation*, and that these relations are supported by the cognitive states of the agent.

The debate is complex and interesting, but for the purposes of the present work we will ignore most of the issues, and will accept, by way of summary, that representing PAs should entail representing explicitly the *agent* to whom the PAs are ascribed ("Floriana"), the *type of relation*, the attitude the PA expresses ("belief"), the *content* of the PA, that is a sentence in some language ("the thesis is complete"), which will refer to the *objects* of the PA, that is the "things" the content of the PA is expressing some property about ("the thesis"). We will also accept that the way in which an agent can reason about a PA is idiosyncratic of the particular agent, according to its inclinations and expertise.

4.1.2 Diversity of Propositional Attitudes

Although many scholars claim (or perhaps exactly because of this) that all that can be said of one PA can be applied, *mutatis mutandis*, to all other PAs, the literature does not give justice to the variety of mental attitudes that can be described. We find in fact an abundance of discussion on the "belief" attitude, even concentrating on sutble variations (see Sperber (1997) for example), a good amount of considerations on the attitudes of "want/desire/intend" (Davidson 1975; Millikan 1986; Bratman 1987), "strive/wish" (Quine 1966) etc., and how they interact with beliefs, often in order to lead to action, but very sparse thoughts on other mental attitudes. In particular, it is hard to find works on the particular differences among attitudes in term of both the agent's mental state and reasoning ability. Therefore questions like: "what is it like to 'remember' something?", or "what is the exact difference between 'assuming', 'believing' and 'supposing'?", or "how to represent 'prejudices' as opposed to 'beliefs'?" have not been

[5]"We think that propositions are an artifact of the semantic endeavor and not the sort of things that people usually talk about, even when reporting someone's attitudes." (p. 178).

answered satisfactorily.

There are some hints towards the understanding of this issue. For instance Kelly (2002) recently examined whether and how “Regret”, “Fear”, “Desire” and “Suppose” can be rationalised, and how they differ from “Belief”. The most accurate attempt, to our knowledge, is however the one by Barwise & Perry (1983), who give a list of principles, that could help both characterise attitudes and, by way of doing it, diversify them:

1. *Veridicality:* An attitude verb has this property if the following applies:

   IF “Agent Verb_A *t-clause*” THEN *t-clause*

   For instance, if “Floriana knows that the thesis is complete” then “the thesis is complete”.

2. *Substitutivity:* An attitude verb has this property if the following applies

   IF “Agent Verb_A *t-clause(X)*” AND “X is Y” THEN “Agent Verb_A *t-clause(Y)*”

   For instance if “Floriana sees that the thesis is complete” and “the thesis is report XYZ” then “Floriana sees that report XYZ is complete” (though she might not know it).

3. *Existential Generalisation:* An attitude verb has this property if the following applies

   IF “Agent Verb_A *t-clause(X)*” THEN there is some x such that “Agent Verb_A *t-clause(x)*”

   For instance, if “Floriana hopes that the thesis is complete” then there is something that Floriana hopes it is complete.

4. *Negation:* An attitude verb has this property if the following applies

   IF “Agent Verb_A *t-clause(not X)*” THEN NOT “Agent Verb_A *t-clause(X)*”

   For instance if “Floriana believes that the thesis is not complete” THEN “Floriana does not believe that the thesis is complete”.

5. *Conjunction Distribution:* An attitude verb has this property if the following applies

   IF “Agent Verb_A *t-clause(X AND Y)*” THEN “Agent Verb_A *t-clause(X)*” AND “Agent Verb_A *t-clause(Y)*”

   For instance if “Floriana believes that the thesis is complete and submitted” then “Floriana believes that the thesis is complete” and “Floriana believes that the thesis is submitted”.

6. *Disjunction Distribution:* An attitude verb has this property if the following applies

   IF “Agent Verb_A *t-clause(X OR Y)*” THEN “Agent Verb_A *t-clause(X)*” OR “Agent Verb_A *t-clause(Y)*”

   For instance if “Floriana sees that either the thesis or the journal paper are complete” then either “Floriana sees that the thesis is complete” or “Floriana sees that the journal paper is complete”.

7. *Distribution of Indefinite Descriptions:* An attitude verb has this property if the following applies

   IF “Agent Verb_A *t-clause(x)*” THEN there is some y such that “Agent Verb_A *t-clause(y)*”

Principle	*See that*	*Know that*	*Believe that*	*Assert that*	*Doubt that*
1	Y	Y	N	N	N
2	O	O	O	O	O
3	Y	Y	Y	Y	Y
4	Y	Y	S	S	S
5	Y	Y	Y	Y	N
6	Y	Y	N	N	Y
7	Y	Y	N	N	N
8	Y	Y	Y	Y	N

Legenda: Y = always applies; N = never applies; O = often applies; S = seldom applies.

Table 11: Behaviour of some attitudes with respect to Barwise and Perry's 8 principles

This is similar to the Existential Generalisation principle, but in this case both x and y are variables, and do not commit the agent to any reference. For instance, if "Floriana sees that some report is complete", then there is some report that Floriana sees it is complete.

8. *Weak Substitution:* An attitude verb has this property if the following applies

   IF "Agent Verb$_A$ *t-clause(X)*" AND "Agent Verb$_A$ that X is Y" THEN "Agent Verb$_A$ *t-clause(Y)*"

   As opposed to the Substitutivity principle, here the concern is on the agent's mental state, and not on the observer's conclusion. For instance if "Floriana believes that the thesis is complete" and "Floriana believes the thesis is report XYZ" then "Floriana believes that report XYZ is complete".

Barwise and Perry propose a preliminary analysis of the behaviour of some attitude verbs on the basis of these principles, and their intuitions are shown in Table 11[6]. The authors also briefly analyse the attitudes of Saying and Inferring, though they do not compare them against the 8 principles.

Attitudes less closely related with epistemic issues, such as emotional attitudes, are perhaps the most neglected in the philosophy of mind, although a recent example is the work by Gilbert (1997), within the development of his account of "Coalescent Argumentation" (Gilbert 1995). There is, however, a vast literature, in psychology, on the study of the cognitive basis of emotions, and how emotions not only are influenced by, but in turn influence beliefs and goals, their content and strength (de Sousa & Morton 2002; Frijda, Manstead, & Bem 2000).

4.1.3 Attitudes about other Agents

The experience of other agents' mental attitudes can only be, as observed by Russell (1948), a second hand one, and one by analogy: we ascribe beliefs, desires and the like to other people on the basis of what

[6]This is in fact a simplifi cation of the original table (Barwise & Perry 1983, p. 195), as our re-elaboration does not show the differences Barwise and Perry make in the way attributive noun phrases can vary (p. 153ss).

we might believe, desire, and so on in the same circumstances. Of course, as we cannot have a precise image of other people's mind, our PA ascription to other agents is doomed to be partial and by approximation. In this thesis, as far as the general problem of defining mental attitudes is concerned, we are not so much interested in the more philosophical issues of how we can conceive to think about other people's mind (Nagel 1974), but rather we want to focus on the problem of how we can "represent" the image of another agent's mind. This perspective is more useful to our ultimate aim of characterising a rhetorical argumentative agent, able to interact effectively with an audience. On this perspective, Johnson-Laird (1983), among others, integrates the notion of beliefs about other agents' in his theory of mental models: a "recursive embedding of mental models" (p. 433) can be used both to represent complex propositional attitudes (for instance for meta-reasoning) and to represent other agents' beliefs. Therefore, if agent A needs to reason about the beliefs of another agent B, then A will have, in its mental model, a "set" representing B's beliefs, which may be not included in the set representing A's own beliefs, though there might be some overlapping elements. This would solve quite elegantly the problem of referential opacity we have seen above. Johnson-Laird does not push the idea further, but, as there appears to be no limit to the number of embeddings, it seems consequential that the same mechanism can be used to represent A's beliefs about B's beliefs about A's beliefs, and so on. This intuition seems to have been adopted by AI researchers developing models of agents.

But the ability to reason about other's beliefs does not suffice. Grice (1957) was perhaps the first in modern pragmatics literature to spell out that, in order for successful communication to take place, it is necessary that the interlocutors share a common ground of beliefs, and know that they share them[7]. This intuition was subsequently more clearly formalised by Schiffer (1972), and adapted by Clark & Carlson (1982) in a notion of *mutual belief*:

Definition 7 (Mutual Belief) *(Clark & Carlson 1982, p. 3) A and B mutually believe that* $p =_{\text{def}}$:

1. *A believes that* p.
2. *B believes that* p.
3. *A believes that B believes that* p.
4. *B believes that A believes that* p.
5. *A believes that B believes that A believes that* p.
6. *B believes that A believes that B believes that* p.

et cetera ad infi nitum.

[7]This view is not shared by other authors, such as Sperber & Wilson (1995) who base comprehension on the notion of "context", and relevance to it. However, the notion of mutual knowledge/belief has proved very useful computationally, as explained in the next Section.

To represent mutual belief, it would therefore seem to be necessary for an agent to keep track of an infinite number of beliefs. Clark & Carlson (1982) object that this is only an apparent problem: agents do not, in fact, store *all* the beliefs of the chain, but they have a way to "compile" the procedure in a mental primitive of the form "A and B mutually believe that p", which they can use by means of inference rules of the like "if A and B mutually believe that p then: (a) A and B believe that p and believe that (a)" (p. 5). This remarkably computational oriented intuition, based on one by Lewis (1969), means that agents have a "mutual belief induction schema" that can be applied whenever they need to know whether they share some common ground, thus overcoming the problem of infinite recursion.

Much debate has been devoted to this issue (Smith 1982), though once again, we leave aside inherently philosophical problems, and we only take note that in order to represent the mental model of an agent we will need to account for some sort of mutuality. We also add that this mutuality should not only take place among beliefs, but, at least in principle, among various other attitudes, so one should be able to talk about mutual desire, etc.

4.2 Computational Approaches

As in the other Chapters of this thesis, we review here the main computational oriented works that have dealt with artificial agents' mental attitudes. We mainly concentrate on the representation of mental attitudes, rather than how they are used for reasoning, and in particular, as we are especially interested in a dialogic situation, we focus in more detail on works which have dealt with mutual beliefs.

4.2.1 Representing Propositional Attitudes

Artificial Intelligence research has by and large embraced the view that a propositional attitude is a relationship between an agent and a "mental object" (Russell & Norvig 2003). The philosophical tension between the sentential and the possible worlds views discussed above is reflected in the presence of two main approaches to the representation of mental attitudes.

The first approach, which implements the possible world view, is through the use of *modal logics* (Hintikka 1962; Moore 1985), that is logics where the predicates representing the propositional attitudes[8] become special "modal" operators, which are defined to be referentially opaque. The semantics of these logics is defined in terms of possible worlds in the sense that it is a model where an agent can conceive of a set of alternative states of affairs, or worlds, according to what the agents has or has not information

[8] In fact these logics typically only implement the epistemic attitudes of *Belief*, *Knowledge* and sometimes *Intention*, but it is, as always, assumed that, by aptly changing the axioms, results can be extended to any attitude.

about (Fagin *et al.* 1996). The modal logic approach is very powerful, as there are robust mathematical/logic bases on which it is grounded, but it does have some shortcomings. First and foremost, from the computational point of view, most of the problem solving methods are inherently intractable (Fagin *et al.* 1996). Secondly, agents modelled in this framework are "perfect reasoners": they believe all the consequences of their beliefs, they believe all valid formulae, they believe all the equivalent versions of the same propositions and so on, a property which is not necessarily desirable in every context. Thirdly, modal logics cannot very well express quantified sentences, such as "Agent A believes all that Agent B believes". To solve the latter two problems, different, "nonstandard" versions of modal logics have been proposed, such as logics including a third attitude of "awareness", or logics of "possible and impossible" worlds, and so on (Fagin *et al.* 1996).

The second approach, which implements the sentential view in the philosophy of mind, consists of *syntactical theories* (Kaplan & Montague 1960; Konolige 1982), where the mental object is represented as, indeed, a sentence. The language used for expressing these sentences varies: it can be first order logic, Horn clauses, default statements, probability statements, semantic networks, and so forth. In these contexts, mental attitudes are identical only if they are represented by identical sentences. Therefore "Believe(Floriana,Complete(Thesis))" and "Believe(Floriana,Complete(ReportXYZ))" cannot be the same belief, because even if Thesis = ReportXYZ, the sentences "Complete(Thesis)" and "Complete(ReportXYZ)" are different. A variation of this approach recognises the fact that not all beliefs are the same, and some are in fact "more basic" than others. So much so that it should be possible to identify *belief bases*, that is minimal, finite belief sets, whose logical closure, that is the set of all their logical consequences, contains the actual beliefs the agent possesses. Belief bases are supposed to be easier to handle when reasoning on, or revising, the mental state of the agent (Nebel 1990).

Some authors augment the expressiveness of the belief representation with some measure of the strength of the belief, the "epistemic entrenchment", as it is named by Gärdenfors (1992b). This can be a *probability* measure, that may express how much the agent is "attached" to the belief in question, or how resistant the agent is to change such belief (Shafer 1976). Or, it can be qualitatively realised as an ordering relation among beliefs (Dubois & Prade 1992, as also proposed by Lewis (1973)). Alternatively, the measure can express some *endorsement* level, for instance the reliability of the belief's source (Galliers 1992; Logan *et al.* 1994). Whichever the interpretation, characterising epistemic entrenchment becomes useful especially when the mental state of the agent is not static, but is subject to update and revision. Also, it can be used to implement a notion similar to the belief bases one, with the same computational benefits (Gärdenfors 1992b).

Syntactical approaches have the important advantage of being computationally tractable, in addition to being more expressive. The main disadvantage, at least for the logic based ones, is that if the sentences they express are "self referential", that is they are about the sentences themselves, the theory produced is inconsistent, and gives rise to paradoxes (Montague 1963; Thomason 1980). This problem was addressed by Morreau & Kraus (1998), and partially solved by characterising a set of sentences which are not self referential.

However, the issue of tractability must be crucial, for, as Gärdenfors (1992b) observes, the sentential is the most common representation of propositional attitudes, especially considering that it has been maintained that "there is a close correspondence between belief sets and possible worlds" that makes the two models "basically equivalent" (Gärdenfors 1992b, while citing Grove (1988))[9].

Representing Diverse Propositional Attitudes

An interesting review on how various mental attitudes have been represented and used by various works is due to Shoham & Cousins (1994). The brevity of the review testifies, as the authors comment, to the scarce interest in this issue by AI developers. The paper identifies four types of mental attitude:

1. *informational* attitudes, including mainly "knowledge" and "belief", but also "awareness" of concepts, and so on.

2. *motivational* attitudes, concerning what drives an agent's action, for instance: goal, plan, intention, want, desire, choice, commitment.

3. *social* attitudes, concerning an agent's rights and obligations, when "living" in a society of agents.

4. *emotional* attitudes, concerning various extra-rational factors.

The informational attitudes, as Shoham & Cousins (1994) observe, are the most studied and understood. Apart from work already discussed above on Belief and Knowledge, it is worth mentioning, in this category the seminal theory by Grosz, Joshi, & Weinstein (1995) on "attention", as opposed to "belief", and the analysis of "doubt" by Carberry & Schroeder (2001). The distinction among various types of beliefs has been analysed by Galliers (1992) and by Reed *et al.* (1996). In specialising the belief attitude, they both propose new categories of informational attitudes: the former considers "communicated" beliefs as opposed to "first hand" and "hypothetical" ones, while the latter consider "factual beliefs" as opposed to "opinions" and "cultural beliefs" (sociocultural maxims).

In the motivational category, most notably, study has been undertaken in the field of agent technology, under the umbrella of the *Belief-Desire-Intention* (BDI) model of agenthood (Rao & Georgeff 1991), a

[9] For a more detailed comparison between the two approaches, see Davis (1990).

model inspired by the work by Bratman (1987) on human rationality. Agents are characterised in terms of the three attitudes of Belief, Desire and Intention, which are typically modelled in terms of modal logics axioms (Wooldridge 2000). On the basis of these attitudes, agents can reason about their world, take decisions on the course of their actions, and negotiate with other agents if they need to collaborate, or if their plans interfere with each others (Parsons, Sierra, & Jennings 1998). Among motivational attitudes, we would also include evaluative mental attitudes, like “desirable that X” (Huang & Bell 1996; Bell & Huang 1996), although more often these evaluations are implemented, similarly to epistemic entrenchment, by “weighing” or “labelling” other mental attitudes (Fox & Parsons 1997; Carenini 2001; Galliers 1992).

Social attitudes are almost exclusively the realm of *deontic logic* (von Wright 1951), a logic aimed at modelling attitudes like “ought to”, “allow to”, and the like, which has found vast application in legal informatics (see Wieringa & Meyer (1993) for a survey of computational applications of deontic logic).

As to emotional attitudes, these have recently attracted the attention of the community working on the “believability” characteristics of agenthood (Cañamero 1998; Picard 1996; Velásquez 1999; Johnson 2001). The most established works are grounded on the theory by Ortony, Clore, & Collins (1988), and include, among others, the works by Bates (1994),Elliott (1992), Sloman (1999).

4.2.2 Changing Attitudes

In this thesis we do not focus too much on the process of an agent having to change its mental state to reflect new conditions in the world. We will therefore mention only very briefly the main issues involved in considering this aspect of modelling an agent’s mental state.

The problem of changing view can be reduced to the problem of deciding how an agent’s mental state has to be held together. There are historically two views on the matter. These views concern mainly the attitude of “believe”, with the usual assumption they could be adapted, with some modifications, to other attitudes as well. The *justifi cations*, or *foundations* view holds that an agent maintains a belief as long as its mental state “provides” a reason, a justification, for this belief. Thus, as soon as all the justifications for a belief cease to exist in the agent’s mental state, for instance because the agent comes to know that some premises were false, the belief is discarded. Conversely, the *coherence* view holds that an agent maintains a belief as long as it does not make its belief base contradictory. This view advocates a minimal change approach: even if the agent undermines all justifications for a given belief, if the belief is still coherent with the rest of its belief base, it will not be discarded. The most widely known theory elaborating the coherence approach is due to Alchourrón, Gärdenfors, & Makinson (1985), who define a series of rules (*AGM postulates*) to deal with the modification of a belief base.

The postulates regulate three operations that could be performed on a belief base: expansion (adding a belief, no matter if it is inconsistent with the rest of the base), contraction (removing a belief), or revision (adding a belief but taking care to maintain the belief base consistent). The foundations approach is mainly elaborated as *truth*, or *reason*, *maintenance systems* (Doyle 1979), that is systems that keep track of the inferences used to generate new information, by also updating them when contraction or revision is needed. The output of this housework is that at any moment the system can tell whether a belief is held or not, on the basis of the belief's grounds[10]. The tension between the two approaches is testified by the defences each of the main authors has put forward of the preferred one (Gärdenfors 1992c; Doyle 1992), but it is interesting that the reason maintenance systems seem to be much used in application problems, where performance is an issue (see Taylor, Carletta, & Mellish (1996a) for example), while the AGM approach is preferred in more theoretical environments (see Friedman & Halpern (1999) for example).

An alternative view to belief revision is the one by Galliers (1992), who combines the foundational and the coherence approaches in a model where agents can autonomously decide not to revise their state when new information occurs. The decision is based on coherency matters, but also on the strength, or endorsement of a belief. The agent will therefore distinguish between first-hand and second-hand (communicated) beliefs, and in the latter case the agent will note the reliability of the source, which might also be generic, for example for widely known facts. All these assumptions are not given, but are subjective to the individual agent. The logical framework has been tested in an information retrieval scenario (Logan *et al.* 1994).

4.2.3 Representing Mutuality

The problem of representing mutuality, while avoiding the computational complexity of infinite recursion, has been dealt with by many researchers. The main difference among the models lays precisely in the way the recursive embedding has been modelled, that is whether they allow an explicit representation of any possible embedding level, or whether they adopt the view of the "schema", a label that identifies once and for all whether something is mutually believed[11].

One of the earliest computational models of mutual belief is due to Cohen (1978). The model assumes that, as soon as agreement is reached on a particular level of embedding, all subsequent levels are in fact equivalent, and interchangeable. In a two agent case, A and B, this means that when there is no

[10]There are several more or less elaborated versions of reason maintenance systems, which differently define the "grounds" on which a belief is based. One of these versions, the Assumption-based Truth Maintenance System, is detailed in Ch. 7.

[11]An alternative way to represent mutuality is by approximation, an approach which computationally is mainly based on the notion of *fixed point* (Fagin *et al.* 1996, Ch. 9).

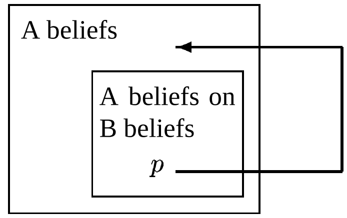


Figure 6: Mutual beliefs in Cohen (1978)

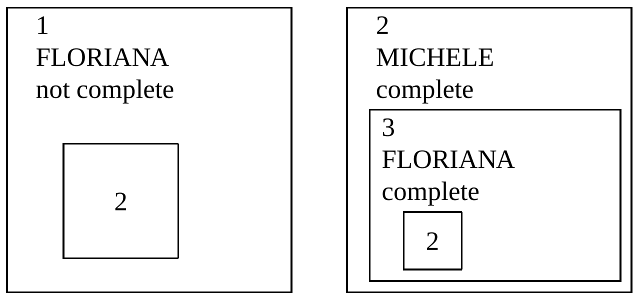


Figure 7: Mutual beliefs in Taylor & Whitehill (1981)

difference between what A believes of a particular proposition p and what A believes that the B believes that A believes about p, it can be assumed that all further expressions of the mutual belief that p can be collapsed. As shown in Fig. 6, this is implemented by creating a loop in one agent's set of belief, where the belief space which is one level deeper than the deepest containing non-mutual beliefs points to its parent space. As Traum (1994) observes, this makes it easier to determine whether a situation of mutual belief exists, by simply looking for loops.

A similar assumption is made by Taylor & Whitehill (1981), who also represent nested belief spaces as cyclic pointers between spaces considered identical. For instance, Fig 7 represents the situation of the mental state of the two agents Floriana and Michele, when Floriana, who has not completed the thesis, tells Michele that she has. Here, Floriana's beliefs on Michele's beliefs correspond exactly to Michele's belief, so a pointer (box numbered 2) is included in Floriana's model. On the other hand, Michele's beliefs about Floriana's beliefs are different from Floriana's actual beliefs, so a new belief box is created (numbered 3). Obviously, this model has sense only if it is made from the point of view of an external observer to the situation.

An important model of belief representation comes from the field of *user modelling*, not surprisingly: systems which have to interact with a human user are likely to store a more or less sophisticated model of the user, including beliefs and goals, and also beliefs of the user about the system itself. BGP-MS

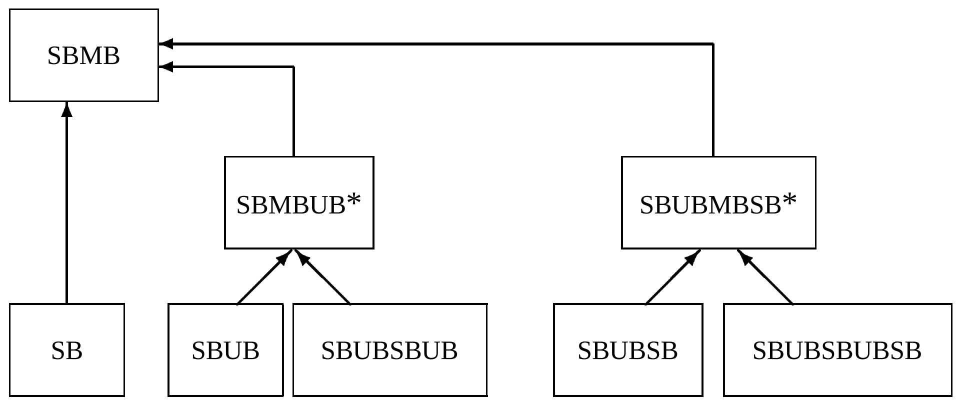


Figure 8: Mutual beliefs in BGP-MS

(Kobsa 1990) does this by organising knowledge in hierarchies of *partitions*, an approach also used by Cohen (1978). This organisation has the advantage of being economical: if two partitions share some content, this content is not reproduced twice, but is transferred to the common superordinate. Partitions are employed for any knowledge requirement of the system, and in particular also for representing the user's beliefs. New partitions are created whenever a new level of nested beliefs is required, for instance SBUS ("system believes that user believes") records "private" assumptions about what the user believes, and similarly for SBUBSB, SBUBSBUB etc, if the need arises. All these partitions are considered at the same level, while anytime the system can assume a mutual belief it creates a higher level partition. Three of these super-partitions have a special meaning. The usual notion of mutual belief is stored in the partition SBMB ("the system believes it is mutually believed"). But a weaker notion of mutual belief is also implemented by creating two subpartitions of SBMS consisting respectively of the "odd" and the "even" hierarchies of mutual beliefs (see Fig. 8). These are:

- SBMBUB, containing all the mutual beliefs of the system and the user about what the user believes on the domain, that is the even nesting levels (SBUB, SBUBSBUB, SBUBSBUBSBUB etc);
- SBMBSB, containing all the mutual beliefs of the system and the user about what the system believes on the domain, that is the odd nesting levels (SBUBSB, SBUBSBUBSB, SBUBSBUBSBUBSB etc).

This is a nice implementation of Lewis's concept of "schema" to immediately identify mutual beliefs, while allowing one to create any level of nesting required by the application.

The RHET system (Allen & Miller 1991) also has a hierarchical organisation of contexts, but with a different parent-child relationship (see Fig. 9). This hierarchy is transparent to the user, who operates in

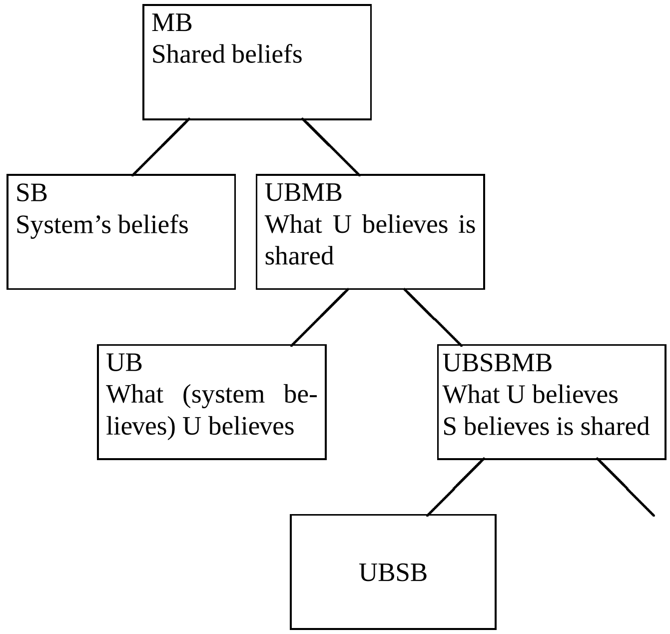


Figure 9: Mutual beliefs in RHET

terms of the usual modal operators to represent beliefs. RHET, conversely, implements the operators in terms of context shifts. Thus asserting "[BEL user P]" is implemented by adding [P] to context UB, or "User Believes". A major limitation in this representation is that shared beliefs can be asserted only in the root context, therefore it is not possible to represent shared knowledge about an individual's beliefs (for instance, it is not possible to assert that MBSB(P), but only that MB(P)).

Ballim & Wilks (1991) propose a model of agents' beliefs, ViewGen, that addresses the need of generating belief spaces of arbitrary depth. Once again, belief spaces are not given *a priori* but are generated whenever needed. A mechanism of ascription is provided to place each new belief in the appropriate space, and generating new space when needed. The various nesting level are represented as in Fig. 10. In the system's notation, beliefs of an agent are collected into boxes having the name of the agent in the middle of the bottom line. Agents can have beliefs about the domain of conversation, or about other agents, each contained in the appropriate box, where beliefs about a topic are collected together into boxes with the name of the topic on the top left corner. The types of nesting are diverse: in Fig. 10, for instance, the beliefs of the agent SYSTEM are divided into two sets: those about pneumonia and those about John. Among those about John, John's belief space is included (in the system's view), where in turn John's beliefs about pneumonia are represented. An interesting property is that there is no partition generation for beliefs that can be assumed as default. In the example, if John was a doctor, the "beliefs of John" partition would not be generated for beliefs like "John believes pneumonia is a disease"

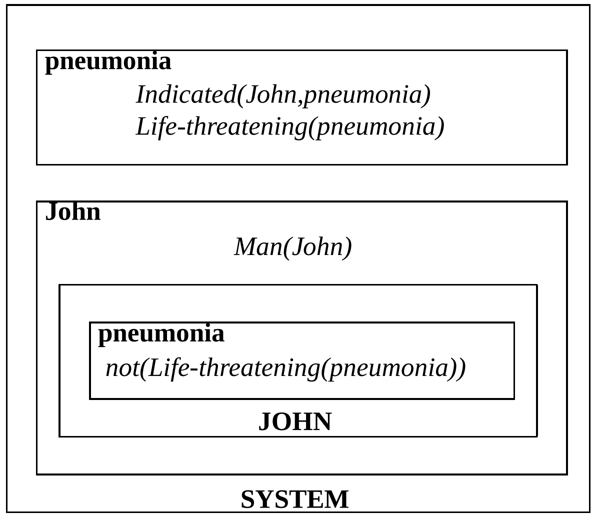


Figure 10: Mutual beliefs in ViewGen

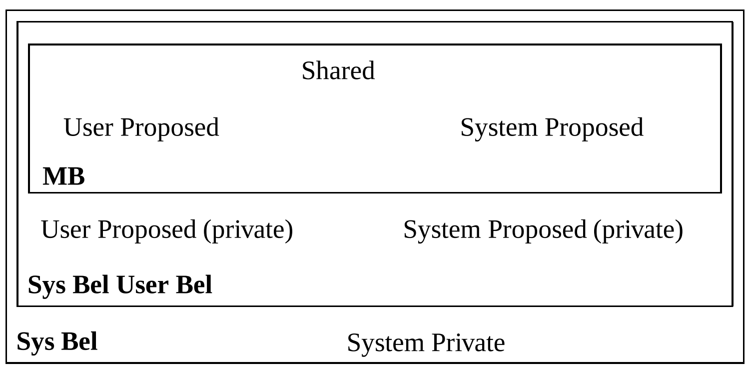


Figure 11: Mutual beliefs in TRAINS

(it would be among the beliefs *about* John, rather than the beliefs *of* John).

TRAINS (Traum 1994) has a similar representation of belief nestings (see Fig. 11). An interesting feature in this system is the separation between actual beliefs and *proposals*, which may be seen as belief which still need to be consolidated. By separating these two sets, insincere utterances may be dealt with. Also, the system distinguishes between just being aware of something and actually believing it. The belief and proposal modalities are implemented as RHET belief contexts (see Fig 9), where the root context represents general (assumed) mutual beliefs, the Shared context is root's child, and the mutually believed user and system's proposals inherit directly from Shared. The other belief spaces have their own context, inheriting from either user or system's proposals.

Taylor, Carletta, & Mellish (1996b), by concentrating on cooperative dialogues, note the "apparent irrelevance" of representing deeply nested beliefs: the assumption is that, if no deception is expected or

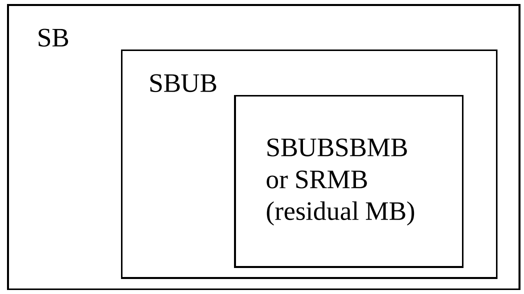


Figure 12: Mutual beliefs in Taylor, Carletta & Mellish (1996b)

allowed, only a few levels of nesting (three, in fact, they propose) is needed to completely model the dialogue. They report on a previous attempt (Taylor 1993) on adding a variable limit on nesting levels to a system for cooperative dialogues, JAM (Carletta 1992), noting that no difference in the performance of the system was obtained as a result. The authors' position is therefore that the representation of deeply nested beliefs is really needed only in well defined problems, like the interpretation of referring expressions, deceptive or "political" dialogue, and utterances explicitly mentioning other's points of view. This position was later confirmed by Barrington & Sonenberg (1997). The resulting proposal is for a system to be able to accept and reply on beliefs on any nesting level, but "allow deeper belief nestings to be conflated with shallower ones". The proposed belief "groups" are (see Fig. 12): the private beliefs of the agent (SB), the belief of the agent about the partner's beliefs (SBUB), and the *residual mutual beliefs* (SRMB), conflating all other belief nestings.

We agree with Reed (1997) that, for each example where a certain number of belief nestings is needed, one can always create a new one needing one more. We are however also convinced by Taylor, Carletta, & Mellish (1996b) and Garagnani (1997) that, for systems which do not need to show an unlimited reasoning ability, reducing the number of sets is a wise choice, both from the computation and from the clarity points of view. We will return to this issue in the next Section.

4.3 A Mental Model for a Rhetorical Arguer

We describe here a mental model for a rhetorical arguing agent. The model is meant to be as general as possible, as we discuss what we believe is important to include, rather than try and solve issues of representation and reasoning. We discuss the model by considering in turn the main issues we have seen so far: how to model the mental state, how to account for multiple attitudes, and how to represent mutuality.

4.3.1 Representing Mental Attitudes

We would like our agents to be able to reason about any mental attitude. We do not wish to state in advance the types of mental state an agent can be in, like for instance in the BDI framework, but we will simply write:

$$\beta_a(x)$$

to mean that agent a is in the propositional attitude of β towards the state of affairs x. In the terminology seen at the beginning of this Chapter, β is an attitude verb, and x is a *t-clause*.

We said we do not want to make any decision on the number of propositional attitudes to represent. We expect our agents to show informational, motivational, social and emotional attitudes as needed by the argumentation, and indeed we expect rhetorical argumentation to make use to a great extent of a variety of attitude nuances. In particular, a careful account of emotions ought to play a crucial role in rhetorical argumentation, as Aristotle pointed out in *Rhetorica*. However, we do not propose, in this work, a theory of mental attitude: how they come to be, how they change, how they influence each other. While this may appear disappointing, it is also true that the only reasonable way to approach this problem is by investigating each and any of the mental attitudes with as much effort as there has been in the epistemological/informational attitudes analysis, and this is definitely outside the scope of this work.

We propose a general framework to reason about mental attitudes, which is inspired by work on contextual reasoning (Benerecetti, Bouquet, & Ghidini 2000). The intuition, in contextual reasoning, is that agents are not very likely to use their complete mental state to solve a particular reasoning task at hand, but they are able to isolate a context that, while containing all necessary information, is more manageable. This notion, we find, implements very nicely one intuition that we have seen emerging throughout the review above, both in philosophy of mind and in computational research. The intuition, that is, that we may represent the mental state of an agent as a collection of "boxes", each with a specific meaning, and typically with its own rules of reasoning. Of particular interest, to us, is the notion of MultiContext systems (Giunchiglia & Bouquet 1998): this adds to the locality assumption that every context contains all that is needed to solve the task, the "compatibility" assumption that all contexts have to be related in some way in the agent's mind. The latter assumption addresses one of Fodor's requirements for a theory of mental attitudes, the already mentioned "Aristotle condition". The two assumptions are implemented by providing, in addition to the rules which are valid only within a context, and which might be of different nature, *bridge rules* that can be used to reason across contexts.

More formally, a context is a triple:

$$\langle L, \Omega, \Delta \rangle$$

where L is a language (e.g. first order logic), Ω is the set of axioms of the context and Δ is the set of inference rules that applies in the context.

A MultiContext system is a pair:

$$\langle C, B \rangle$$

where C is the set of all contexts, and B the set of bridge rules.

A bridge rule is in turn a rule of the type:

$$c_1 : \phi_1, c_2 : \phi_2, \ldots c_n : \phi_n \rightarrow c_{n+1} : \phi_{n+1}$$

where $c_i : \phi_i$ means that the formula ϕ_i is true in the context c_i (note that the notation does not imply an ordering on contexts).

We see the agent's mental state as a set of "attitude contexts", where different contexts are used to represent different mental attitudes. An attitude context contains, for the single agent, not only the objects, the *t-clauses* the agent happens to have as objects of that particular attitude, but also the reasoning rules that the agent uses to manipulate the propositional attitudes in the "box". We therefore give the following:

Definition 8 (Mental State) *A mental state for a rhetorical arguer $\mathcal{M}$ is a multicontext system:*

$$\mathcal{M} = \langle B, R \rangle$$

where B is a set of attitude contexts, and R is a set of bridge rules among the attitude contexts. An attitude context $\beta \in B$ is a triple:

$$\beta = \langle \lambda, \alpha, \varrho \rangle$$

where λ is a language, α a set of axioms in λ and ϱ a set of inference rules internal to β.

The definition above does not commit to the use of a possible worlds rather than a sentential approach to reasoning within or across attitude contexts. We would tend to adopt a sentential view for a possible implementation, as we believe that the balance of advantages and disadvantages is biased on the sentential side, but the choice is essentially arbitrary.

The definition allows us to postulate that different attitudes can have different reasoning styles. The 8 principles of Barwise & Perry (1983) mentioned earlier can be an example of rules internal to a context. In addition to internal rules, we also assume that the agent has a set of bridge rules, that allow ascribing new propositional attitudes of one sort from other attitudes of different sorts. Similar to what has been hypothesised for intentions and beliefs (Cohen & Levesque 1990), we can therefore imagine an agent having rules to decide, for instance, when a doubt can become a belief, and vice versa, or when fear leads to intention to act, and so on. Again, we do not want to make assumptions on what these bridge rules may look like, and we leave this issue to further works.

A decision that we did take, however, was to see both the attentional stance and the evaluative stance an agent can have towards a concept as something different from a mental attitude. We refer here to the attitudes of "having in focus that x", and of "valuing x". While they might be considered as contexts, we prefer to see them as meta reasoning abilities of the agent, as we see that agents "use" them differently. In fact, the attentional attitude ("having in focus that") is something which makes sense only while a conversation is active. It is therefore more volatile than the other propositional attitudes, that are expected to be a property of the agent, not of the conversation, and will last after the conversation has terminated. In fact, one can consider the focus as the subset of the mental state of the agent which is "active" at a particular time. Strictly speaking, it should be a multicontext containing subsets of the attitude contexts that are relevant at that moment. As to the evaluative attitude, we agree with the line of reasoning mentioned in Sect. 4.2.1 that this is best be seen as a "label" that can be associated to other mental attitudes.

This distinction has already been used in the definition of Rhetorical schema (Def. 3, page 33). In the definition, we describe the evaluation of an element as a basic notion of the rhetorical argument description. Such evaluations are distinguished from the references to other propositional attitudes in the agent's mental state, denoted again with the symbol β. In the definition, we also use a special constraint type, the "Relevance constraints", to express what the audience has to have "in focus" for the argument to be successful, thus separating this attitude from the others.

4.3.2 Representing Mutuality

We have discussed above the tension between computational systems implementing the "infinitely complete" version of mutuality, and those which adopt Lewis's point of view that there must be a "schema", a way to collapse the notion of mutuality from a certain level on, and just assume the agent has "reached" that level. We tend to agree, as we anticipated, with the latter view. The problem for us is then shifted to when we can assume mutuality has been reached. In other words, we need to establish what is the ideal number of nesting levels that are needed for rhetorical argumentation.

In a situation of cooperative dialogue, agents are only interested in communicating what they believe, and assume their opponent does the same. This is the simplest communicative situation, and we have seen in the previous Section that authors like Taylor, Carletta, & Mellish (1996b) or Garagnani (1997), or indeed Kobsa (1990) with their "special partitions", argue that three levels of belief nesting are sufficient to represent it, where the third level is really only needed to represents misconceptions.

If one needs to model deceptive communication (Cohen & Perrault 1979; Lee & Wilks 1997), or communication where it is needed to “plan for plan recognition” (Taylor, Carletta, & Mellish 1996b), that is to figure out what the other agent’s plan might be, deeper nesting levels are needed. It is however difficult to imagine real life situations where more than a few more levels of nesting need to be added, as humans would tend to lose track of the connections.

But what are the characteristics of our arguing agents? Can they be defined as cooperative?

We would not want to go as far as saying that some sort of deception is in a sense inherent in the definition of rhetorical argumentation. We have seen in Ch. 2 that it is precisely this misconception that has diminished the importance of the field. We have however to concede that the fact that a rhetorical argument is based on the audience’s beliefs, and not on what the arguer really thinks on the matter, needs some of the devices of deception. In particular, when the arguer makes a standpoint which is only a projection of the audience’s mind, like saying *Eating fruit is important because it helps slimming* to an audience who gives high value to being slim, we cannot exclude that the audience could ascribe this belief to the speaker, who on the other hand might have little interest in being slim. The ascription may well be “for the sake of conversation” (Stalnaker 1974), as the audience may not necessarily believe that the speaker *believes* that slimming is important, but it is needed to create the grounds on which subsequent arguments can be discussed and evaluated.

So what do we need our agent to reason about? Let us consider the attitude of believing. We definitely need an agent to reason about private beliefs. We need the agent to reason about beliefs of other agents. And we need mutual beliefs. However, mutual beliefs depends on the interlocutor: agent A can reason about what is mutually believed with agent B, which may be different from what A believes is mutually believed with agent C. We represent this by creating separate environments in agent A’s state of mind:

1. what A believes B believes is mutually believed
2. what A believes C believes is mutually believed
3. etc.

This, however, still does not capture the expressiveness we would like. We have always maintained that rhetorical argumentation is not about absolute facts, but about what agents perceive the reality is. It has therefore little sense to say that agents A and B have a mutual belief *that X*, as in this case X would be a given fact. What we really would like to represent is that agents A and B have a mutual belief *that someone believes X*, where “someone”, in the discussion between A and B, can be either of them. We see, in other words, the third belief nesting mentioned by both Taylor, Carletta, & Mellish (1996b) and Kobsa (1990) as in fact a double environment, which is labelled with one of the two interlocutors.

This allows representing situations in which an agent realises that what the other has perceived from the conversation is not exactly what is in fact believed, because of a misconception, or because of a statement of a "belief for the sake of conversation". An agent A saying "eating fruit helps slimming" may assume that its interlocutor will believe that "A believes slimming is important", while at the same time A might believe that this is not case. We then have:

1. A believes that "not X";
2. A believes that B believes that it is mutually believed that A believes that "X";

Symmetrically, agent A can ascribe a "belief for the sake of conversation" to agent B (including it in the mutual belief section) while not really ascribing the belief to B:

1. A believes that B believes that "not X";
2. A believes that B believes that it is mutually believed that B believes that "X".

Therefore the situation of *"mutual belief that X"*, as envisaged by Clark & Carlson (1982), corresponds to the set of beliefs (from agent A's point of view):

1. A believes that "X";
2. A believes that B believes that "X";
3. A believes that B believes that it is mutually believed that A believes that "X";
4. A believes that B believes that it is mutually believed that B believes that "X".

4.3.3 Mutuality across Attitudes

We assume that in principle mutuality can be expressed for any attitude (though we are prepared to accept that for some attitudes it will have less sense). For instance, a "mutual desire", as a paraphrase of the definition by Clark & Carlson (1982) happens when:

1. A desires that p.
2. B desires that p.
3. A desires that B desires that p.
4. B desires that A desires that p.

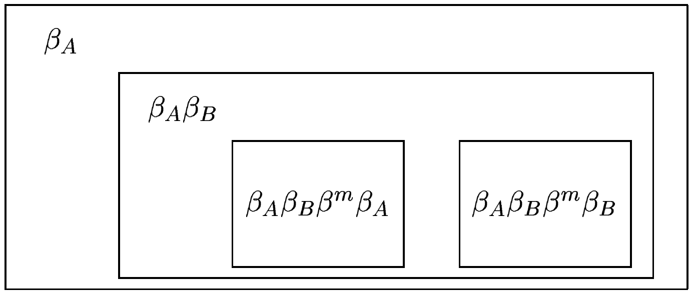


Figure 13: Model of Belief Nestings

5. A desires that B desires that A desires that p.

6. B desires that A desires that B desires that p.

et cetera ad infi nitum.

Furthermore, as we assume mutual attitudes are expressed towards agent's attitudes, and not mere facts, we want to represent that it is "mutually believed that A believe X", or that is it "mutually desired that A believes X", or that it is "mutually believed that A desires X", and so on.

We assume, in other words, a generic multi attitude scenario, where inner attitudes, and mutual attitudes, can be defined as needed. Figure 13 shows a graphical representation of nesting levels in a generic multi attitude scenario, where:

β_A is agent A's own mental state (A believes X, or wants Y or doubts that Z, etc.)

$\beta_A\beta_B$ are attitudes ascribed by A to agent B (A believes B believes X, or A believes B wants Y, or A wants B believes Z, etc).

$\beta_A\beta_B\beta^m\beta_A$ represents A's view on B's view on what mutually happens "as regard to A" (e.g. A believes B wants that it is mutually believed that A believes X, or A wants B believes that it is mutually wanted that A believes X, etc.).

$\beta_A\beta_B\beta^m\beta_B$ represents A's view on B's view on what mutually happens "as regard to B" (e.g. A believes B wants that it is mutually believed that B believes X, or A wants B believes that it is mutually wanted that B believes X, etc.).

Mutuality as Attitude Contexts

Going back to the problem of representing the mental model, we consider the levels of nesting as if they were different propositional attitudes (Thomason 1988; 2000). In other words, we assume that, for instance, "believing" and "believing that agent X believes" are two different attitude contexts, and we assign them different "boxes". This makes sense if it is noted that agents may use reasoning rules, when trying to put themselves in other agent's shoes, which may well be different than those used when reasoning on their own mental state. Moreover, we can assume these reasoning rules are dependent on the particular agent they reason upon, so "believing that agent X believes" may work differently than "believing that agent Y believes". As described above, bridge rules are used to reason across contexts, allowing, for example, to conclude something on what another agent believes on the basis on one's own beliefs.

In this hypothesis, the number of attitude contexts that there are in an agent's mental state depends on the number of agents to reason about, and the number of propositional attitudes to consider. For instance, if two attitudes only are considered (say "belief" and "want") in a two agent (A and B) scenario, then: An agent A has a context for each private mental attitude:

1. what A believes;
2. what A wants.

Then, agent A might be in one of the two attitudes towards any attitude of agent B:

3. what A believes B believes;
4. what A believes B wants;
5. what A wants B believes;
6. what A wants B wants.

Finally, agent A needs to store mutual believes and wants, with different combinations[12]:

7. what A believes B believes is mutually believed B believes;
8. what A believes B believes is mutually believed B wants;
9. what A believes B believes is mutually wanted B believes;
10. what A believes B believes is mutually wanted B wants;
11. what A believes B wants is mutually believed B believes;
12. what A believes B wants is mutually believed B wants;
13. what A believes B wants is mutually wanted B believes;
14. what A believes B wants is mutually wanted B wants;
15. what A wants B believes is mutually believed B believes;
16. what A wants B believes is mutually believed B wants;

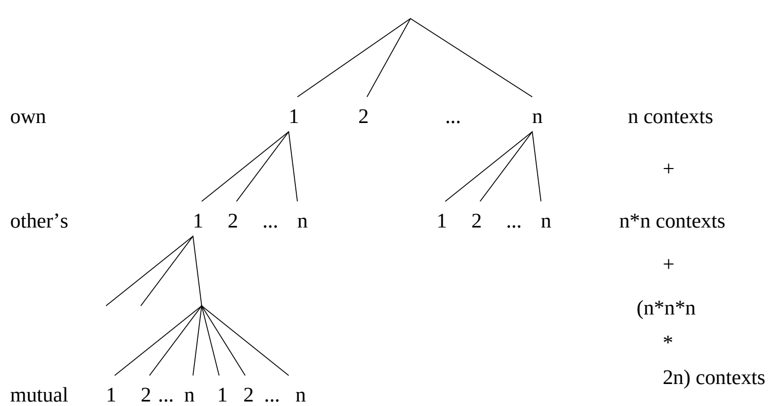
[12]Remember that mutual beliefs are not defined as such, but only as a reflection of the other agent's attitude.

17. what A wants B believes is mutually wanted B believes;
18. what A wants B believes is mutually wanted B wants;
19. what A wants B wants is mutually believed B believes;
20. what A wants B wants is mutually believed B wants;
21. what A wants B wants is mutually wanted B believes;
22. what A wants B wants is mutually wanted B wants.

and:

23. what A believes B believes is mutually believed A believes;
24. what A believes B believes is mutually believed A wants;
25. what A believes B believes is mutually wanted A believes;
25. what A believes B believes is mutually wanted A wants;
27. what A believes B wants is mutually believed A believes;
28. what A believes B wants is mutually believed A wants;
29. what A believes B wants is mutually wanted A believes;
30. what A believes B wants is mutually wanted A wants;
31. what A wants B believes is mutually believed A believes;
32. what A wants B believes is mutually believed A wants;
33. what A wants B believes is mutually wanted A believes;
34. what A wants B believes is mutually wanted A wants;
35. what A wants B wants is mutually believed A believes;
36. what A wants B wants is mutually believed A wants;
37. what A wants B wants is mutually wanted A believes;
38. what A wants B wants is mutually wanted A wants.

In general, for two agent scenarios, with n attitudes, the number of attitude contexts for one agent is:



That is:

$$n + n^2 + 2n^4$$

If k agents are to be considered, and we assume agents do not have a complete picture of what other agents think of each other mental state (that is we do not represent, say, that agent F believes that agent G believes that agent H believes...[13]), the number of attitude contexts for one agent is:

$$n + k(n^2 + 2n^4)$$

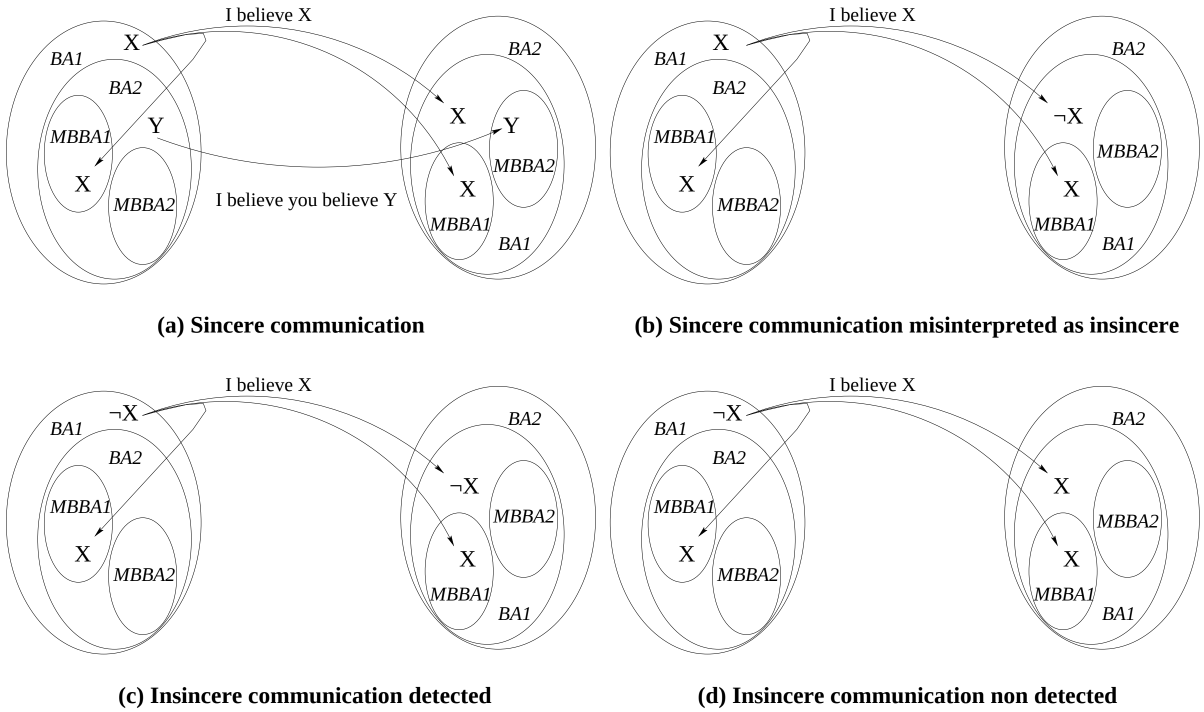
4.3.4 Use of Contexts in Dialogue

We can now detail better, from the point of view of the mental state, what happens in the dialogue game when a move is made. We mentioned in Sect. 3.3 that, after a dialogue move, something is entered into the agent's *commitment store*. We see the commitment store as the multicontext mental state defined above.

We assume that an agent, when performing a dialogue move, is confident in the "success" of the move, and therefore stores, in the self related mutual attitude context, that the move has taken place. The attitude invoked will depend on what the agent has expressed: whether a belief, or a wish etc. A similar assumption is made by the agent when listening to a move performed by another agent, that is that the other agent will store what has been communicated in the same way. Therefore, whatever is "said" is by default included in the appropriate mutuality contexts of the dialoguing agents. On the other hand, what the speaking agent's real point of view is, or what the listening agent really ascribes to the speaker after a communication, depends on the agent's relevant bridge rule and will be stored in the "private" sections of the mental state.

Figure 14 represents how the various parts of the model of two agents, A1 and A2, are updated after different sorts of communication, showing that the four sets proposed are necessary and sufficient to represent the different situations. The figure only considers one attitude (belief) for the sake of clarity, but this can be extended to multi-attitude models. The piece of communication is shown as a label on the edge between the two agents' models. This can be either sincere, or insincere, not in the sense of being maliciously deceptive, but in the sense that it is not faithful to the actual agent's belief. When a communication is sincere, and is perceived as sincere, all the sections of the belief model of both agents contain the same piece of information (case a). In cases in which either a misinterpretation, or a belief

[13] We are not aware of any study of multiagent dialogues which takes this kind of situation into account, but we suspect that conclusions on the ability of humans to keep track of chains of beliefs can be made which are similar to those on infinitely nested mutual beliefs.



(a) Sincere communication

(b) Sincere communication misinterpreted as insincere

(c) Insincere communication detected

(d) Insincere communication non detected

Figure 14: Communication types

for the sake of conversation (or indeed a lie) occur, differences exist between the mutual and the private sections of both agents, where the mutual sections mirror what has been communicated, and the private ones express the beliefs which are actually held.

It should be noticed that case (a) is the only situation in which the communication concerns a second level belief (that is a belief about the other agent's beliefs). We assume in fact that when agents talk about each other beliefs, only a straight, truthful communication can take place (one agent cannot say "I believe you believe X" if this is not the case). Similarly, we also note that we cannot represent more sophisticated styles of communication, such as irony, in which the speaker says false statements with the assumption that the hearer will realise that. In both cases, in order to be able to represent this properly we would need extra levels of attitude nesting (Lee & Wilks 1997).

4.4 Conclusions

We have explored, in this Chapter, the problem of representing the mental state of a rhetorical arguer, with the intent of better specifying both the content of the admissibility constraints of the rhetorical schemata we proposed in Ch. 2, and the notion of commitment store we have introduced in Ch. 3. In the spirit of all that has been done so far, the exploration has produced an ideal framework, rather than a specific, implementation oriented design for the agent's mental state. By drawing upon fundamental notions of philosophy of mind, and by analysing how these notions have been implemented, we have proposed a clear picture of what is needed to represent an agent able to argue rhetorically. From this, the reader interested in actually implementing the agent, can benefit from off the shelf formal models, whether from logic, probability or constraint reasoning fields, both to reason within the contexts and for implementing bridge rules. We will not choose a preferred one, and we leave the investigation on whether one can be better than another outside the scope of this work.

Chapter 5

An Application Scenario: Health Promotion Dialogues

> The trees went forth on a time to anoint a king over them; and they said unto the olive tree, Reign thou over us. But the olive tree said unto them, Should I leave my fatness, wherewith by me they honour God and man, and go to be promoted over the trees?
>
> - Judges, *9,9:10*

It is often useful, while devising a general model, to detach for a while from the theoretical, sometimes abstract constructions, and try and focus on a domain of application, in order to think of specific examples to test the model upon. This is not meant to limit the generality of the theory, but, by applying it to a specific problem, to spell it out more clearly, and hopefully obtain new insights to improve it. If the ultimate aim is the one of building an artificial agent able to argue rhetorically, we need to make sure that the grounds on which we base the agent's characterisation are sound and robust.

The remainder of this thesis, starting from this Chapter, accomplishes this task, and concentrates on the field of Health Promotion as one possible scenario of application. The choice was driven by two factors, as will be explained in the first two Sections of the Chapter. Firstly, the scenario presents interesting manifestations of rhetorical argumentation. Secondly, health education/promotion has historically been a fertile application field for research in discourse processing, so we find more terms of comparison in this area.

The core of this Chapter concerns the analysis of health promotion dialogues, which we perform by means of a new annotation model that we introduce, based on the theoretical work described so far in the thesis.

5.1 Health Promotion as an Argumentative Activity

Health Promotion can be described as "the process of enabling people to increase control over, and to improve, their health" (World Health Organization 1986). It is mainly concerned with developing an individual's personal skills, through both providing information, and advocating healthier behaviours. A crucial issue for the health adviser is, therefore, understanding the mechanisms of behaviour change. In particular, the adviser must consider that several barriers can prevent the health promotion message from getting accross, and being accepted. These principles can be seen to apply clearly in promotion of healthy nutrition. Despite the generally accepted importance of a balanced diet, and the fact that several health organisations have pointed out that healthy nutrition can have a crucial role in decreasing the incidence of illnesses such as cardiovascular disease or cancer (Department of Health 1992; US Department of Health and Human Services 1991), changing people's diet has turned out to be a hard challenge. A number of studies tried to analyse this phenomenon. Among others, Fries & Croyle (1993) showed that stereotypes influence people's reaction to nutrition education. For example, a sample of college students interviewed described people eating low-fat food as *picky*, whereas people with a high fat diet were pictured as *easygoing* and *more likely to attend parties*. As a consequence, students who held a negative image of people eating a low fat diet, were more sceptical when presented with information about its benefits. Similar stereotypes can be found in a survey by Sadalla & Burroughs (1981). Another study by Ziebland *et al.* (1998) identified some of the barriers which prevent people from changing their diet, covering a range of factors both internal, such as enjoyment or lack of willpower, and external, such as lack of money. It turned out that internal barriers (such as *I enjoy eating foods that are bad for me*) were the most frequently mentioned by the interviewed. Similar findings appear frequently in other studies (Brinberg & Durand 1983; Brug, Lechner, & de Vries 1995; Fagerli & Wandel 1999; Potts & Wardle 1998; Wardle *et al.* 1997).

Fries & Croyle (1993) argue that this kind of situation is similar to that faced in advertising: the health promotor should therefore be armed with the same type of argumentative skills. This turns out to be very interesting for our purposes. We have seen that people are likely to have strong beliefs and preconceptions on what a healthy diet is. These are not necessarily wrong, and people are confident in what they believe, therefore the relationship that is established between the nutrition adviser and the advisee is more balanced than in typical educational scenarios. This means that, especially in a situation of "health behaviour", that is when people are not forced by illnesses to change nutrition habits, the advisee's beliefs are very likely to foster an argumentative attitude, standing for what is believed, and having to be persuaded to change. Moreover, as seen in the literature, this scenario is more likely to involve discussions on extra-rational factors, like moods, emotions, and, especially, values, something we are interested in observing.

We, therefore, focus our attention on argumentative dialogues in healthy nutrition promotion. We will try to look for "patterns" in these dialogues that confirm the theory on rhetorical argumentation discussed so far. But before that, we need to have a deeper understanding of the goals of the dialoguing partners in this scenario. That is, we need to look into the factors that influence people's behaviour change, and the way in which these can be exploited for argumentation.

5.1.1 Models of Change

Two well established models, in health education and promotion research, that explain the mechanisms of behaviour change are the Health Belief Model and the Stages of Change Model.

The **Health Belief Model** (HBM) (Becker 1974) assumes that there are "regions" in people's life, which are associated with a value, or valence, either positive, or negative, or neutral. Negative valued regions cause a person to take actions in order to move away from that region, while obviously avoiding ending up in a worse one. Diseases are the most immediately perceived negative regions, and actions (therapies) will be undertaken to move away from them, provided that their side effects are not worse than the disease itself. The assumption, however, is that motivation is a necessary condition to action, where motivation is both the state of being *ready* to take specific action and the state of believing that such action will be beneficial. This happens if people believe:

- that they are personally *susceptible* to the negative valued region, rather than just admitting that it could "statistically" happen;

- that its occurrence would have at least moderate *severity* on some components of their life, although the degree of seriousness is a matter of subjective intepretation;

- that taking a particular action would in fact reduce their susceptibility to it or, if it occurs, would reduce its severity;

- that taking a particular action would not cause them overcoming important (for them) *barriers* (e.g. cost, convenience, pain, embarrassment and so forth).

The HBM also argues that, all these factors being present, a *trigger* to act is still necessary, like a mass media campaign, advice from others, illness of a family member or a friend etc. The way this trigger works depends on the scenario to be confronted. For instance, in a situation of good health the behaviour to induce is a preventive one, so triggers emphasising the perceived severity of not adopting the new behaviour are less efficacious than triggers emphasising the perceived susceptibility to a bad consequence. This does not apply when people already feel or know there is something wrong with them.

While the HBM is concerned with the beliefs on which a person's behaviour is based, another important model, the **Stages of Change Model** (SCM) (Prochaska & Clemente 1992) is concerned about how change can actually take place. The model assumes that people progress through very distinct stages of change on their way to improve health:

1. *precontemplation*: people at this stage see no problem with their behavior and have no intention of changing it. They mainly lack information, in the sense that they have not been presented yet with any convincing reason to change their behaviour. Often people are not very open to receiving advice.

2. *contemplation*: in this stage, people come to understand their problem, its causes, and start to think about taking action to solve it, but have no immediate plans. This is a delicate stage, as there is always the risk to miss the opportunity, and go back to precontemplation, because of laziness or old influences.

3. *preparation*: people are planning to take an action, and are putting together a plan, but have not taken any step yet. This is a sort of transition stage between the decision to act and the action itself. Often one of the causes of going back to a previous stage is that the plan is too ambitious, and the life style change planned is too drastic.

4. *action*: people are actually in the process of actively making behaviour changes. The concern here is to pay attention to negative emotions: anger, depression, anxiety, apathy, insecurity, etc., in order to prevent relapse.

5. *maintenance*: health behaviour continued on a regular basis. The state is more stable than the action one, but there is always the possibility of relapse.

6. *(termination)* at this stage, the former problem no longer presents any temptation or threat. Many people never reach this stage.

The SCM has been successfully applied in various campaigns, and researchers have elaborated the model in order to provide concrete strategies to recognise and deal with each stage of change. Table 12, summarised from Barrie (1991) shows an example of these, while Table 13 shows a list of activities to be performed for encouraging each stage movement, proposed by Grueninger (1995).

	Signs	Advice aims	Methods
PREC	see benefits rather than costs; may see other things as problematic; use fewer change strategies; may avoid help and info	provide info on related problems; give opportunities to identify problems; emphasise harm reduction and health maintainance	information giving; costs-benefits; identify harm reduction methods
CONT	consciousness arising; self re-evaluation; looking for information and help;	raise consciousness about problem; increase commitment towards changing	info on health and social consequence; identify problems and facilitate expression of concern; advice on the nature of help
ACTION	self re-evaluation; self liberation ('I can do it"); helping relationship; rewards; alternative activities	maintain and enhance commitment; facilitate behaviour change; teach strategies to prevent and manage breakdown	identify problems; facilitate expression of concern; enhance the perception of need to change; teach problem solving techniques; identify benefit of change; identify harm free activities
MAIN	remove things from environment which remind old habit; engage in new activities to relax	maintain behaviour; prevent and manage relapse; encourage general life style change;	explore reasons for change and reinforce; consider benefit; encourage vigilance; encourage alternative activities

Table 12: Characteristics of SCM states

Furthermore, the SCM and the HBM have been successfully combined in many health promotion interventions, as constructs of the HBM are very closely associated with, and can help identify, the movements through the stages in the SCM (Werch & Clemente 1994). The combination of the two models has proved to be most effective in healthy nutrition promotion in particular (Roe *et al.* 1997). For example, Campbell *et al.* (1994) presents a study whereby dietary advice was produced on the basis of both the SCM and the HBM. Messages were created by nutritionists and health educators, and fragments of these texts were selected by a computer system, on the basis of the addressee's provided characteristics, and merged to form personalised newsletters. The study showed that participants receiving tailored messages were twice as likely to remember them, and fat intakes were significantly reduced, when compared with those who received messages identical in shape but not tailored. The authors conclude that health advice must effectively integrate behavioural and psychosocial factors. The paper also provides an interesting example of how a computer system can help in nutrition-related behaviour change programs. The next Section will provide other examples, though seen from the computer science perspectives.

5.2 Computation and Health Promotion

Researchers in health education have recognised that computer technology could offer powerful solutions to many important issues, provided that the systems are not merely didactic, and that they tailor

- From Precontemplation to Contemplation
  - assess prior knowledge
  - clarify understanding
  - transfer up-to-date information
- From Contemplation to Preparation
  - clarify feelings, attitudes, believes, motives and potential intention
  - assess and build instrumental skills
- From Preparation to Action
  - help to choose among options
  - anticipate problems, develop individual solutions
- From Action to Maintenance
  - check mutual understanding of agreed plan
  - assure continuity of collaboration

Table 13: Activities in the SCM movements

information to the individual (Skinner *et al.* 1993). Researchers in discourse processing and generation, on their part, have historically shown interest in applying their insights to health education and promotion. Notably, in the comprehensive collection of references to natural language generation projects put together by Bateman & Zock (2002), 19 systems were in one way or another related to health education issues, making the medical the most popular domain of application.

Migraine, a patient education system for migraine patients, is perhaps one of the most documented AI projects in this respect, with several publications on various aspects of the project (Buchanan *et al.* 1995). It aimed at generating tailored handouts to be given to patients. The intuition was that one should avoid presenting patients with large amount of data, most of which not interesting for them, as happens with standard leaflets. Ethnographic techniques (Forsythe 1995) were used to observe physicians providing explanations to patients in clinic settings, and to conduct interviews with migraine sufferers and physicians. A patient model was used to tailor information to the individual patient, which contained symptoms, prescribed drugs, etc. A natural language generation component produced an interactive information sheet, which was shown to the patient in the form of a hypertext document. The user/patient could ask follow up questions by clicking on portions of text and selecting one from a menu of available questions. The system focused on informational issues, so the model of the patients only considered their medical state and their factual knowledge, on the basis of the preliminary interview. No account is given of persuasive techniques, or techniques to reduce anxiety, in the text production.

Similar aims are shared by the PIGLIT project (Binsted, Cawsey, & Jones 1995; Cawsey, Binsted, & Jones 1995), which too aimed at building a computerised program to present cancer patients with tailored information in a hypertext fashion. A more practical approach was preferred to sophisticated AI techniques: the text generation component was more domain-oriented, strictly based on the medical record of the patient. The explanations started from the patient record, so that only information relevant to the user was presented first. By means of hypertext links, the patient could obtain more general information on specific problems or therapies. A knowledge base was used for general information, which was an "is-a" hierarchy of objects, based on a modification of the Read Code hierarchy of medical term classification (Read 1993). The system was once again mostly an information tool, therefore it was mainly based on factual knowledge. Evaluation studies with a sample of patients (Cawsey, Jones, & Pearson 2000) showed that, albeit limited, the use of personalisation made the users feel more comfortable in using the system, and perceive that the information provided was just related to them. While the user's satisfaction seemed affected by the system, the experiment however showed that using the system did not significantly improve the level of users' anxiety.

OPADE (De Carolis *et al.* 1996) was a European project aimed at developing an intelligent, multilingual, computerised system for drug prescription, adapted to different European countries. A byproduct of the system was a text generator component, with the aim of explaining the prescription by retrieving information from a set of heterogeneous databases: a medical record, a drug database and a prescription record. There was no direct interaction with the user: the explanation produced was printed out as a leaflet accompanying the prescription. The text generation process was informed by the results of a preliminary study, which collected a corpus of explanations from various doctors to fictitious patients. As opposed to Migraine and PIGLIT, texts in OPADE were not only concerned with informational goals, but were mainly aimed at ensuring that the prescription was properly followed, by also persuading the addressee that the therapy was efficacious. Also, texts were adapted not only to the recipient (the patient) but also to the "sender" of the message, that is the doctor who made the prescription, who was in fact the only user directly interacting with the system. Two user models represented the main attitudes of the two users, and the double user model meant that two, possibly conflicting, points of view about how the explanation should be produced needed to be reconciled (Berry *et al.* 1997; de Rosis *et al.* 1995). The texts produced have been evaluated both analytically, by comparison with the training set of texts, and empirically, in experiments with a sample of about 500 users (Berry, Michas, & de Rosis 1998). The factors evaluated were both user's recollection of the texts, and the "perceived likelihood" of the patient taking the medication, as a function of the negative (e.g. side effects) and positive (e.g. drug efficaciousness) information in the texts, as well as the order in which this information was presented. The findings have given rise to a reconsideration of the text generation policies (de Rosis & Grasso 2000), but no formal account was given of how the argumentative nature of the original texts could be rendered automatically.

A more recent project, STOP (Reiter *et al.* 1997; Reiter, Robertson, & Osman 1999) produced tailored letters to encourage people to stop smoking. This project is most interesting, from our point of view, as it recognised that in settings like health promotion the mere provision of information is not the main task. The project explicitly applied the principles of the Stages of Change Model for deciding what to include in the letters, which were produced by means of a standard natural language generation system (Reiter & Dale 2000). Notably, STOP has been thoroughly evaluated with a clinical trial, in which more than 2500 smokers participated. The evaluation however produced the counterintuitive result that tailored messages were not significantly effective, in terms of smoke cessation (Reiter, Robertson, & Osman 2003). This is in contrast with other studies, like the aforementioned one by Campbell *et al.*. The authors attempt various explanations for this negative result, but it did not seem to occur to them that one of the reasons might have been the lack of recognition that this was likely to be an argumentative scenario, and therefore the lack of impact might be due to the lack of an explicit argumentative ground for the model.

The possibility of combining the Stages of Change Model with argumentative techniques in the domain of health communication was also suggested by Marcu (1996b), even though only at a conjectural level. The paper is also sceptical at the possibility that "conceptual structures that are proposed by research on argumentation" could be adequately used to build discourse generation systems, mostly because they deal with factors, like "cognitive stages", that could not be "easily accomodated" in computational systems. We try to address this criticism with the theoretical model proposed in this thesis. With a view to the production of such a system, we start, in the tradition of the projects reviewed above, with getting more insight on how the system should behave, by analysing how humans behave in similar circumstances.

5.3 Analysis of Health Promotion Dialogues

The main task of this Chapter is the one of analysing a set of dialogues in the nutrition promotion domain, with the aim of capturing their fundamental peculiarities. A systematic analysis of dialogues, or in general human produced texts, is useful when creating models, whether or not computational, both because it produces training material for developing the model, and because it provides benchmarks for the evaluation of the model itself. In the analysis, we hope to find confirmation to our hypotheses that (1) the theoretical models of behaviour change are in fact used in naturally occuring health promotion dialogues, and (2) that these dialogues present a component of rhetorical argumentation, which can be described in terms of our characterisation of the New Rhetoric. We will start by introducing issues in conversation analysis, then we will propose our analytical framework, also describing the corpus of natural occurring dialogues which we test our framework against.

5.3.1 Discourse Analysis

The analysis of discourse is aimed at capturing those characteristics which are beyond issues of grammatical soundness. The discourse analyst looks at how the speaker, in case of a monologue, or the partners, in case of a dialogue, develop their ideas in the flow of various sentences. The aim is to produce models of the structure of the entire monologue or conversation. Coulthard (1985) provides a valuable introduction to the analysis of discourse, from the point of view of various disciplines. Particularly interesting is a survey on how various theorists have hypothesised a conversation structure, from simple bipartite models distinguishing *acts* and *events*, to elaborate discourse structures accounting for various ways of combining moves. The most sophisticated model, among the linguistic ones, is perhaps the one by Sinclair & Coulthard (1975), which comprises five levels:

1. *Act*: the simplest units of discourse, like “ask”, “accept”, “reply” etc.

2. *Move*: one complete contribution of a speaker to the conversation, typically comprising several acts.

3. *Exchange*: a sequence of moves related to the same activity in the conversation. Typically, one exchange will comprise three moves, one that initiate the activity, a response one, and a follow up one. For example, an information seeking exchange will typically have a question that initiates the activity, the reply, and an acknowledgement of some kind.

4. *Transaction*: a sequence of exchanges related to the same topic. They typically begin with a “boundary” exchange, that marks the start of a new topic, starting with a “marker” such as “well”, “OK”, etc.

5. *Interaction*: the complete conversation, made up of one or more transactions.

Models of discourse are typically accompanied by annotation schemes, that help the analyst to label naturally occurring discourse on the basis of the model. The annotation of a dialogue according to a code scheme is aimed at reconstructing the structure of the dialogue, and the participants' goals. What is interested in the dialogue depends on the researchers' interest: for instance Sacks, Schegloff, & Jefferson (1974) concentrate on the sole issue of turn taking in dialogues.

Carletta *et al.* (1997) describe an annotation scheme based on the three middle levels of the dialogue structure by Sinclair & Coulthard mentioned above. The coding scheme has been specifically designed for dialogues in a given domain, that is task oriented conversations aimed at finding a route on a map. The authors name *conversational moves* the basic units of the dialogue structure, like Instruct, Check, Query, and so on. Sequences of moves give raise to *conversational games*, which are played to fulfil a specific goal. Game are named according to the move that initiates them, so there is an Instruct, Check, Query game, and so on. As an extension to the parallel definition of "exchanges", games can be embedded, for instance to represent clarification subdialogues. Finally, *transactions* are sequences of games which satisfy a given, higher level purpose. In the domain, these correspond to sub-dialogues aimed at agreeing on a specific segment of the map.

Mann (2002c) provides, conversely, an attempt to code dialogues in any domain, with his *Dialogue MacroGame Theory* (DMT), a substantially revised extension to his own notion of Dialogue Game that we have discussed in Ch. 3 (page 51). As in the original work, the challenge is to identify and characterise the notion of "coherence", as an "integrity of intentions" (Mann 2002b) that can be ascribed to the dialogue participants[1]. The annotation captures these intentions by explicitly expressing, for each "game" or "macro-game" the goals of the initiator, of the respondent, and a joint goal (see Table 6, page 53). The annotation scheme is based on the metaphor of negotiation, whereby games are initially *bid* for by the initiator. The bid can be accepted or rejected by the respondent. After the game has been pursued, either participant can *bid termination* of the game, where the bid can again be accepted or rejected. All these bidding activities are labelled as "game acts" by the theory, and are used to frame the exchanges related to the actual game. DMT's notion of game emphasises its collaborativity: every act that cannot be thought of as collaborative, like "tell", is labelled as "unilateral", and placed outside games.

The works of both Mann and Carletta *et al.* try to capture, in different ways, the goals of the speakers, in a more thorough way than the original model of Sinclair and Coulthard. However, they still do not account for the *ways* in which these goals are pursued by the speakers, in a way which is perhaps beyond the scope of classical discourse analysis, and is rather in the realm of the analysis of critical thinking (Groarke, Tindale, & Fisher 1997). A reconciliation of dialogue analysis with RST, as for instance in the work by Stent & Allen (2000), or in the proposal of Core & Allen (1997) to integrate RST-like relations in a taxonomy of "communicative functions", could be a possible avenue for capturing the argumentative

[1]This is also the case in the author's already mentioned Rhetorical Structure Theory, see page 18.

aspect in the analysis of discourse. However, if the "argument schemas", as named by the authors, are in fact merely the basic RST constructs, one has again to face the problems that RST has in representing argumentative discourse (see Sect. 2.3.1). Another solution could be to take a more global view, as mentioned by Teufel, Carletta, & Moens (1999)[2], and represent the whole text as one argument. This provides an accurate description of high level regularities of argumentative texts, especially if specific genres are considered (Teufel, Carletta, & Moens are inspired by Swales (1990)), but it loses specificity at the low, sentence level, that theories like RST can provide.

We seek to combine the typical discourse analysis aim of marking fine grained conversation events, and the rhetorical analysis aim of marking how the speakers' goals are pursued in the conversation. The annotation model we propose incorporates the theoretical models we have introduced so far in the thesis, and is presented in the next Section.

5.3.2 A Three-Layered Coding Scheme for the Analysis of the Corpus

We have developed an annotation scheme that captures three aspects of the dialogues to analyse: the discourse analysis aspect of how the dialogue is structured; the rhetorical aspect of how the participants develop their arguments; and the strategic aspect of how the participants' goals guide the interaction.

Terminologically, we call the basic unit of discourse a *move*. This is in contrast with the terminology in Sinclair & Coulthard, or in the Dialogue Macrogame Theory of Mann, although it is in line with the one used by Carletta *et al.* Our choice is merely a continuity one, by analogy with the literature on dialogue games that we mentioned in Chapter 3, which assumes that dialogue moves are in fact the basic units of discourse. Consequently, we call one complete contribution of a speaker a *turn* in the conversation, where a turn can comprise one or more moves.

We propose a three-layered structure to represent dialogues, where each dialogue move can be seen from three perspectives:

1. a strategic level, looking at the ultimate goals of the dialogue for each participant, that is the reason why the dialogue has been initiated;

2. the rhetorical level, looking at the argumentative goals/strategies a portion of the dialogue shows;

3. the dialogue game level, looking at the dialogue moves performed by the participants.

[2]The work proposes an annotation theory for research papers, whose argumentative nature is captured by marking sections like the description of background research, or the comparison with other works. The authors address, therefore, a monological situation, rather than a dialogical one.

We elaborate on the characteristics of the three levels below, while an example of their application will follow.

Strategic Moves

Strategic, or meta-dialogue moves identify the dialogue's higher order goals. Typically there will be one single goal per dialogue, although this is not prescribing. The argumentative dialogue will be therefore instrumental to the satisfaction of this goal. The strategy is likely to be peculiar of the particular domain of conversation. In the health promotion scenario we have chosen to analyse, a typical high order goal will be for the adviser to make the advisee move one stage further in the SCM sequence, for instance from precontemplation to contemplation. In other scenarios, such as in a negotiation one, the strategic goal could be to induce the opponent to perform some action, or for the speaker to obtain some goods. In the map task dialogues of Carletta *et al.*, the strategic goal is to come up with a common map.

We will identify two main activities for each agent in the dialogue with respect to this main goal: a *preparatory* activity, in which an agent wants to "feel the ground" to understand what is the opponent's stance with respect to the goal, and the proper *goal satisfaction* activity, when the goal is actively attempted. The two activities give raise to a response on the opposite partner, who will perform a *declarative* activity when the position with respect to the main goal is clarified, and a *responsive* activity, in which the opponent may either allow for, or resist, the goal's fulfilment. In a general situation, the dialogue is not necessarily asymmetric: the two participants may both have a strategic goal to achieve, and strategies to pursue it. Therefore, both participants can in turn be in any of the two roles ("attack" and "defend") and both can play any of the four moves, according to the situation.

We want to explicitly mark the following strategic moves:

- From the "attacking" agent's point of view:

  1. *exploratory moves*: the portions of the dialogue used to establish he opponent's position with respect to the goal at hand;

  2. *action moves*: the portions of the dialogue in which the agent applies one of the strategies to pursue the goal;

- From the "defending" agent's point of view:

  3. *declarative moves*: the portions of the dialogue in which the agent gives information on its position; this could happen either voluntarily, or unconsciously;

4. *responsive moves*: the portions of the dialogue in which the agent either facilitates or impedes the other agent's goal.

Therefore, the *exploratory* and the *declarative* are mainly information exchange moves, while the *action* and the *responsive* are moves directly related to the goal satisfaction activity.

The terminology should not suggest that we are only interested in competitive activities: in an information seeking scenario, the preparatory move may involve asking for the level of expertise of the opponent, who may reveal it, and the action move will involve just asking for information, which can be provided in a responsive move.

We have said that we assume one strategic goal per dialogue. However we also allow for this goal to be structured, or decomposed into subgoals. All the subgoals, however, must be at the same level of abstraction with respect to the dialogue structure, and no one will refer to the rhetorical or dialogue level. For instance, in information seeking dialogues one might have several informative needs on the same subject. Or in a negotiation dialogue one might need to obtain some goods and at same time keep the expenditure in budget. The fact that these goals will be realised by means of questions, or persuasion acts, is not considered at this level.

In our health promotion scenario, we hypothesise that the strategic goals correspond to specific strategies associated with the health behavioural models, and in particular the SCM. We will mark therefore:

- From the adviser agent's point of view:

  1. *exploratory moves*: the portions of dialogue used to establish in which stage of change the advisee is;

  2. *action moves*: the portions of the dialogue in which the adviser applies one of the strategies to let the advisee move one stage further; this goal might have specific subgoals according to the strategies suggested by literature on the SCM (like those shown in Table 12);

- From the advisee agent's point of view:

  3. *declarative moves*: the portions of the dialogue in which the advisee gives away clues for the adviser to establish the stage of change;

  4. *responsive moves*: the portions of the dialogue in which the advisee, again according to the literature's suggestions, show readiness or resistance to progressing to the next stage.

For each of the possible subgoals of the main strategic goals, we assume a similar pattern of exploration followed by action. For instance, when trying to move the advisee to the action stage with respect to healthy eating, the nutrition adviser could propose an option, like cooking at home rather than buying pre-packed meals. To pursue this subgoal, the adviser may start asking what the advisee's stance is with respect to cooking, before providing easy recipes to try out.

Rhetorical Moves

By rhetorical moves we mean the moves which are specifically used for the argumentation. These can be either the application of a rhetorical schema, or moves in which a relation among values is expressed (for example, "health is more valuable than social life"). The former are based on the definition of the rhetorical schema given in Chapter 2. The definition is given in terms of 6 components (Definition 3, page 33):

1. the name of the schema,
2. the claim the schema supports,
3. the ontological constraints the schema is based on, that is which relations should exist in the domain ontology among the concepts used in the schema for the schema to be applicable;
4. the acceptability constraints, that is which beliefs the audience has to possess for the schema to be effective;
5. the relevance constraints, that is which beliefs the audience has to have "in focus", in other words which beliefs are relevant to the argument;
6. the sufficiency constraints, that is which premises should exist in order for the argument to be "less attackable".

We want to mark schema applications occurring in the dialogue, and in doing so we will explicitly mark the move or set of moves in which the claim is put forward, and those where the various constraints are specifically verified (either by means of a question/answer pair, or by the speaker saying something to make them true). We do not of course expect that, in natural occurring dialogues, each schema application will faithfully follow the definition: we will therefore allow some of the constraints to be expressed as "implicit premises". This means that there will be no specific annotation, but we assume that the speaker has made hypotheses on these constraints' satisfaction.

Dialogue Moves

At the lowest level, the dialogue structure comprises single dialogue moves, typically extracted from a pre-established set of basic moves. We base ourselves on the move definition given in Chapter 3 (see page 65), and we will therefore annotate the following moves:

1. *Claims*, any statements, either to make a new point, or to back/dispute a previous one;
2. *Acknowledgements*, that is statements which agree/accept another agent's point, or "re-state" one's own point.
3. *Replies*, that is statements which reply to questions previously posed.
4. *Disputations*, that is statements which disagree on previously made points.
5. *Withdrawals*, that is statements which deny one's own previously made points.
6. *Open Questions*, that is requests for information, on items the querying agent does not suppose previous knowledge.
7. *Requests for Argument*: that is requests for support to a new claim.
8. *Challenges of Claim*: that is requests for support to a claim previously made by another agent.
9. *YN queries:* that is questions whose only reply can be either Yes (True) or No (False).
10. *YN answers:* answers to a YN-question.

Note that at this level we do not annote in a special way whether a claim supports or disputes another claim: this judgement is in the realm of the rhetorical level, and will not appear at the dialogue level.

As noted in Chapter 3, unlike other authors, we do not explicitly distinguish among several games. The notion of "game board" we gave in Sect. 3.3.3 (page 72) allows us to note well formed sequences of moves, as well as "changing game" moves, but we do not name different games differently. While it is useful to distinguish between game types when such notion carries most of the information on the dialogue dynamics, we also believe that our three layers allow us to distribute this information better. Therefore, the check on the participants' goals, that is embedded for instance in the definition of a game in DMT (Mann 2002c), is done for us either at the rhetorical level for argumentative goals, or at the strategic level, for higher order ones. The notion of game type described by Carletta *et al.*, or by Kowtko, Isard, & Doherty (1992), on the other hand, is monitored by the game board, but not explicitly annotated. This notion corresponds to defining the games on the basis of "typical" moves (so that for instance a "query" game means that there will be a "query" type move followed by a "reply" type move), so our annotation could in fact capture at no extra cost.

Specific guidelines for annotating dialogues according to our schema are described later, and exemplified by means of a collected corpus of natural dialogues. We start, therefore, by describing this corpus.

5.3.3 A Corpus of Dialogues on Nutrition

We sought to build up a corpus of naturally occurring dialogues in the chosen domain of application. Rather than incur the considerable effort of running an experiment with a setting like nutritionist surgeries, we decided for a more immediate, and perhaps more interesting solution. We opted for collecting e-mail conversations, rather than face to face ones. Besides the obvious advantage of being able to contact people from virtually any place in the world, at basically no cost, we hoped that the e-mail setting would make the participants in the experiments more comfortable, as they could decide when and how to reply, they could think a bit more on what to say, and they could stop at any time without feeling the pressure to go on that face to face interaction can produce. Moreover, as we are interested mainly in the content of the message to communicate, we could concentrate completely on the text of the dialogues, without having to accomodate other interfering factors like intonation, gazing, engaging moves (like “mm”, “yeah”, etc.), and the like (Coulthard 1992). The mechanism of turn taking was also greately clarified, with one turn corresponding to one e-mail message. To better control the experiment, we were acting as one of the dialogue partners.

A preliminary corpus of five dialogues was created with the involvement of five different researchers from the Department of Human Nutrition of the University of Glasgow. The theoretical aim of the experiment was to establish whether real nutritionists explicitly use models of behaviour change and, more importantly, whether an argumentative style of discourse could be detected and analysed in terms of the New Rhetoric constructs. Another, more practical, aim of this experiment was to assess whether an overall structure of a nutritionist’s session could be identified, and a list could be collected of specific pieces of advice.

After having the aims of the project explained, the nutritionists were asked to engage in an e-mail exchange with an advisee on healthy nutrition matters. The experimenter “role played” five different advisees as five fictitious characters, each with nutrition attitudes corresponding to one of the five stages of change. Nutritionists were not constrained in the length or style of their messages, as the aim of the experiment was to keep the dialogues as natural as possible.

The dialogues comprised respectively 34, 16, 14, 11, and 4 e-mail messages, therefore turns. The 4-turn dialogue was not particularly informative, as there was no scope to develop a proper conversation. Two dialogues were initiated by the advisee, and three by the nutritionist, with the 4-turn dialogue in the latter group.

From time to time, the nutritionists were also asked to comment on stages of the dialogue, explaining why they posed a particular question, or whether they had an overall plan, and so on. These were treated as meta-comments on the moves, and not included in the dialogue analysis, although the comments were referred to in the analysis process. Two nutritionists, for example, explicitly mentioned in the meta-comments that they were using a stages of change approach to the conversation. A general overall policy in the conversation was for the nutritionist to focus on the advisee's replies, and build the session on these.

The preliminary corpus was used in the development of the annotation framework. In order to test the framework, data was used from a second corpus of e-mail dialogues. This corpus was collected in a second experiment, which involved participants recruited from a mailing list with an interest in nutrition. The recruiting message explained that a new nutrition education computer system was being tested, but that, in the experiment, a real person would in fact interact via e-mail, simulating the behaviour of the system. No mention was made, however, of the interest in the argumentative nature of the interaction, to avoid the participants paying too much attention to this aspects. The mechanics of the dialogues were the same as the first experiment, but in this case the experimenter played the role of the adviser. The experiment lasted a month, involving 46 participants, with more than 500 e-mail messages exchanged overall. The average dialogue length was 11 turns, with the longest dialogue consisting of 45 turns.

Excerpts of dialogues coming from both experiments are shown in Fig. 15. In the figure, the adviser's turns are labelled AR, and the advisee's ones are labelled AE. One complete dialogue is shown in Appendix B.

One notable feature common to basically all dialogues (in both experiments) was that there was no proper conclusive message/exchange: at some point the conversation "faded away", typically after an experimenter's turn. Participants in both experiments were not pushed to continue any longer, as it was thought that this fuzzy stop to the conversation, although unthinkable in voice dialogues (where even a party suddenly leaving or hanging the phone does mark the end of the conversation), is conversely typical, and well accepted, in the e-mail environment.

5.3.4 Analysing Argumentative Dialogues

The annotation of a dialogue according to a code scheme is aimed at reconstructing the structure of the dialogue, while recording the goals of the participating agents. Mann (2002a) remarks that the process of dialogue analysis is in fact a process of "affirming statements" on a dialogue, which are a matter of personal judgement. It is therefore important that the definitions that have to fit the portions of the dialogue are very clear to the analyst.

Mann (2002a) also provides a useful set of guidelines for dialogue annotations with DMT, which can be generalised to other annotations schemes. The guidelines recommend that the first task to perform should be the division of the dialogue in basic units, to avoid circularity of interpretation. The decision

AR *Do you like cooking?*
AE *Not especially. There are three or four basic meals I know how to cook and I don't have much interest in learning anything new (even though I know it's important). Cooking feels to me like a lot of effort for something (ie. eating) that's over quite quickly. Also, I often feel tired at the end of a day's work and don't want to spend too much time in the kitchen.*
AR *You do not cook just because you have to eat! Cooking can be a very relaxing and interesting activity, better than watching TV!*
AE *I know you're right but that still doesn't make it easy to do!*

AR *Have you ever considered having some fruit for breakfast or as a snack?* (instead of chocolate)
AE *I should do that, yes. I'll have to go and buy some....*
AR *Don't you have to go and buy chocolate as well?*
AE *I didn't mean it would take an extra effort to buy fruit on my trips to the supermarket. However, if I've had to leave home in the morning without having anything to eat, it's much easier to get hold of unhealthy snack food near work than it is to get fruit.*

AE *I do enjoy fruit, but it is not as appealing as say a biccie, and vegetables in summer aren't really the go, I would eat vegies in the winter time at least 4 times a week.*
AR *Maybe pears and apples are not that appealing, but what about satsumas, or a cup of strawberries, or a bunch of grapes? Nice and refreshing, especially in summer!*
AE *Yummy, I think if someone was to give me a plate of cut up fruit ie: like below then I would definitely eat it, it is always more appealing when it is all done for you.*

AE *When I eat with my friend I just do not take health into account: it's just to have fun, I don't want to spoil the evening out.*
AR *Is there any particular reason for feeling that being 'healthy' is not fun? do you feel that your friends will think you are being 'fanatical' about diet and health?*
AE *It's not that... but if, say, everyone wants to go for an Indian, I don't want be the one who says "no, thanks, I'd rather have a salad...".*

Figure 15: Extracts from the corpus of e-mail dialogues

of applying a given definition to a portion of the dialogue can be then made on the basis of three main heuristics:

1. *Whole interaction:* the complete dialogue is taken into account to interpret a single turn or move.
2. *Left to Right:* the past moves provide the context for interpreting the move under analysis.
3. *Moment of use:* as the turn immediately succeeding the move to analyse can shed some light on how *the other party* has interpreted the move, this information can be used in retrospective. This works because the speaker of the move to analyse will necessary use this context when starting to speak again.

The guidelines suggest the list above is a ranked list, the third heuristic being the best option, as the analyst should try and adopt the perspective of the speakers, while the dialogue unfolds.

Our three-layered analysis can be done on a table as shown in Table 14. The middle part of the table (headed *Dialogue*) lists the numbered dialogue moves, as broken down by the annotator into basic units of discourse. A move can be an entire speaker's turn, or a sentence in a turn, or smaller parts of a sentence, according to the analyst's judgement. The moves of the two agents are annotated separately, in the left and right sections of the table. This separation makes the table clearer in the cases where high level moves of a speaker span over several turns, and overlap those of the other speaker.

The annotation of the three levels of move can be done either bottom up, starting from the move level, or top down, starting from the strategic level. There seem to be no suggestion that one approach is better than another, apart from the analyst's own preferences.

Assuming a bottom-up approach to the analysis, starting from the move level, up to the strategic level, the analyst will annotate, in the two columns headed *Move*, the specific dialogue move used, according to the dialogue move list. For each move, the move type is indicated, as well as, if appropriate, the number of a referenced preceding move. As indicated in Sect. 3.3.2 (page 65), all moves but Claims, questions (either Open or YN) and Requests for Argument, refer to a preceding move.

In the second phase, the analyst will look for occurrences of rhetorical schemata. The main task, in this process, is to look for the speaker's affirmation, either explicit or implicit, of the claim, and the verification of the satisfaction of the ontological constraint of the schemata, for instance the fact that pragmatic argumentation has to be applied to action. The second important task is to evaluate the quality of the schema application, by verifying the admissibility constraints. Naturally, not every speaker will use an argumentation schema in its complete form: we have seen in Ch. 2 that the schema verbalisation can be shortened if some of the constraints are assumed to be true, or if the speaker is less attentive with providing good arguments. In any case, instances of the speaker specifically testing (e.g. with a question), or actively making the constraints of a schema true (e.g. by communication), are identified and marked by the annotator. The two columns headed *Schema* will be filled in at this stage, with the specific mention of the constraint satisfaction process. Note that a schema can span over several dialogue turns: this will most certainly happen when the constraint checking is explicitly done, as the speaker has to allow for the hearer to reply to the question.

The final phase looks for manifestations of the two agents' higher order moves, and depends on the specific domain about which the dialogue takes place. In our health promotion dialogues, the strategic moves will be labelled with the stage of change that is currently being acted upon: a list like the one in Table 13 can be used to mark specific subgoals.

Agent 1			**Dialogue**			**Agent 2**		
Goal	Schema	Move	No.	Content	No.	Move	Schema	Goal
Exp		YN-Q	1	*Do you like cooking?*				
				Not especially.	2	YN 1		Decl
				There are three or four basic meals I know how to cook and I don't have much interest in learning anything new (even though I know it's important).	3	Claim	By Means	
				Cooking feels to me like a lot of effort for something (ie. eating) that's over quite quickly.	4	Claim		
				Also, I often feel tired at the end of a day's work, and don't want to spend too much time in the kitchen.	5	Claim	By Effectiveness	
Act	By Ends	Disp 4	6	*You do not cook just because you have to eat!*				
		Claim	7	*Cooking can be a very relaxing and interesting activity, better than watching TV!*				
				I know you're right	8	Ack 6	Establish Hierar.	Resp
				but that still doesn't make it easy to do!	9	Ack 4		

Table 14: An annotated dialogue fragment

5.3.5 An Example of Annotation

Table 14 shows an example of annotation, applied to Dialogue 1 of Fig. 15.

Dialogue Moves

The first level of the annotation does not present major difficulties. The breaking up of turns into moves follows the grammatical structure of the message, where one move corresponds to one complete sentence. While a finer grained decomposition could be possible, in this case it does not add new information, as most of the moves in the example are in fact Claims. Had sentences included several move types (e.g. an Acknowledgement, followed by a Disputation, and a Question all in the same sentence), they would have been decomposed into several moves. In the Table, it is worth noting that there are moves referring to preceding ones: the YN Reply in Move 2 to the YN-Question in Move 1, the Disputation in Move 6 of the Claim in Move 4, and the two Acknowledgements in Moves 8 and 9, the latter as a restatement of a speaker's own claim. We point out once again that at this level we do not annotate how claims can be used to back or to dispute other claims, as this is in the scope of the second level analysis.

Rhetorical Moves

At the rhetorical level, moves are interpreted according to the list of rhetorical schemata. For the sake of clarity, the table does not show where the application of the argumentative schema has been checked against the constraints, but we detail this task below. For the same reason, in the table we put the label with the schema name where the main claim is, and we explain here where the constraint checking has taken place. While describing the analysis at this level, we also take the opportunity to introduce a few more argumentative schemata, besides those alredy detailed in Ch. 2.

After the preliminary exchange, which does not provide any argument, Agent 2's whole turn is based on a single causal relation. The causal one is a relation of "the structure of reality" that links an "act", or a "means", to a "consequence", or "an end". In the example, Agent 2 exploits the link between cooking and eating. From the causal relation, several argumentation schemata can be constructed, which may emphasise either the act/means, or the consequence/end side of the link, and may aim at valuing or devaluing this side. The schemata are listed in full in Appendix A. In the example, Move 4 is an argument "By Means". This is a way to devalue an act by reversing it into a means: cooking is only done because one needs to eat. The admissibility constraints for this schema ask that the end achieved is recognised as valuable by the audience, but from a low ranked perspective. This is obtained by stressing that eating is over soon, therefore implying that is not enjoyed as such. Move 5 backs the devaluation of "cooking" by stressing that cooking is not even the best of the means to achieve the end. The argument "By Effectiveness" aims at valuing or devaluing acts/means according to whether they are convenient for the job. Admissibility constraints ask that the audience believes the act/means is not appropriate, and this is obtained in various ways: in Move 3 it is remarked that the actor is not an expert in the act itself, and in Move 5 it is claimed the act is tiring and time consuming (re-elaborating on Move 4, where it is explicitly claimed it takes effort).

Agent 1, in Moves 6 and 7, builds upon the causal link provided by the opponent, and responds with the classical counterargument to the schema "By Means": the argument "By Ends" values an act by reversing it into an end. Cooking is valuable as such, not just as a means to eat. The valuation is done from the perspective suggested by Agent 2 as one of interest: cooking is relaxing.

Finally, the double Acknowledgements by Agent 2 in Moves 8 and 9 serve to re-establish the hierarchies of values: although it is recognised that cooking could be interesting, this cannot outweigh the fact it is difficult to do.

Strategic Moves

At the strategic level, in Move 1 Agent 1 makes an Exploratory move to establish the stage of change of the other agent with respect to the habit of preparing one's food, rather than buying pre-packed meals. This is a subgoal of making the advisee progress to the Maintenance stage with respect to healthy eating (in the preceding moves of the dialogue, Agent 2 had already shown to be in the Action stage).

Agent 2 makes an extensive Declarative move, by replying to the question (Move 2) and providing extra information in support of his health beliefs.

Agent 1 triggers the Action goal and starts providing new information/points of view on the activity of cooking, but Agent 2 in a Responsive goal, shows that the advice is not yet accepted.

5.4 Conclusions

We have presented a coding scheme to annotate argumentative dialogues in the domain of health promotion. Our approach to annotation, to the best our knowledge, is the only one which combines the need for precisely capturing the unfolding of the dialogue game moves, the argumentative nature of the dialogues, and the way in which the arguments serve higher level goals of the participants. Most importantly, every notion in the scheme is grounded on well established discourse analysis, rhetorical and behavioural theories.

We have applied our scheme to a corpus of dialogues, in which we have indeed found occurrences of rhetorical moves. As the domain was set as an example, independently from the definition of the schemata, we may attempt a general consideration that our New Rhetoric schemata can be used to describe naturally occurring dialogues, provided that the scenario is an argumentative one, and that one could identify strategies, like the behavioural theories ones, that lead the two speakers to make use of the schemata.

An important test for any annotation scheme is its reliability, that is the fact that it may be consistently applied and used by people other than the developers, and that the same analyst gives similar analyses over time (Carletta 1996). This is the single most important step that is still in progress: so far the analyses have been done by the developers, although sometimes by consultations with colleagues[3] . A trial study is planned, with a set of annotators, to be trained on the coding scheme, but not necessarily familiar with either the domain or the theories behind the scheme. The task of training annotators is not to be underestimated: as explained above, the annotators should be very clear on the meaning of the definitions that have to fit the dialogue moves. The rhetorical moves are particularly deemed to

[3] We are profoundly indebted with Bill Mann for comments and advice on annotating the dialogues.

create difficulties, as the rationale behind Perelman and Olbrechts-Tyteca's original definition, and its adaptation, can be difficult to interpret.

A major drawback of our scheme, as it is at the moment, is that it does not propose any "constructive" theory for evaluating the coherency of argumentative dialogues, like for instance RST does for general (especially explanatory) discourse. We do not, therefore, address completely the criticisms that RST attracts when applied to argumentative discourse. A more complete annotation would capture how arguments can be built upon other arguments, in order to evaluate the quality of the whole argumentation, as well as the coherency of the dialogues. We believe, however, that such theory can only be built on grounds like those we propose, that is by accurately providing the annotation process with a rhetorical argumentative connotation.

Chapter 6

Modelling Knowledge

To speak of knowledge is futile. All is experiment and adventure.

- Virginia Woolf, *The Waves*

We continue in this Chapter the exercise started in the previous one, focusing our considerations on how to design a rhetorical argumentation system operating in one specific domain, the promotion of healthy nutrition. In this Chapter, we explore the knowledge representation requirements, and we proceed by applying insights from a relatively new branch of knowledge representation research, that is ontology engineering. In doing so, we also make some observations concerning the kind of knowledge needed to implement a rhetorical argumentation system in general[1].

6.1 Introduction

Reiter, Robertson, & Osman (1999) have noted that in order to produce effective advice in a health promotion setting, different types of knowledge are required:

- knowledge of how people change addictive behaviours;
- practitioner knowledge about empathy;
- linguistic knowledge and
- medical knowledge about the health behaviour to promote.

We too have noted, in the previous Chapters, that in order to argue rhetorically one needs to reason about values and opinions. We have also seen that one needs to know basic facts in the domain of discourse,

[1]This Chapter has immensely benefited from numerous discussions with Valentina Tamma.

in our case at least basic nutritional notions. While Reiter, Robertson, & Osman concentrate on the ways in which such knowledge can be acquired from experts, for example via "think aloud" techniques (see also Reiter, Robertson, & Osman (2000)), we are in the somewhat opposite position of having various material to start with, that has to be organised. This material consists of the notions in the New Rhetoric theory described in Ch. 2, the concepts specific to the health promotion model we use, described in Ch. 5, and basic nutrition information, all of which has had some substantiation in the e-mail dialogues presented in Ch. 5. We need to reconcile in a coherent and harmonised knowledge base all the different kinds of expertise. In one word, we need an orderly *conceptualisation* of the domain, that is a definition of the concepts and the relationships among them that are presumed to exist in the domain (Genesereth & Nilsson 1987).

In pursuing this goal, we benefit from best practices coming from one of the most recent, and vibrant, fields of research in Knowledge Engineering: *Ontologies*. The combination of different representations of various aspects of reality is among the issues that researchers in Ontologies have been studying in recent years. The aims are diverse, the most important of which is to build knowledge that can be re-used. Many researchers in this field have proposed several methodologies, both to build and to integrate knowledge. In designing the knowledge base of our adviser, we therefore take advantage of this experience, and use one of these methodologies as a guideline.

In this Chapter this exercise is described, after introducing the major issues in research on ontologies.

6.2 Ontologies

Ontology, as a branch of metaphysics, deals with the nature of being: the study of what exists.

Aristotle's *Categories*, reorganised by Brentano, and Kant's *Triads*, analysed and developed by the like of Hegel, Peirce, Husserl, Whitehead, and Quine, among others, are the milestones in this effort to understand how everything we experience could be categorised and characterised. In the words of Sowa (2000), classifications, or ontological categories, "provide hooks to which the definitions and axioms of a knowledge base are attached" (p. 76). One can therefore distinguish precisely, for instance, what is Physical from what is Abstract, by explicitly describing what it means for something to occupy some space, last some time, change, be represented, be caused by something else, and so on.

As noted by Sowa (2000), while the philosophers are interested in the general, grand categories, AI researchers tend to start from the bottom, and scale up knowledge bases that are created for single applications. Because of the massive proliferation of information available in recent times, it has therefore become crucial that the abundance of efforts to build knowledge bases, or indeed databases, for a given domain should not be limited to one application only, but should benefit many others. Thus, despite its ancient origin, and despite the fact that knowledge representation and processing researchers have always had to deal with understanding the nature of what needs to be represented and processed, ontology has only relatively recently become an explicit field of research in AI (Guarino & Poli 1994).

The attention of this new field, which has been named *Ontologies* to distinguish it from the classical philosophical trend, has turned towards both general ontology issues and knowledge engineering design methods. The hypothesis is that any hope for communication and sharing between knowledge bases lies in requiring that different representations of knowledge have to agree on a set of "ontological commitments", which constrain what to see about the world and how to see it (Davis, Shrobe, & Szolovits 1993).

The most widely cited definition in this new community, due to Gruber (1994), is that an ontology is an *explicit specifi cation of a conceptualisation*. This definition has been spelt out in detail by Studer, Benjamins, & Fensel (1998), by saying that:

> "An ontology is a formal, explicit specification of a shared conceptualisation. A conceptualisation refers to an abstract model of some phenomenon in the world by having identified the relevant concepts of that phenomenon. Explicit means that the type of concepts used, and the constraints on their use are explicitly defined. [...] Formal refers to the fact that the ontology should be machine readable. Shared reflects the notion that an ontology captures consensual knowledge, that is it is not private to some individual, but accepted by a group."

The reason why it is crucial to come to an explicit and agreed conceptualisation is to "reduce or eliminate conceptual and terminological confusion and come to a shared understanding" (Uschold & Gruninger 1996). This could allow knowledge engineers to build up on existing works, by using, and possibly extending, other products (Gruber 1992). The effort of agreeing on the ontology has therefore the payoff that the task of building knowledge is greatly facilitated: if a knowledge engineer needs to be able to represent, for instance, the notions of Space and Time within the domain of application in which the knowledge base has to be built, an appropriate off-the-shelf ontology, perhaps also with useful inference axioms, allows the effort to be confined to just the main domain. In this perspective, particular attention is given to situations in which overlap and problems of possible misunderstandings might exist because knowledge is provided by different sources (people, theories, etc.).

Research on ontologies is not limited to methodologies to reuse pre-existent represented knowledge, but is also devoted to "developing theories, methodologies and tools to elicit and organise domain knowledge, in a reusable and 'transparent' way" (Guarino 1997). We survey some of them in the next Section.

6.2.1 Ontology Development Methodologies

Despite the importance, and the historical character of ontological issues, not very many methodologies have been proposed to guide the ontology conceptualisation process, and research in this field still has to agree on a standard approach (Jones, Bench-Capon, & Visser 1998; Gómez-Pérez *et al.* 2002; Gómez-Pérez, Fernández-López, & Corcho 2002).

Gruber (1994) first proposed some useful criteria to evaluate the quality of an ontology, especially if it is meant for knowledge sharing:

- **clarity**: the definitions in the ontology should be clear, objective, and, if possible, complete (i.e. defined by “necessary and sufficient” conditions);
- **coherence**: the ontology should prevent inferences that contradict the definitions;
- **extendibility** the ontology should grow monotonically, with new concepts that can be built on existing definitions, while preserving coherence;
- **minimal coding bias**: the conceptualisation should not depend on the final encoding,
- **minimal ontological commitment**: the ontology should make the least possible commitments on the world to model, so that the agents accessing the ontology can specialise it and instantiate it as needed.

While these criteria can provide guidance in the construction of the ontology, for instance to test preliminary versions, they do not offer a specific methodology that could guide the building process. In addressing this point, Jones, Bench-Capon, & Visser (1998) observe that proposals can be classified into two categories: stage-based models and evolving prototype models.

The most significant **stage based model** was proposed by Uschold & Gruninger (1996). The model suggests that the development of an ontology should follow the following steps, or stages:

1. identification of purpose and scope: it should be clearly spelt out the reason for building the ontology in the first place, and how it is going to be used;
2. construction of the ontology, possibly integrating existing ontologies: this is the main activity, that requires two main tasks:
   (a) ontology capture: the identification of concept and relations, and
   (b) ontology coding: the definition of concepts and relations in a chosen (formal) language
3. evaluation of what has been produced against the initial requirements

4. documentation.

These notions have been developed and put into practice by Gruninger & Fox (1995) in the construction of an enterprise ontology, as part of the ToVE (Toronto Virtual Enterprise) project. In this task, the original set of stages were refined, and a more complete set is presented. The final set of steps involve:

1. capturing *motivating scenarios*: these are story problems, examples of applications, or any other collection of real case scenarios that can suggest both general and particular needs for the ontology to address;

2. formulating *informal competency questions*, that is a set of informal queries the ontology has to be able to address;

3. specification of the terminology of the ontology within a formal language;

4. formulation of more formal competency questions, in the terminology of the ontology;

5. specification of axioms and definitions for the terms in the ontology in the formal language;

6. justification of the axioms and definitions, perhaps by solving characterisation problems.

A different approach to the construction of ontologies is the **evolving prototype model**, where ontologies are built by means of a set of intermediate representations, of increasingly higher level of formality. Deliverables are issued at the end of each stage that are used as bases for the the subsequent stages.

The most significant approach in this category is METHONTOLOGY (Blázquez *et al.* 1998), which is grounded on a software engineering concept of life cycle. The approach identifies stages in the designers' development, each one corresponding not only to specific activities, but also to documents to be provided describing these activities:

1. specification of the ontology, that is identification of purposes, intended users, degree of formality, and so forth;

2. knowledge acquisition, running mostly in parallel with the first activity;

3. conceptualisation, that is identification of terms, etc. by means of a suitable informal representation;

4. integration with other existent ontologies, if applicable

5. implementation, in a formal language;

6. evaluation;

7. documentation.

Indeed, a great emphasis is placed by the authors on the affinities between METHONTOLOGY and the IEEE Standards for developing software (Fernández 1999), not least the great importance given to the evaluation phase.

A similar life cycle model is proposed in the IDEF5 project (KBSI 1994), where the identified activities are:

1. Organising and scoping of the ontology: establishing purpose and content;

2. Data collection: generally by means of typical knowledge acquisition techniques;

3. Data analysis: extract the ontology concepts from the raw data collection;

4. Initial ontology development: concepts extracted are put together in a first prototype of the ontology, with an initial description of kinds, relations and properties;

5. Ontology refinement and evaluation: the prototype is iteratively refined and tested, by "instantiating" concepts with actual data.

Jones, Bench-Capon, & Visser (1998) suggest that stage based approaches are more appropriate when the purpose and the requirements of the ontology are clear from the beginning, while evolving prototypes approaches are applicable when no clear purpose has been identified[2]. In the remains of this Chapter we will show how we interpreted these guidelines in the design of our knowledge base.

6.3 An Ontology Engineering Approach to Design the Knowledge Base

The problem we needed to tackle was not so much an engineering one, but one related to the more philosophical issues in ontology: deciding the top level ontological concepts and how they relate. However, we had also to deal with the issue of integrating suggestions coming from different and heterogeneous sources, not least the collection of e-mail dialogues presented in Ch. 5. In order to have an orderly approach to the task, we therefore felt that a well defined methodology could be of great help.

[2]Besides these two mainstream approaches, others deal with situations where it is not appropriate to engineer a new ontology from scratch, but existing ones can be integrated and modifi ed. Others still are based on the cooperation of several sites to the construction of the ontology. A comprehensive list of the various approaches is given by G´omez-P´erez, Fern´andez-L´opez, & Corcho (2002).

In choosing which approach to use, we felt that methods based on prototyping were less appropriate to our specific problem. While they provide significant help at the subsequent stages of implementation and maintenance, they offer less guidance at the conceptualisation stage, where one needs first and foremost to establish a principled way to look at the reality to represent. Although no approach explicitly deals with this problem, the stage approaches seemed to be closer to our purposes. In particular, the model developed by Gruninger & Fox (1995) in the realisation of the ToVE ontology, with its emphasis on real case scenarios and competency questions, seemed the natural way to proceed, with the sources we had at our disposal. This Section explains how we applied the methodology to the development of our own ontology, by following the proposed decomposition in stages.

6.3.1 Motivating Scenarios

The first step of the methodology advises collecting a series of "motivating scenarios" in order to understand the kinds of issue the ontology has to address.

The e-mail dialogues collected in order to test the appropriateness of the argumentative and behavioural theories to the case study domain, presented in Ch. 5, were particularly well suited to act as a source of realistic motivating scenarios. The dialogues could in fact be considered as typical situations, scenarios, that we wanted to be able to replicate in our system. Therefore the specific topics of these dialogues, the way they were conveyed by the participants, the questions asked and how they were addressed, and so forth, all represented examples of notions that the ontology had to represent, in order to build a system able to engage in similar conversations.

As an example of motivating scenario, let us consider one of the dialogue excerpts we have already shown in Ch. 5 (see page 126):

Advisee: *When I eat with my friend I just do not take health into account: it's just to have fun, I don't want to spoil the evening out.*

Adviser: *Is there any particular reason for feeling that being 'healthy' is not fun? do you feel that your friends will think you are being 'fanatical' about diet and health?*

Advisee: *It's not that... but if, say, everyone wants to go for an Indian, I don't want to be the one who says "no, thanks, I'd rather have a salad...".*

Each of the dialogue turns can show how the domain is perceived by the dialoguing agent, from very basic notions, such as "eating can be done with friends", or "Indian is a type of cuisine", to more abstract concepts, such as valuing something for being "fun", or proposing an argument by incompatibility between paying attention to health issues and being easy going when out with friends. These are captured, at different levels, by the dialogue annotation scheme described in Ch. 5. The content of the ontology can therefore be inspired by these scenarios, as shown in the second stage below.

6.3.2 Informal Competency Questions

The second stage of the methodology requires to formulate, on the basis of the motivating scenarios, some informal competency questions (ICQ). As suggested by Uschold & Gruninger (1996), ICQs can be defined in a *stratifi ed manner*, by identifying high level questions first, and lower level questions needed to solve them. In order to have a more organised approach, we defined an:

> **Definition 9 (ICQ schema)** *An Informal Competency Question (ICQ) schema is a structure comprising the following slots:*
>
> **ICQ number:** *a number to identify the question;*
> **Question:** *an expression in natural language of the question;*
> **Rationale:** *the reason why such a question was needed. This may either refer to another ICQ, to which this question provides an answer, or to an excerpt of a motivating scenario, or to a New Rhetoric schema which the question helps to implement.*
> **Decomposition:** *what needs to be known in order to answer the question.*

A set of ICQs was created on the basis of the schema above by analysing the corpus of dialogues in the light of the three perspectives we needed to represent:

- the argumentative perspective, that is how people talked about values and how they put forward their arguments;
- the behavioural change perspective, that is how people talked about their health states and how they can be modified;
- the nutritional perspective, that is how people talked about food and diet.

The annotation of the dialogues, which was performed exactly following these three perspectives, was naturally reused for defining the ICQ as well. It is important to emphasise that we paid attention to *how people talk* about things, rather than to the reality itself. In this sense, our effort goes in an opposite direction to what classical ontological studies aim at, that is to obtain a description of the reality as objectively as possible. The final product is nevertheless “a description of the reality”, only it is given from one point of view.

Three examples of ICQs are given in Fig. 16[3]. Some ICQs are “top level” ones (like the one at the top of the figure), expressing the need for the ontology to support rhetorical argumentation. The first turn

[3]The complete list of high level ICQs can be found in Appendix C.

of the dialogue example above shows a motivating scenario for such an ICQ: the advisee expresses that health has “no value” when going out with friends, that is from a “social life” perspective. If asked, she should be able to support this claim with an appropriate argument.

Some other ICQs are more directly related to New Rhetoric schemata (e.g. middle ICQ in the Figure), and their definition mirrors the definition of the schema itself. In the dialogue example above, the second line is a motivating scenario for including in the ontology tools to argue by Incompatibility, as the adviser is challenging the claim that staying healthy and having fun cannot be compatible.

The ontological constraints defined by the New Rhetoric schema are then used to establish another set of ICQs (e.g. bottom ICQ in the Figure) which express the fact that relations among notions of the reality have finally to be included in the ontology in order to be able to implement the schema itself.

6.3.3 Specification of the Informal Terminology

This stage of the ontology construction is aimed at listing, starting from the ICQs, the specific concepts that the ontology will represent, with their relationships. The way in which the ICQs were formulated suggested that the ontology could be produced in a modular fashion, where the main sub-ontologies that need to be included, as shown in Fig. 17, are:

- an **ontology of health beliefs**. This is needed for representing the stages of change and the beliefs about health. Having as basic concepts *actions* (e.g. “eating fruit”), *states* (e.g. “having high cholesterol values”) and *persons*, we want to be able to express that:
  - an *individual* can be $\left\{ \begin{array}{l} susceptible \\ insusceptible \end{array} \right\}$ to a *state*;
  - an *individual* can be $\left\{ \begin{array}{l} Precontemplator \\ Contemplator \\ InPreparation \\ Active \\ HasHabit \end{array} \right\}$ with respect to an *action*
- an **ontology of argumentation** is also needed, and especially a way to express the fact that a topic is “valuable”, at least from a certain perspective. Having as basic concepts *topics*, *values* (Good, Bad, Indifferent) and *perspectives* (e.g. QualityOfLife, Health) we should be able to express the following relations:
  - a *topic* can be considered from a certain *perspective*

ICQ number:	1
Question:	Is there an argument in favour of a topic T?
Rationale:	Main aim.
Decomposition:	One of the following: 1. the topic T is valuable 2. a New Rhetoric schema can be applied in favour of T

ICQ number:	7
Question:	Is there an argument by Incompatibility in favour of an action A?
Rationale:	Solve ICQ 1 and implementation of New Rhetoric schema.
Decomposition:	One of the following: 1. Action A leads to a state S, and there is an argument by Incompatibility in favour of S. 2. There is an action A_1, which is valued negatively, and A and A_1 are not compatible

ICQ number:	47
Question:	are states S and S_1 incompatible?
Rationale:	Solve ICQs 6, 5, 50
Decomposition:	One of the following: 1. S prevent S_1 from happening, or vice versa 2. S leads to a state S_2 which is incompatible with S_1 3. S_1 leads to a state S_2 which is incompatible with S 4. S leads to S' and S_1 leads to S_1' and S' and S_1' are incompatible.

Figure 16: Examples of Informal Competency Questions

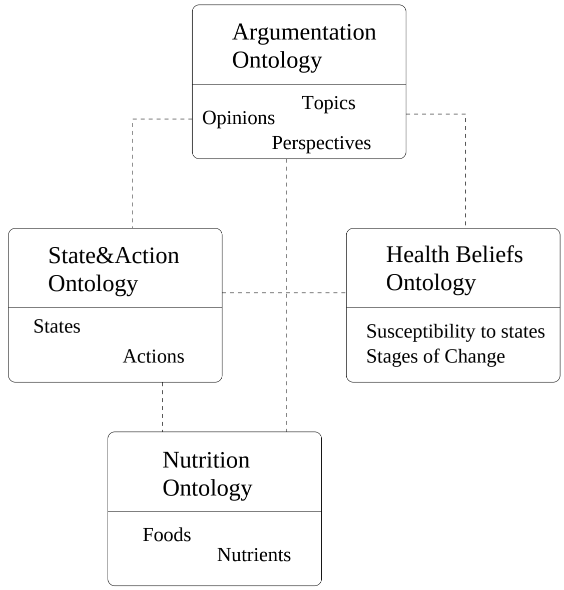


Figure 17: Sub-Ontologies

- a *topic* can have a *value* from a certain *perspective*;
- a triple {*topic*, *perspective*, *value*} can be more valuable than another similar triple (this encodes the New Rhetoric's concept of *hierarchy of values*)
- a *perspective* can be more valuable than another *perspective* (this encodes the New Rhetoric's concept of *locus*)
- a *perspective* can be more general than another *perspective*

A topic can be an action (*Eating Fruit*), a state (*Having High Cholesterol*) or an instance of a relation (*EatingFruit can Prevent Cancer*). We wanted, in summary, to be able to express sentences like:

- *EatingFruit can be seen from the Health perspective*
- *Cancer is Bad from the Health perspective*
- *the fact that EatingFruit can Prevent Cancer is Good from the Health perspective*.

Less crucial to our research, but necessary for a complete implementation of the adviser, at least two other sub-ontologies are needed:

- an **ontology of states and actions**
- an **ontology of food**.

Many so called "upper level", or "top level" ontologies could provide the former (Bateman, Magnini, & Fabris 1995; Cycorp 1997; Miller 1995, for example): the principle of *minimal ontological commitment* stated at the beginning of this Chapter should guarantee that "plug-in" other ontologies is achieved with little or no effort. We have some requirements, derived from the analysis of the motivating scenario and the theories, that a possible ontology needs to satisfy to be safely integrated in our system. Apart from being able to generally express action, states, steps, length of an action, pre-requisites, and so on, at a higher level of abstraction, we need an ontology of actions and states able to express that:

- an *action* $\left\{\begin{array}{l} can \\ cannot \end{array}\right\} \left\{\begin{array}{l} achieve \\ maintain \\ favour \\ strengthen \end{array}\right\}$ a *state*;
- an *action* $\left\{\begin{array}{l} is \\ is\ not \end{array}\right\} \left\{\begin{array}{l} usually \\ always \end{array}\right\}$ done at a certain *time*, and/or in some *location* and/or on some *object* and/or involving some *effort*.

As to a food ontology, the task is more problematic. Several medically or biologically oriented ontologies exist, such as GALEN (Rector *et al.* 1995), for instance to represent nutrition disorders, or nutrients contained in food. Also, some general concepts could be integrated from upper level ontologies, like what it means to cook, digest, etc. We were however not able to find a comprehensive food ontology, in which characteristics of food are expressed in term of serving types, how they are transformed when cooked, which properties they acquire or lose with time etc. The development of such an ontology is a challenge in itself, which we will not even approach here. We limit ourselves to expressing a few basic concepts that we need. As mentioned before, the designed ontologies, and the food one in particular, should be seen as the description of what people think about the reality, rather than the reality itself. From this perspective, looking at the e-mail corpus, we need at a high level to be able to express that:

- food can be prepared in various ways, so becoming a "meal"
- nutrients are found in meals in various proportions.

6.3.4 Specification of the Formal Terminology

From the informal terminology, the methodology suggests specifying the actual ontology in some formal language. In describing the cognitive ontology of our nutrition adviser, we make use of Object Oriented (OO) representations of concepts and relationships. The notation for the OO representation we use is

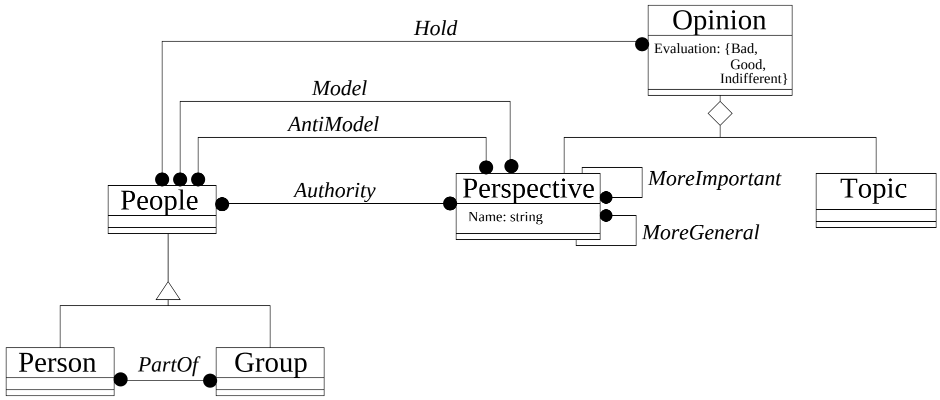


Figure 18: Ontology of Arguments

the one by Rumbaugh *et al.* (1991), and is summarised in Appendix D. Let us proceed looking at each sub-ontology in turn.

Argument Sub-ontology (Fig. 18)

Central to this ontology is the concept of *Opinion*. An opinion exists whenever an association can be made between a given topic and a certain perspective, with an evaluation (Good, Bad or Indifferent).

An essential element in describing an opinion is the specification of who holds it: the spirit of our adviser is that there are no incontrovertible truths, and each opinion is related to somebody, some *People*. People can be either single *Individuals* or *Groups* (such as "Italians"), where individuals can be part of a group. Various relations may hold between people and perspectives, in order to be able to express, for example, arguments by model/anti-model, by authority etc.

An important restriction introduced, as already pointed out in the general theory of argumentation description (Ch. 2), is to constrain the attribution of a value to a topic to be made via a perspective. In this view, for example, having high cholesterol values is not good or bad *per se*, but it is so from, say, the health perspective. This, we repeat, facilitates a more natural style of argumentation, in which the same topic can have simultaneously good and bad values with respect to different points of view.

Health Belief Theory Sub-Ontology (Fig. 19)

The basic concepts of the Health Belief Model and the Stages of Change model are expressed in this sub-ontology:

- a person can (feel to) be susceptible or not to a state
- a person can be in one of the Stages of Change towards an action

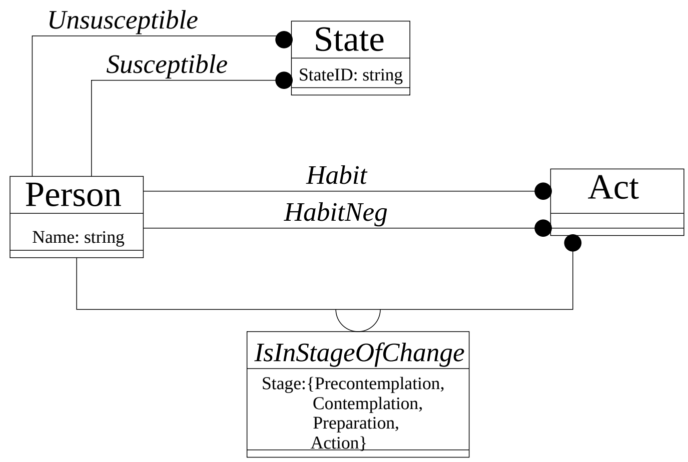


Figure 19: Ontology of Health Behaviours

The final stage of change as hypothesised by Prochaska & Clemente (1992), the "maintenance" stage, has been separated from the other stages, and represented as a relation of *Habit* between the person and the act. The decision was mainly due to the fact that the expression of a person's habit has strong links with argumentation techniques, such as the one by model and anti-model, something that does not happen with other stages of change. Incidentally, there is still controversy among behavioural researchers as to whether maintenance has really to be considered as a stage, therefore treating this stage of change in a special way has some grounding in the literature as well.

Act/States High Level Sub-Ontology (Fig. 20)

An *Act* is defined by one *Action* plus zero or more extra pieces of information about this action. These may concern the *Location* where the action is undertaken ("eating at home"), or the *Time* ("eating in the morning"), or the *Object* one can act upon ("eating apples"). These were the most frequent modifiers in the corpus of e-mail dialogues, but, as said before, we expect to be able to integrate a much comprehensive theory of action by using off-the-shelf ontologies. We also defined an "effort", or a cost for the action, in terms of different measures. This helps in representing scenarios in the corpus like "eating mangos is expensive" (measure = money) or "having breakfast is time consuming" (measure = time) etc. A relation of *MoreGeneral* among acts represents for example the fact that "eating oranges" is in fact a more general way to "eat VitaminC" (as oranges contain it), or that "swimming" is a more specific way to "exercise", while the relation of *NotInvolves* allows us to represent, for instance, that "eating fruit does not involve eating fat". Finally, a *Modality* can be added to the act concept, formed by a modality type and a sign which, in their four possible combinations, represent the notions of *always, usually, never, rarely*.

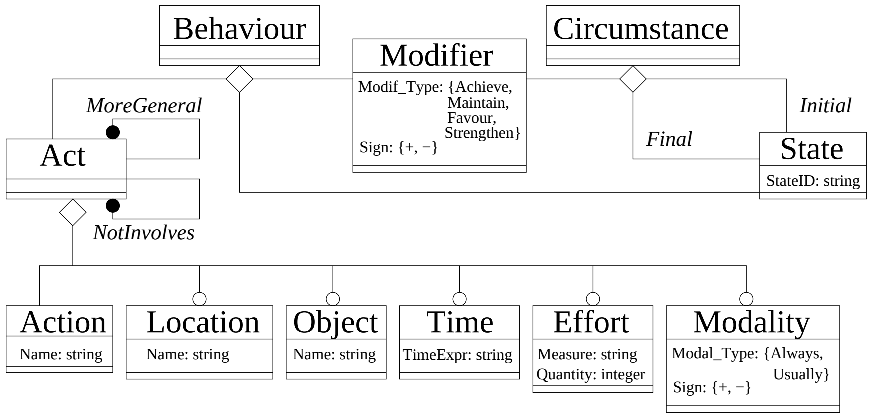


Figure 20: Ontology of Actions/States

Two higher level ontological notions express relationships between actions and states:

- **Behaviours**: this class expresses the notion of undertaking an act for a purpose. The type of this implication is characterised by a modifier (*achieve, favour, maintain, strengthen*) and a sign. The modifiers, associated with the "-" (minus) sign represent respectively the notions of *prevent*, *obstruct*, *lose* and *weaken*. Therefore we can say that "eating fruit can help in staying healthy" by using a modifier with Type = "Maintain" and Sign = "+" (plus).

- **Circumstances**: with this class we wanted to express the notion of two states being related. The same class of modifiers is used to connect an *initial* state and a *fi nal* state, so that we can say that "having high cholesterol levels obstructs being healthy" with Type = "Favour" and Sign = "-".

Food High Level Sub-Ontology (Fig. 21)

As noted above, much work is needed to define an appropriate ontology that can represent the whole range of human experiences of eating. As a sketch of the main concepts to elaborate upon, we define the notion of *Food*, that can be related to some *Preparing* method, thus acquiring the status of a *Meal*. Meals can be associated with *Nutrients*, in various proportions.

Integration of the Sub-Ontologies

Figure 22 shows how the various sub-ontologies are related to one another, and in particular the fact that:

- opinions can be expressed on one of the following topics:
  - Acts (e.g. "exercising is good from the health perspective");
  - States (e.g. "having high cholesterol values is bad from the health perspective");

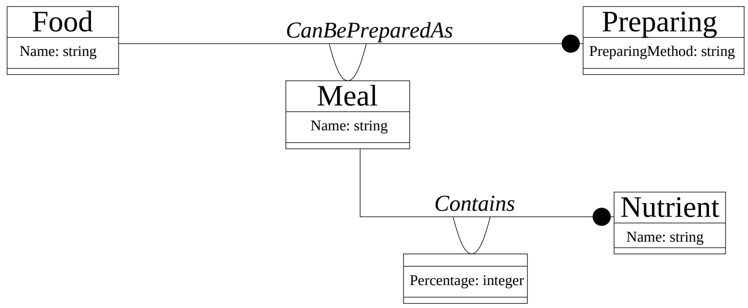


Figure 21: Ontology of Food

- Behaviours (e.g. "the fact that eating fruit favours lowering of cholesterol levels is good from the health perspective");
- Circumstances (e.g. "the fact that having high cholesterol levels obstructs being healthy is bad from the health perspective");

- one possible subclass of action is the set of preparing activities for meals;
- possible subclasses of objects for actions are meals, nutrients and food.

A complete ontology will integrate in various ways an appropriate action/state ontology, and ontologies for basic notions like space, time, and so on. An example of an instance of "opinion" is given in Fig. 23.

6.3.5 Formal Competency Questions

The final task for the ontology development concerns the formal specification of the behaviour of the concepts in the ontology. First, this involves the translation of the informal competency question in a more formal language. For instance, we can produce something like the following, where parameters in the form $?x$ represent variables and parameters in the form $!X$ represent constants:

1. We can argue in favour/against of a topic if there exist a perspective from which the topic has a Good/Bad value:
   - $HasVal(?topic) \subset$ $Perspectives(?p), Opinion(?topic, ?p, !Good)$
   - $HasValNeg(?topic) \subset$ $Perspectives(?p), Opinion(?topic, ?p, !Bad)$
2. A state or an action leads to a state of affairs if the state or the action achieve, favour, maintain or strengthen the state:

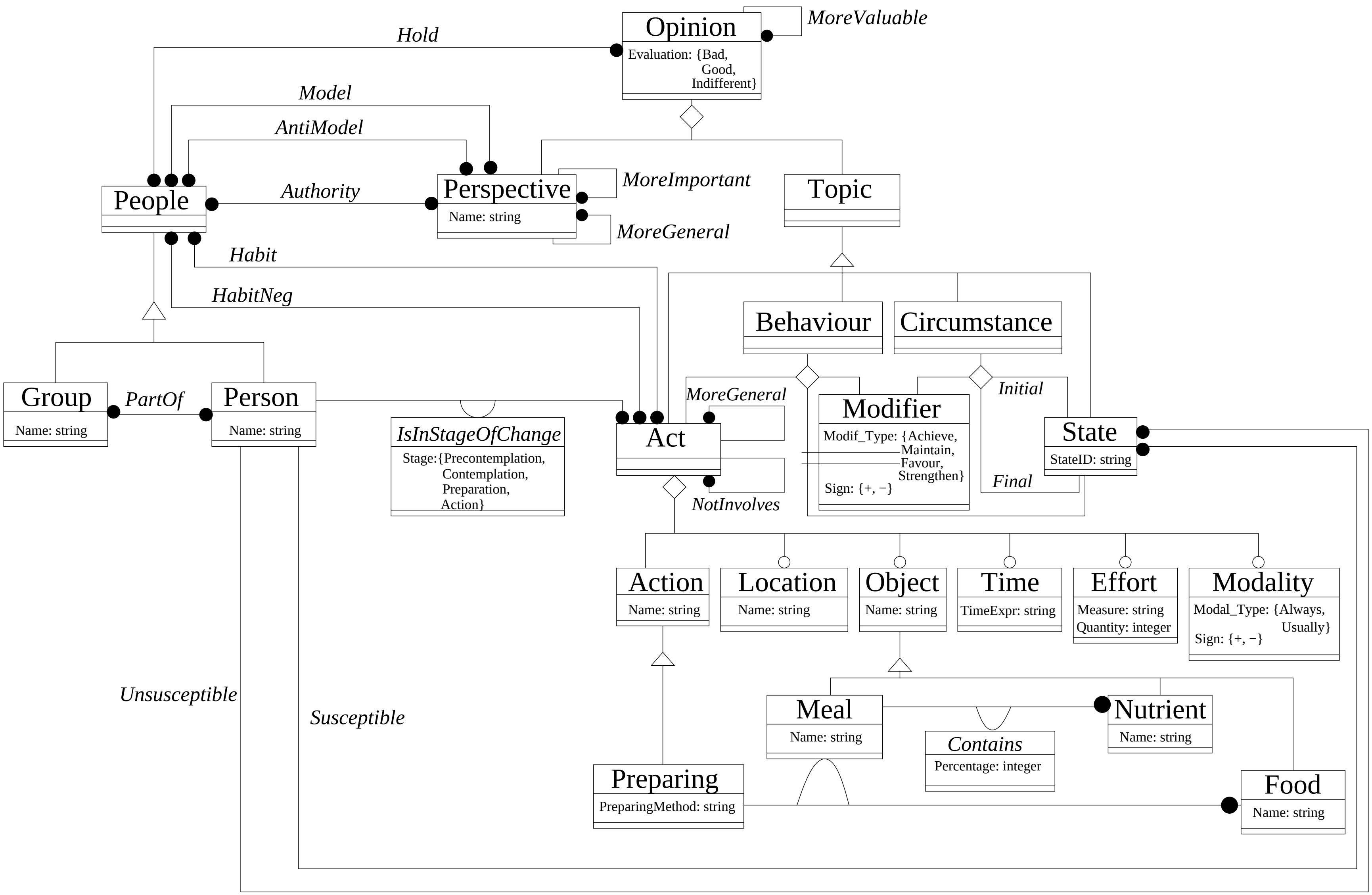
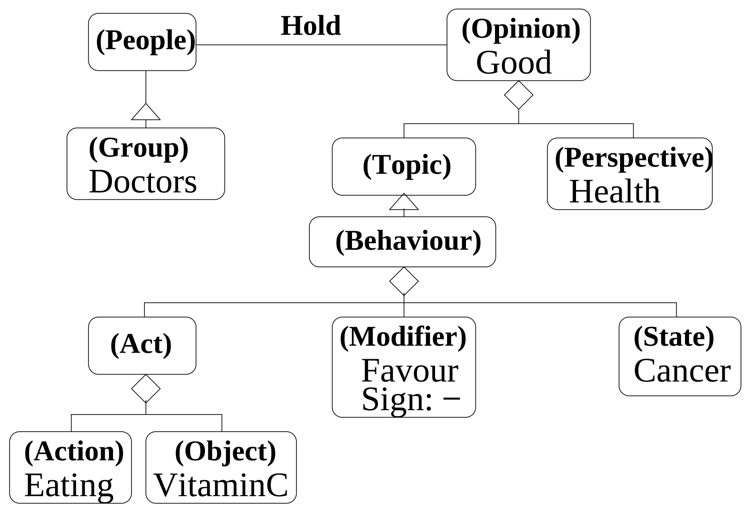


Figure 22: The Integrated Ontology



"Doctors think that the fact that eating Vitamin C can prevent cancer is good from the health perspective"

Figure 23: Instantiation of an Opinion.

- $ModifierPos(?x, ?y) \subset (Achieve(?x, ?y, +)$
- $ModifierPos(?x, ?y) \subset Favour(?x, ?y, +)$
- $ModifierPos(?x, ?y) \subset Maintain(?x, ?y, +)$
- $ModifierPos(?x, ?y) \subset Strengthen(?x, ?y, +)$

3. We can use an argumentation by authority in favour/against a topic if there exists an authority who has a good/bad opinion on the topic:

    - $HasAuth(?topic, ?perspective) \subset People(?authority),$
      $Authority(?authority, ?perspective),$
      $Hold(?authority, Opinion(?topic, ?perspective, !Good))$
    - $HasAuthNeg(?topic, ?perspective) \subset People(?authority),$
      $Authority(?authority, ?perspective),$
      $Hold(?authority, Opinion(?topic, ?perspective, !Bad))$

4. We can use an argumentation by model in favour/against an action if there exist a model who has/has not this action as a habit:

    - $HasModel(?action, ?perspective) \subset$
      $Action(?action), People(?model),$
      $Model(?model, ?perspective), Habit(?model, ?action)$

All other informal competency questions can be re-elaborated similarly, by establishing the appropriate relations among predicates.

6.3.6 Axioms Specification

Finally, the mechanisms should be put in place for the ontology to actually be able to address the competency questions, therefore to “reason” about the objects represented. This is done by defining axioms. From the various ICQs, we have seen that the main effort of a rhetorical arguer is to operate on the audience’s values and let them pass from one topic to another, in the hope to make the latter more acceptable. The central set of axioms we need, therefore, is aimed at establishing how such “value passing” occurs, for example:

- $HasVal(?topic, ?perspective) \subset$
  $HasAuth(?topic, ?perspective)$
- $HasVal(?state, ?perspective) \subset State(?state_2),$
  $LeadsTo(?state, ?state_2), HasVal(?state_2, ?perspective)$
- $HasVal(?action, ?perspective) \subset State(?state),$
  $Prevents(?action, ?state), HasValNeg(?state, ?perspective)$

To express the above axioms we also need a set of axioms to relate chains of states to a state or to an action, for example:

- $LeadsTo(?state_1, ?state_2) \subset$
  $ModifierPos(?state_1, ?state_2)$
- $LeadsTo(?state_1, ?state_2) \subset State(?state_3),$
  $ModifierPos(?state_1, ?state_3), LeadsTo(?state_3, ?state_2)$
- $Prevents(?state_1, ?state_2) \subset$
  $ModifierNeg(?state_1, ?state_2)$
- $Prevents(?state_1, ?state_2) \subset State(?state_3),$
  $ModifierPos(?state_1, ?state_3), Prevents(?state_3, ?state_2)$

These two sets of axioms enable us to create chains of arguments which can be presented to the advisee within one of the New Rhetoric schemata.

With the specification of all axioms needed to solve the formal competency questions, the ontology is ready to be deployed.

6.4 Discussion

The use of an explicit guideline, from an ontology conceptualisation methodology, proved to significantly help in focusing our effort in the construction of a well designed knowledge base for our domain. However, in some cases the methodology was not detailed enough to give sufficient advice. From the point of view of mere "users" of the methodology, we would have been helped, for example, by some more guidelines on how to collect and organise motivating scenarios, and how to use them to generate informal competency questions. In fact, we found that the early stages the methodology suggests, the ones we were most interested in, were not discussed in sufficient detail. We therefore had to define our own protocol for collecting and classifying informal competency questions, as no clue was provided by the methodology, and to decide a policy for establishing which aspects of the competency questions needed to be considered when defining the basic concepts of the ontology.

As pointed out by Jones, Bench-Capon, & Visser (1998), methodologies rarely address the problem of how to identify ontological categories[4]. We also suspect that such identification depends strongly on the theories that the ontology needs to account for: we wonder whether our conceptualisation would have been different had we used a different behavioural or argumentation theory, even with the same corpus of motivating scenarios as a starting point. This, we believe, does not depend so much on having a functional notion of knowledge, something that Guarino (1995) warns as not adhering to the principles of ontology engineering, but rather on seeing the "nature of the real world" from a different point of view. As our ontology needs to be used mainly in an advice giving environment, the purpose of our conceptualisation is to describe how people talk about things, rather than the things themselves. As Reiter, Robertson, & Osman (2000) maintain, it is difficult to formalise these notions because there is still a very poor understanding of how natural language really works.

We agree with Guarino (1995) that the most interesting issues are at the highest levels of the conceptualisation process, and that research on these aspects should benefit from the intervention of many and diverse disciplines. We have provided our own contribution to this issue.

6.5 Conclusions

The design of an effective advice giving system in a specific domain requires us to represent expertise coming from different fields, besides the application domain itself. We identified the argumentation and the behavioural science domains as crucial in the production of effective advice, as well as specific

[4]General, more philosophical issues of how to better define ontological categories have been addressed more recently by Guarino & Welty (2002) and Tamma & Bench-Capon (2002).

knowledge on the health and food domains. The problem of reconciling some very different theories in a unified knowledge framework suggested the use of a principled approach to knowledge engineering, and we drew on research on Ontologies for advice. The final product of the conceptualisation process, when compared to the formal definition of ontology given above (page 134), lacks of the property of being "shared": while the ontology is based on well established theories, it is still a private interpretation of these theories, and needs to be subjected to the evaluation of a larger community.

A predefined methodology was used to guide the conceptualisation process, and this exercise has been presented in this Chapter as a case study. Other case studies that demonstrate the application of the various methodologies for building ontologies are concerned with domain, or task ontologies. We have instead shown the application of one of these methodologies to a more subjective area, for the definition of concepts like "arguments", "values" or "opinions". We believe this was an interesting exploration of how the methodology could stand up to a very different application: there are few existing reports that test the reliability of an ontology engineering methodology by authors other than those who designed the methodology itself, and in domains that are very different than the one the methodology was designed for.

In addition to the production of a better crafted knowledge base, another important issue attracted our attention towards ontology research. The existence of so many advice giving or tutoring/educational systems in the medical domain[5] shows a clear need for a generalised ontology of health education concepts. Furthermore, notions related to the presentation of information, and subjective notions like "values", "opinions", "arguments" and so on, are widely used, more or less explicitly, in both dialogue and monologue generation systems[6]. Some concepts and relationships are surely shared by many of these systems, even if exploited differently, or represented differently, perhaps embedded in the behaviour of various components. An ontology engineering approach may help in isolating portions of knowledge that can be used in other contexts, therefore contributing to the more general effort of defining libraries of knowledge that can be semantically interoperable (Neches *et al.* 1991).

As Jones, Bench-Capon, & Visser (1998) observed, more of these efforts are needed, so that an ontology engineering practice may be established, based on the experience of as many projects as possible.

[5]See, among others: Buchanan *et al.* (1995), Binsted, Cawsey, & Jones (1995), Carenini, Mittal, & Moore (1994), Cawsey, Binsted, & Jones (1995), Hirst *et al.* (1997), Cawsey, Grasso, & Jones (1999), De Carolis *et al.* (1996), Miksch, Cheng, & Hayes-Roth (1996), Reiter & Osman (1997).

[6]For example: Cohen (1987), Elhadad (1995), Huang (1994), Karacapilidis & Papadias (1998), Marcu (1996b), Maybury (1993), Reed (1998b).

Chapter 7

Daphne: A Model of a Dialectic Adviser

Ah, lest some thorn shou'd pierce thy tender foot,
Or thou shou'dst fall in flying my pursuit!
To sharp uneven ways thy steps decline;
Abate thy speed, and I will bate of mine.

- Ovid, *Metamorphoses (Daphne and Apollo)*

This Chapter concludes the part of the thesis devoted to the specification of a practical implementation of the rhetorical framework defined so far. After taking inspiration from natural occurring dialogues in a nutrition promotion setting, and after having discussed the type of knowledge necessary to implement all the notions introduced in the first part of the thesis, we now concentrate on the architecture of a system for producing rhetorically driven advice on healthy nutrition. The aim of this Chapter is more to produce an initial demostration that the framework can be used to guide an example implementation, rather than to present a full scale system. Indeed, despite its position in the thesis, this Chapter does not chronologically follow the previous ones. The various parts of the architecture proposed here have been, in fact, prototyped in parallel with the development of the framework, an exercise which greately contributed to further refine the theory.

The Chapter starts by describing the proposed overall architecture, using the metaphor of an artificial agent. Then, it details the main components of the agent in turn.

7.1 The Architecture of an Advisory Agent

Daphne[1] is an artificial agent designed on the basis of the framework described in this thesis. An artificial

[1]Dialectical Argumentation for Providing Healthy Nutrition Education.

agent is a computer program that (Wooldridge 2000):

- can act "autonomously", that is the decisions to act are under its own control;
- is "proactive", that is it exhibits a goal-directed behaviour;
- is "reactive", that is its behaviour is responsive to changes in the environment;
- has some "social ability", that is it can interact (communicate, negotiate, collaborate, etc.) with other agents in the environment, each with its own agenda.

A basic agent could be summarised as the following loop (Wooldridge 2000, p. 26):

```
while true
    observe the world;
    update internal world model;
    deliberate about what intention to achieve next;
    use means-ends reasoning to get a plan for the intention;
    execute the plan
end while
```

As an agent, ***Daphne***'s main activity is to produce messages to satisfy a communicative goal: social ability is therefore the main *raison d'être* here. ***Daphne*** interacts with another (possibly human) agent. The observation of the behaviour of this interlocutor (i.e. other messages) will cause a revision of the beliefs about the environment. ***Daphne*** controls various strategies on how to accomplish goals, at various levels. The execution of the chosen plan will cause a reaction in the interlocutor (a response message) which lets the loop start again.

The architecture of ***Daphne*** is depicted in Fig. 24. The main components are:

- a **Dialogue unit**, responsible for managing the actual communication between the two parties;
- a **Reasoning unit**, responsible for managing the mental attitudes of the agent;
- a **Planning unit**, responsible for devising and executing the plan for action.

Notably, each component of the system can be in turn described as an autonomous agent: the components are independent on one another, they have their own strategies and "beliefs", and communicate with the other components (and with the external world where appropriate) to decide what to do. In fact, ***Daphne*** can best be seen as a *Multi-Agent* system (Weiß 1999). Let us examine in turn the behaviour of the various component-agents.

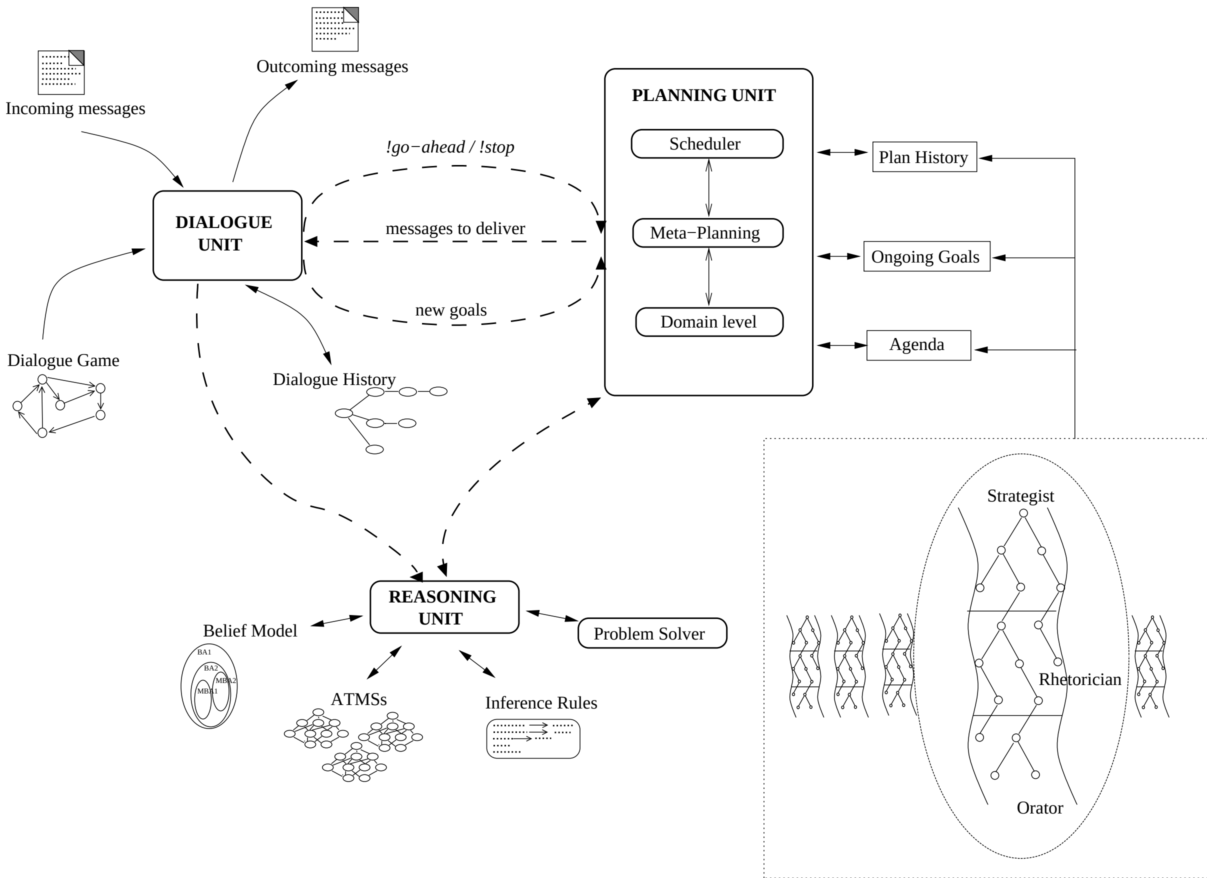


Figure 24: ***Daphne***'s architecture

7.2 Dialogue Unit

The Dialogue Unit is in charge of managing the conversation, therefore it basically deals with both incoming messages and the messages that are to be put forward. Some assumption can be made to make this task more manageable:

- *non ambiguity*: the two participants are assumed to agree on the meaning of the concepts being communicated;
- *limited set of utterances*: specific communication acts are defined, and utterances including other types of communication acts are not considered;
- *explicit back reference*: whenever an agent wants to refer back to a previous dialogue move (e.g. to answer a question or to express disagreement), this previous move has to be explicitly indicated.

These assumptions permit us to put aside, for the time being, the complex issues of dealing with natural language.

One of the tasks of the Dialogue Unit is to ensure that the dialogue overall coherency is preserved. This property is not judged on the *content* of the messages, but it is a syntactical property derived from the Dialogue Game Board (DGB). As explained in Sect. 3.3.3 (page 72), a DGB is a transition network in which each node represents the state the system is in when one agent just completed a dialogue move of a given type. From each state, an agent is allowed to perform a particular dialogue move if:

- an arc in the network connects the current state with the one representing the new move, and
- the player is allowed to traverse this arc.

Two types of arc distinguish between "coherent" moves, which follow naturally from the previous move, and "changing game" moves, which signal interruption to the normal flow, for instance when a question goes unanswered. To monitor this situation, the dialogue manager maintains a *dialogue history*, which keeps track of the references among moves, as well as the chronological succession of moves in the dialogue.

The behaviour of the Dialogue Unit is similar both when it is ***Daphne***'s turn to speak, and when it is the other party's, except that, in the latter case, if the Dialogue Unit decides that the move is not allowed, it will ask the contributing agent to reformulate it. It also updates the dialogue history. If no more moves are allowed to the current speaker, the Dialogue Unit warns the speaker to stop, and signals the other party to take the turn.

The Dialogue Agent's Cycle

The Dialogue Agent's cycle described above can be summarised as:

```
1. while true
2.     listen to the next message;
3.     update dialogue history;
4.     deliberate about whether the current speaker can hold the turn;
5.     broadcast the message to the other agents;
6. end while
```

7.3 Reasoning Unit

The Reasoning Unit deals with the agent's mental attitudes about the domain and the opponent. It consists of a problem solver, coupled with a reason maintenance system to help dealing with potentially withdrawable information.

The problem solver is based on a clausal form logic (Richards 1989), but restricted to *Horn clauses*. These are clauses formed by a conjunction of positive elements as premises, and either zero or one positive element as conclusion:

$$a_1 \wedge a_2 \wedge \ldots \wedge a_n \rightarrow b$$

Horn clauses have long been proved to make the problem solving process more efficient (Dowling & Gallier 1984). Moreover, it has been argued (by Taylor, Carletta, & Mellish (1996a) among others) that reasoning by means of Horn clauses is more "natural", as people are notoriously bad at dealing with disjunctions: the sentence "Either it rains or I go out" may be hard to grasp immediately.

As we have seen in Ch. 4, a reason maintenance mechanism is needed when one has to be able to deal with mental attitudes that can change over time. The property of a mental state to allow retraction of previously accepted attitudes is called *non monotonicity*. Non monotonicity can concern both time (e.g. what is believed at a certain interval of time can be disbelieved later on) and logic (inferences made with incomplete information, that is on the basis of *assumptions*, may need to be withdrawn when new evidence arises). ***Daphne*** accounts only for the logic type of non monotonicity: the time span of a conversation is not long enough to assume that, without new evidence, a belief held at the beginning of the dialogue ceases to be held later on[2].

The mechanism used to deal with reason maintenance is an Assumption-based Truth Maintenance System (ATMS) (de Kleer 1986a). An ATMS deals with the second type of non-monotonicity, helping in identifying at each stage all the different sets of consistent beliefs. Each set is a possible world, so in fact the ATMS helps in creating different contexts among which to choose. An ATMS is not a reasoning mechanism itself, but rather a "database" for problem solver data, keeping track of the links between beliefs and assumptions in a network structure.

The basic object in an ATMS network is a *node*, which is represented as the triple:

$$\langle datum, label, justification \rangle$$

where:

- the *datum* is just an identifying name for the problem solver datum;
- the *label* is a set of *environments*, each representing one context in which the datum holds.
- the *justification* shows how the problem solver derived the node from other nodes (the chain of

[2]Typical beliefs that can change with only time are volatile properties of concepts. For example, "being a teenager" will cease to be true after the appropriate period of time.

inferences used);

The *environment* is the key notion here. These are minimal sets of validating assumptions for the datum. In other words, each set is necessary and sufficient to make the datum true in that environment. In a way, the label represents how defeasible the datum is. For instance the node[3]:

$$\langle x, \{\{A, B\}, \{C\}\}, \{(b), (c, d)\}\}\rangle$$

represents the fact that datum x can be derived in two ways, either from b, or from $c \wedge d$. Furthermore, it holds in the environments containing either assumptions $\{A, B\}$ or assumption C.

Special nodes can be represented, such as *premises*, that is nodes that are always true, *contradictions*, that is nodes that are always false, or *assumptions*, that is nodes which "justify themselves". Assumptions are created, for instance, for defaults in the user's profile. New environments and new justifications can be added while reasoning progresses.

If the problem solver discovered a contradiction, the set of assumptions causing it is added to the justification set of a special node representing falsity. The list of inconsistent environments, that can be thought of as the label of datum $\perp$, together with the list of assumptions, can be used to determine all the possible *extensions*, that is the "possible worlds" that the ATMS describes.

Each time a new inference is made by the problem solver, the justification is passed to the ATMS, which updates the label for the datum corresponding to the inference's conclusion. A propagation algorithm is used to change labels of affected nodes, if appropriate, so that each label is sound, complete and minimal. This means that: each environment in the label does derive the datum; every environment which derives the datum is a superset of one of the environments in the label; and no environment in the label is superset of any other. These properties are achieved by removing from a label environments that turned out to be inconsistent, and environments that are subsumed by other environments. The latter step is what makes the ATMS efficient, as it exploits the property that if a datum is derived from a set of assumptions, it is derived from all the supersets of this set.

The specific algorithm for the implementation of label propagation used in ***Daphne*** is described by de Kleer (1988), and will not be detailed here[4]. ***Daphne***'s implementation of the ATMS differs from the standard one in two aspects. Firstly, there are no premises, as in ***Daphne***'s philosophy any datum is in principle falsifiable. Secondly, the assumptions are in fact considered as "defaults", that is rules of the type "In the absence of information about x, assume x". This is because we want to be able to

[3]In de Kleer's notation, we indicate assumption with upper case letters, and derived data with lower case ones.
[4]A sketch can be found in Appendix E.

"withdraw" x, as well as finding new justifications for it. This simply means that everytime ***Daphne*** decides to make an assumption a, a new dummy datum is created, A, and the justification $A \rightarrow a$ is added to the ATMS. Therefore, the only real assumptions in ATMS terms are in fact the dummy nodes, created by ***Daphne***.

7.3.1 Dealing with Negations

We have observed above that natural sentences are well represented as Horn clauses, as people do not use disjunctions comfortably. However, people do use negations quite extensively: the sentence "If it does not rain, I go out" is a more natural version, and an extensively occurring one, of the disjunction seen above. If the use of a Horn theory is justified by tractability, naturalness would justify the need for ***Daphne***'s interlocutor to express justifications like the one above, or in general like:

$$a_1 \wedge a_2 \wedge \ldots \wedge a_i \wedge \neg b_{i+1} \wedge \neg b_{i+2} \ldots \wedge \neg b_n \rightarrow c$$

We do not want the problem solver to actually reason with this type of clause, to maintain the tractability property, therefore we only allow clauses like the one above to be communicated to ***Daphne***. We need, therefore, a mechanism to include this information in the ATMS. The justification above is in fact equivalent to a non-Horn clause having a disjunction on the right end side of the implication:

$$a_1 \wedge a_2 \wedge \ldots a_i \rightarrow b_{i+1} \vee b_{i+2} \vee \ldots b_n \vee c$$

The extension to the ATMS proposed by de Kleer (1986b) suggests a way to deal with more general clauses by introducing disjunctions, or *Choices*, among assumptions. The list of Choices is used together with the list of contradictory environments to compute the possible extensions. Some "hyperresolution" rules ensure that, when calculating the extensions, only one element of each Choice is included.

7.3.2 Attitude Spaces

We have seen so far how the ATMS mechanism works on various data. We have not said, however, what kind of object a datum is. In Chapter 4 we have introduced a mental model $\mathcal{M}$ of the dialoguing agents, as a MultiContext environment:

$$\mathcal{M} = \langle B, R \rangle$$

where B is the set of all attitude contexts, and R the set of "bridge rules" that reason across attitude contexts. In ***Daphne***, one attitude context, $\beta \in B$, represents one mental attitude, such as "what ***Daphne*** believes the interlocutor wants", or "what ***Daphne*** wants the interlocutor believes about what is mutually believed ***Daphne*** believes", and the like.

An attitude context, we remind, is a triple:

$$\beta = \langle \lambda, \alpha, \varrho \rangle$$

where λ is a language, in our case in Horn-clausal form, α a set of initial axioms in λ and ϱ a set of inference rules internal to the context. In ***Daphne***, each attitude context is implemented as a separate ATMS, with associated set of initial axioms and rules. We can drop therefore the attitude specification in each separate context, so that for example, to store that "***Daphne*** believes that the advisee wants x", we simply store datum x in the ATMS labelled $b_{Daphne}w_{advisee}$ (where b represents "believes" and w "wants"). We have shown that, if n mental attitudes are considered, with the four types of belief nesting defined in Ch. 4, the number of attitude contexts that ***Daphne*** needs in order to interact with one advisee is: $n + n^2 + 2n^4$. We have therefore that many ATMSs. The problem solver reasons within a context by using the set of rules proper to the context, and the relevant ATMS stores the result of this reasoning process. The initial axioms, as mentioned above, are in fact initial defaults, implemented as assumptions, so they can be withdrawn if the situation arises.

Sometimes it is useful to reason across attitude spaces: for instance, one might want to be able to say that if Agent 1 believes that Agent 2 wants x, then Agent 1 also wants x[5]. As a MultiContext environment, in ***Daphne*** this can be done by means of "bridge rules". The conclusion of a bridge rule (in the example above, that "Agent 2 wants x") is stored in the correspondent attitude context ATMS ("what Agent 2 wants"), but the justification itself is of course not propagated in the usual way, because part of the premises will not belong to the context of the conclusion. It is, instead, recorded as an assumption. While this is not the perfect solution, it is practical enough to allow reasoning within one context to remain a black-box task.

7.3.3 Choosing among Extensions

The ATMS allows us to keep track of how many extensions, or possible worlds, there are to consider. Choosing which is the one to use at any time is done by means of rules which are external to the reasoning process. The choice is needed for instance to establish what ***Daphne*** believes, wants etc. *exactly*, whenever it has to be decided what to say next. We have seen in Ch. 4 different theories to deal with this problem. In particular, the approach used by Logan *et al.* (1994) can be appropriate to our scenario.

[5]This is the typical characterisation of the behaviour of a truly cooperative agent, see for example Castelfranchi (1998).

The approach proposes a combination of heuristics to decide which is the preferred extension. Sets are ordered: (i) according to the endorsement of the elements, (ii) according to the coherence of the set, that is by giving higher value to the sets providing the most proofs for "core" beliefs, and (iii) according to minimal change, preferring sets which are closer to the last preferred one. The three heuristics are evaluated in the above order, to choose the preferred extension. We are not however committed to one specific choice, and the modularity of the system allows experimenting with various theories, provided the appropriate way is used to label the agents' attitudes.

The Reasoning Agent's Cycle

The Reasoning Agent's cycle described above can be summarised as:

```
while true
    listen to request;
    case request is:
       store new justification: update ATMS;
       check datum: choose new preferred extension;
                    use ATMS to tell whether datum is true in the extension;
                    if unknown:
                        use clause reasoning to derive datum in the extension;
                        update ATMS;
                    endif;
    end case;
end while
```

7.4 Planning Unit

The conversation in ***Daphne*** is the result of a planning process, as in several other conversation systems (Allen *et al.* 1995; Carletta 1992; Cawsey 1992; Chu-Carroll & Carberry 1994; Moore & Paris 1993; Taylor 1994). Planners accept a library of *plan operators* and a representation of the domain as input, together with the task to be executed, and produce the plan to execute the task as output. A plan operator is typically defined by three objects:

Goal: the goal to achieve; such as a belief to induce, a topic to elaborate, or a dialogue move to utter;

Prerequisites: the conditions that have to be verified before the operator can be applied;

Decomposition: the description of how the goal is decomposed into subgoals.

The subgoals lead to the invocation of subtasks, which in turn might invoke other subtasks, and so on, till the final goal state is reached. The final output is a combination of all the individual goals achieved by the sub-tasks.

The problem description in our case naturally suggests a conceptual hierarchy among tasks, which can therefore be classified into *abstraction spaces*. Distinct sets of abstraction spaces are historically used (Sacerdoti 1974; Fox & Long 1995; Washington 1994) in order to better organise the solution to a problem: the problem is first completely solved in a high level space. Then the solution is refined at successively more detailed levels. Each abstraction space represents, therefore, a different type of knowledge, which gives a different perspective on the problem at hand. The description of the problem seen in the past Chapters, and especially in Ch. 5, suggest that ***Daphne***'s tasks can be seen from three abstraction spaces:

1. the higher level space, the *strategist*, is concerned with general goals of the system. In ***Daphne***'s case, these correspond to goals concerned with planning a general strategy for the promotion of a healthy behaviour, and embed knowledge coming from the behavioural model seen in Ch. 5. In other systems, this space will correspond to different strategies, for instance tutoring strategies (Freedman 1996);

2. the middle level space, the *rhetorician*, is concerned with planning good arguments to support specific claims, and embeds knowledge from the New Rhetoric, in the form of the schemata. Typically, one schemata, as described in Appendix A, will correspond to one operator at this level. This space, as observed when explaining the dialogue annotation scheme, is independent from the subject of the argumentation, so it is designed to work equally well with a different strategic environment;

3. the low level space, the *orator*, concentrates all the linguistic knowledge on how to express, in natural language, the argument proposed by the rhetorician. This level will know about the basic communicative acts that can be performed (the dialogue moves) and will also know of presentational strategies, like the ones suggested by the Rhetorical Structure Theory.

A further level can be thought of, which is not discussed here, in order to address the problem of choosing the sentence structure in the natural language to be used (words, style etc). Theories like the one by Anscombre & Ducrot (1983) can be of help.

These four levels can be accomodated in the standard three-stage pipeline model of Natural Language Generation systems (Reiter & Dale 2000)[6]:

- planning, where the general structure of the text is established;
- microplanning, where the sentences are decided upon;
- surface realisation, where the sentences are actually produced.

However, they are differently distributed. The *planning* phase correspond in fact to the succession of the strategist and the rhetorician levels. The strategist decides on the plan for the general aim of the discourse, and it does not decide on discourse related matters. Some authors put this "content selection" phase outside the planning process, and the phase per se is generally understudied. The rhetorician is in fact across a content selection and a planning phase, as, while defining the argument line, it both continues to decide on new topics to add to the discussion, and starts organising them in discourse structures. We have not dealt with the problem of natural language generation in this thesis, but at this level, in addition to merely argumentative schemata, one could include schemata built on the Rhetorical Structure Theory constructs we have seen previously, perhaps as a sublevel of the argumentative one. The orator is the level with the expertise on sentence organisation, the microplanning in the pipeline architecture, as from this level the decision comes of which sentence to put forward, and how much of it in one turn.

7.4.1 Planning Process

In each of the three levels of abstraction, the actual planning process is structured in three *planning layers* (Stefik 1981). At the bottom layer (the *domain space*) there are the actual planning operators related to the domain of the abstraction space currently under consideration. On top of this, there is a *meta-planning* layer, whose aim is to control the applicability of domain operators, by checking prerequisites and executing goal decomposition. At the top of the hierarchy, a *scheduler* layer is responsible for controlling the overall process, by activating new goals, choosing among alternative operators, and so on.

Goals are provided to the planner by different sources. The first, top level goal derives from the system's main dialogue aim, expressed by means of a strategic goal. In ***Daphne***'s case, it is typically concerned with moving the user one step forward in the Stages of Change chain. Goals are also set by

[6] The four levels nicely fit also in the four parts of the rhetorical discourse identified in the *Rhetorica ad Herennium* (see page 8): *inventio*, the collection of the arguments, *dispositio*, their arrangement, *elocutio*, their expression in the language, and *pronuntiatio*, their utterance.

the dialogue situation, to fulfill discourse obligations (e.g. answering a question). Finally, goals can be set by the planner itself, as a result of the decomposition of goals in subgoals.

At each planning step, a goal is selected from an *agenda*, and, unless it can be directly executed (that is, it is a primitive action), an attempt is made to plan for it. The appropriate operator library is considered in order to find an operator whose goal can match the goal at hand. At the moment there is no specific search strategy, and operators matching the goals are attempted in sequence. If the operator's prerequisites are not satisfied, the planner may try to plan for them as well, or may decide to discard the operator altogether. The prerequisites can have three levels of stringency:

1. *mandatory* prerequisites have to be true for the operator to be considered in the first place; for instance, the ontological constraints of a rhetorical schema could be at this level;

2. *necessary* prerequisites can be false, but the system must try to satisfy them by planning; for instance, the acceptability, or the relevance constraints of a rhetorical schema could be at this level;

3. *desirable* prerequisite can be false, and the system may decide to ignore them and use the operator anyway; for instance, the sufficiency constraints of a rhetorical schema could be at this level.

The level of stringency of a constraint is not a property of the operator, though, but rather of the system's "attitude" towards planning: an over cautious planning will require constraints of all types to be satisfied prior the application of the operator; a more daring one could be more lenient. Also, the way in which the constraints are labelled can change according to the system's attitude: for instance, as seen in Ch, 2 when describing the admissibility constraints of rhetorical schemata, only a completely honest argumentative attitude will require constraints of all types to be at least mandatory.

If the prerequisites are attempted as goals, they will be put in the agenda, together with the subgoals in the decomposition slot, and the process will start again.

7.4.2 Planning Decisions

The planner has to perform a choice in two occasions: it has to choose the goal to accomplish when several goals are present in the agenda at the same time, and it has to choose the operator to apply when several operators are equally suitable to accomplish a goal.

Goals are kept in the agenda partially ordered with respect to several different priority scales. Many scales can be defined: for instance, the goals proposed to fulfill the system's high level aims might have a high value with respect to a *strategy* scale. The goals set to fulfil dialogue obligations might have, on the other hand, a high value with respect to a *focus of attention* scale, and so on. The choice of which goal to activate first is made by means of metarules, and depends on how the system prioritises the different scales. We can then run ***Daphne*** with a higher emphasis on the strategy or we can choose to be more or less cooperative, with respect to the discourse obligations, by preferring or otherwise the *focus* scale. This mechanism allows the presence of any number of scales, as long as the appropriate metarules exist to evaluate new goals.

If several operators are available to solve a goal, other metarules are used to choose among them. The set of metarules is specific to the operator library. This is reasonable if it is accepted that three different theories are expressed in the three libraries: the parameters influencing the choice of a strategy in the Stages of Change theory are likely to be different than those influencing the choice of an argumentative schema, or a linguistic construct. General heuristics may concern the reliability of the operator (how many of the desirable constraints are met), the novelty of the operator (especially at the rhetorician level one would not want to use the same schema too often), or a measure of its complexity, based on the subgoals, and so on.

The planning process is incremental, as execution and planning are interleaved. This approach, first envisaged by Sacerdoti (1974) and thoroughly developed by Ambros Ingerson (1987) for general planners, is particularly useful in the planning of dialogue moves, where one may want to wait for the interlocutor's reaction before saying too much (Cawsey 1992; Freedman 1996). Moreover, speaking without planning too much ahead makes the system's behaviour more natural and humanlike. In particular, as soon as the first dialogue move can be uttered, the planner "proposes" it to the dialogue unit, and stops to listen whether the dialogue game board authorises some new dialogue moves to be performed, rather than passing the turn to the interlocutor.

After an interruption, it may happen that discourse obligations impose the start of a new thread of conversation, or resume an old one. The system will decide to accommodate new goals according to the goal selection rules seen above, and some garbage collection will be done in the agenda to eliminate goals that are not going to be pursued any longer.

The Planning Agent's Cycle

The Planning Agent's cycle described above can be summarised as:

```
1. while true
2.     listen to the next message;
3.     update agenda with new discourse goals;
4.     repeat
5.         choose one goal from agenda;
```

```
        while goal is not a dialogue move:
            repeat:
                choose one operator from correspondent library;
                ask reasoning agent to verify constraint;
            until applicable operator is found;
            update agenda with new sub-goals;
            choose one goal from agenda;
        end while;
        broadcast the dialogue move to the other agents;
    until dialogue agent requests to stop;
end while
```

7.5 An Example

As an example of the envisaged behaviour of ***Daphne***, let us consider the following short dialogue:

1-*Daphne*:	*Have you considered eating more fruit?*
2-User:	*No.*
3-*Daphne*:	*It is good for your health, as it helps prevent cancer.*
4-User:	*Fruit is boring! It's granny's stuff!*
5-*Daphne*:	*Avocados are not boring! Did you know they are Californians' favourite?*

Let us analyse the dialogue a turn at a time.

1-*Daphne*

The first of ***Daphne***'s turns is a result of the *strategist* phase, involving the assessment of the user's stage of change with respect to some "healthy nutrition" action, such as, in this example, the action of eating more fruit. This is done by means of an enquiry about the attitude towards the action, rather than an argumentative discourse. Therefore the *rhetorician* is not actively involved in this stage: it receives the goal, but, as the goal qualifies as a "basic act" for this abstraction level, it is passed directly to the next level. The goal is achieved by the *orator* by means of a Yes/No question, so, according to the dialogue game board in Fig. 5 (page 76), ***Daphne*** is warned by the Dialogue Manager to stop talking and wait for the User's reply.

2-User

The user's reply is a straightforward "No", and follows the dialogue game rules. The reply tells ***Daphne*** that the user is in the precontemplation stage with respect to the activity to promote.

3-*Daphne*

After obtaining the information needed, the actual argumentation begins. The first strategic goal for ***Daphne*** is to let the user move one step forward in the States of Change chain. Here, ***Daphne*** is trying to move the user from the precontemplation to the contemplation stage with respect to eating fruit. There are several ways to pursue this goal, dictated by the Stages of Change model, such as describing the benefits of the action, warning about possible consequence of not performing the action, and so on. Suppose the choice at the strategic level is to describe the benefits of the action, that is to "claim that the action is valuable". As the goal involves rhetorical abilities, the control is passed to the relevant planning abstraction layer, the *rhetorician*.

The *rhetorician* has to choose one of the rhetorical schemata for promoting an action. Suppose the chosen schema is the "pragmatic argumentation". Pragmatic argumentation can exploits several consequences of one action. For example, ***Daphne*** can argue in favour of eating fruit by emphasizing that this is useful to prevent cancer. This can be done if it is assumed that the user will give a bad value to cancer, for instance from the health perspective.

Suppose now that the subgoals of the pragmatic argumentation schema involve claiming that the action is valuable, and describing the positive consequence it has. These goals refer to the orator abstraction level, which in particular will have the linguistic knowledge needed to produce a list of consequences of an action. When the speech acts are produced, the planner stops its activity and the acts passed to the dialogue manager, which communicates them to the user. This happens one act at a time, in order for the dialogue manager to check whether, according to the dialogue game board (Fig. 5), another move is allowed to ***Daphne***. If dialogue manager communicates to the planner that it can carry on with its activity, subsequent speech acts are produced, or, if planner has no other goal to achieve, the turn is passed to the user.

4-User

The user makes two assertions, and the communicated messages are:

- Fruit is boring (because)
- Grannies eat fruit.

We assume in this example that ***Daphne*** considers the user truthful[7]. These two beliefs of the user are then recorded both in the mutual section and the private section of ***Daphne***'s beliefs about the user:

1. ***Daphne*** believes User believes "Fruit is boring";
2. ***Daphne*** believes User believes "Grannies eat fruit";
3. ***Daphne*** believes User believes it is mutually believed User believes "Fruit is boring";
4. ***Daphne*** believes User believes it is mutually believed User believes "Grannies eat fruit".

As the whole argumentation is based on values rather than facts, ***Daphne*** is primarily interested in ascribing an evaluation to what the user says. Therefore, the reasoning module tries to make some inferences with the new pieces information to this aim. This could be done by means of a set of rules which ascribe beliefs to the user. The resulting new beliefs, with their justifications, are recorded in the ATMS.

In the example, with the assumption that boredom is not a positive emotion, the system can infer that the same value is passed to Fruit; it can also be inferred that in, the user's hierarchy, the Social Life perspective has a higher position than the Health one; moreover, as the user said that eating Fruit is a Granny's habit, it may be inferred that Granny is an anti-model for the user from the Social Life perspective. As the inferences are done by the system, they are considered "presumptions" and are stored in its private belief section about the user:

6. ***Daphne*** believes the User believes "Fruit is not valuable from the Social Life perspective";
7. ***Daphne*** believes the User believes "the Social Life perspective is more important than the Health one";
8. ***Daphne*** believes the User believes "Grannies are anti-models from the Social Life perspective".

5-*Daphne*

At the beginning of the new turn, ***Daphne*** realises the user has not still passed to the contemplation stage, as there still are no "good" values attributed to the action of eating Fruit. The strategic goal is therefore still to be achieved, and it is posed again to the planner.

Daphne has to produce a counterargument, and can focus on the sixth belief in the list above:

- ***Daphne*** believes the User believes "Fruit is not valuable from the Social Life perspective"

[7] If this was not the case, the messages would be stored differently, but the behaviour of the system would have probably been the same, in order to exploit what the opponent says.

There are two ways ***Daphne*** can attack a claim about the attribution of a bad value with respect to a perspective:

1. by arguing that the topic has good value with respect to the same perspective;
2. by arguing that the topic has good value with respect to another perspective;

Suppose the first of the two is chosen. ***Daphne*** needs to find a way to be able to say that "fruit has value from the social life perspective". If this cannot be done directly, because there are no arguments "based on the structure of the reality" to be used, ***Daphne*** may look for arguments "establishing the structure of reality". One of these is the "argumentation by example". This style of argumentation creates an evaluation by generalising from a specific case. So ***Daphne*** has to find a type of fruit that can have good value from the Social Life perspective. The schema library is again analysed, and an argumentation by model is chosen, that exploits the fact that Avocados are a habit of Californians. Of course this presupposes that the fact that Californians are a model from the Social Life perspective should be accepted by the user. The two argumentation schemata may therefore be nested: avocados are eaten by Californians; Californians are "good" from the Social Life perspective, and so avocados are good as well from the same perspective; avocados are an example of fruit, so fruit in general is good from the Social Life perspective.

7.6 Discussion

We have presented in this Chapter the architecture of a system which embodies the various aspects of the rhetorical argumentation framework we have seen in this thesis. The architecture has its core in a planning agent, with a hierarchical structure.

Various systems have used similar hierarchical planning frameworks for discourse planning and interpretation. Reed, Long, & Fox (1996), for instance, decompose the task of planning argumentative texts in three subtasks:

- choosing the argument structure (AS level);
- refining the argument by employing stylistic techniques (EG, or Eloquence Generator level);
- refining the interclausal structure, through the use of the Rhetorical Structure Theory.

Our proposal isolates all the knowledge about argumentation techniques and tactics (Reed, Long, & Fox' AS level and part of the EG level) in the rhetorician component, and builds on top of this a strategic component whose aim is to decide "why argue in the first place", by establishing the purposes of the argumentation on the basis of behavioural models of change. All the linguistic knowledge (EG and RST levels) is then concentrated in the orator.

Another similar architecture, but with the purpose of discourse understanding rather than generation, is due to Lambert & Carberry (1991). The authors propose a tripartite model of dialogue, in which three types of action are distinguished:

- domain actions, related to the goals of the participants;
- problem-solving actions, related to the planning process itself (e.g. evaluating alternatives);
- discourse actions, related to conveying or obtaining information needed by other levels' actions.

Our conceptual hierarchy can be seen as a finer grained representation of Lambert & Carberry's domain actions, in which also the discourse level actions are included (the "orator"), whereas all the knowledge concerning planning and problem solving is separated from the application domain and is concentrated in the meta-planning and scheduler layer, in accordance with Stefik's view of the planning process.

One aspect of the system we have not addressed at all is the ability to understand the arguments of the opponent. The communication from ***Daphne***'s interlocutor is used to reason about mental attitudes, but there is no thought on how "plan recognition" (Carberry 1990) can take place. Most importantly, for the argumentative nature of the dialogues, the system should be able to understand which argumentative schema the interlocutor is in turn using, and think of effective moves in reply, perhaps substantiated by a sophisticated user model (Zukerman & Litman 2001). Some suggestions on possible "counter-schemata" can be found in the New Rhetoric, but the formalisation of these suggestions in an effective strategy requires a thorough study that we have to leave to further work.

Daphne, as it stands, is not therefore a full scale, working system, nor has it ever meant to be. As we mentioned at the beginning of this Chapter, what is elaborated here is not the final product of the development of the framework, but rather the other way round. While developing the framework, one natural, test of its practicability was the attempt at implementing parts of it[8]. While implementing, new problems would manifest themselves, whose solution required a new design choice for the framework. In particular, the running of the second e-mail experiment described in Chapter 5 involved extensive evolving prototyping: each reply of the "adviser" was crafted as if produced by the system, by running whatever available, augmenting whatever was needed, and assuming whatever was not implemented (most notably, the natural language front-end interface). The modules that have been implemented are:

[8]The implementation has been undertaken in the ML language (Paulson 1991). The choice of language was merely dictated from the fact that, as a truly functional language, and therefore inherently elegant, ML lends itself very nicely to work as a pseudo-code language, which can at the same time be made to run.

the ATMS manager, the dialogue game board manager, the Horn clauses problem solver, the planner. The communication among the modules is not automated, nor it has been implemented a satisfactory strategy for decision making in various aspects of the system (for instance, the planner simply chooses the first applicable operator, rather than evaluating the relevant ones according to given parameters). None of the implemented parts presented a particularly interesting or new feature, when compared to other implementations of the same problems. It was therefore thought of little interest to detail the process here, as these are well studied problems, whose standard algorithms and complexity have been long known. Indeed, off-the-shelf software could have been used for all these parts, and would have been if the purpose of the implementation was to obtain a working system.

7.7 Conclusions

In a provocative, humorous, and still very spot on essay, McDermott (1981) harshly criticises one attitude of many AI researchers, that he calls of the “wishful control structure”. He refers to the practice of creating, in order to test a theoretical framework, an under-specified, under-equipped, under-achieving “Version I” of a program, identifying its shortcomings, envisioning their solution in a forthcoming “Version II”, and taking this as completing the task of proving that the theory was well formed. We tried not to fall ourselves in this trap by not even attempting to propose a “Version I” of ***Daphne***. What we proposed is an architecture, which we believe can be implemented. Our proposal by no means implies that an implementation would be uninteresting: on the contrary, we believe that a well documented implementation and evaluation of a system like ***Daphne*** is a PhD thesis in its own right. We also believe that the only way to test whether a theory is really meaningful is, to us, its implementation from someone *other* than its author, as a proof of the theory’s reliability. The program should live a separate life from the theory, for, if we subscribe to the criticism above, we also subscribe to McDermott’s warning that it is not necessarily true that “because *you* can see your way through a problem space, your program can”.

Chapter 8

Paralipomena and Final Remarks

> [...] in this pursuit, whether we take or whether we lose our game, the chase is certainly of service.
>
> - Edmund Burke, *A Philosophical Enquiry Into the Origin of Our ideas of the Sublime and Beautiful*

This thesis has reported on an investigation of how rhetorical argumentation can be completely characterised and formalised. By rhetorical argumentation it is meant argumentation which is both heavily dependent on the audience it is addressed to, and is concerned with evaluative judgements, rather than logical proofs.

The contributions of this work can be summarised as follows:

1. a framework for rhetorical arguments, presented in Chapter 2.

   The framework directly implements a philosophical theory, the *New Rhetoric*, and it is the first of this kind. Based on the notion of Rhetorical Schema, and inspired by insights from Rhetorical Structure Theory, the framework precisely articulates the constraints that define the admissibility of argumentation. These constraints are themselves a direct implementation of insights from Informal Logic theories.

2. a dialogue game framework, presented in Chapter 3.

   The framework, inspired by both theoretical and computational works, defines a dialogue game that captures the structure of rhetorical argumentation. A set of moves is defined for the game, and formalised in terms of how they change the commitment store of the participants. The notion of "dialogue game board" provides a general structure for representing and managing not only the proposed game, but also general, similarly defined games.

3. a mental model, presented in Chapter 4.

   The model expands the classical Belief/Desire/Intention model of the mental state, popular in agent oriented computing, and considers general mental attitudes. The model also provides a structure for representing mutual attitudes, that elaborates on the New Rhetoric notions of "facts" and "presumptions" by clearly distinguishing between communication and assumptions. This structure naturally fits the hypotheses on the commitment store update defined in the dialogue game.

4. an annotation scheme for rhetorical dialogues, presented in Chapter 5.

   The scheme addresses several desiderata in annotating rhetorical dialogues, as it not only captures the unfolding of the dialogue game moves, as defined in the dialogue game structure, but also the argumentative aspect of the dialogue, and the way in which the arguments are used by the participants to satisfy higher level goals. The scheme has been applied to a corpus of naturally occurring rhetorical dialogues, collected for the purpose.

5. an upper level ontology model for argument and behaviours, presented in Chapter 6.

   The ontology design has followed an established engineering methodology, which has been applied in a novel way to capture knowledge which is not task/domain oriented, such as the definition of subjective notions like values or opinions.

6. a high level design of the architecture of a rhetorical agent, presented in Chapter 7.

   The architecture harmonises all aspects of the rhetorical argumentation framework defined, and provides a valuable basis for a possible implementation of such a system.

The problem addressed in this thesis, is, not surprisingly, far from being completely solved, and in particular there are several important issues that have not been addressed in this work.

First and foremost, there is no explicit account of the importance of the communication channels by means of which the argumentation can be presented. Natural language is the most obvious vehicle to convey rhetorical arguments, and a wide plethora of issues like style, use of figures of speech, presentation order, emphasis, will play a crucial role on the degree of acceptability of the argument itself. No treatment of these has been attempted in this work, although some hints were given that theories like the one by Anscombre & Ducrot (1983) could be useful in this respect. But natural language is not the *only* communication channel: the metaphor of the advertisement world that we used in the introductory notes teaches us that the rhetoric of images can be very powerful indeed. In fact, philosophers are well aware of these non linguistic devices (Blair 1996; Groarke 1996), and their findings will be valuable in this respect.

The production and presentation of argument is, however, only one half of the communicative process. A second issue that was not considered in this work concerns the *understanding* of rhetorical argumentation. By this we mean the definition of mechanisms that can enable an agent to recognise that a particular rhetorical device has been used by the "opponent", and react to it with strategies for counter-moves. This is a well studied problem in formal argumentation frameworks (Dung 1995), but is at its infancy in naturally occurring dialogues, and its solution is bound to be delayed until the broader problem of understanding natural language, as posed by Hume (1758), has been more completely solved. A restricted version of this problem, however, can be tackled, by making use of insights from plan recognition research (Carberry 1990), while strategies for counter-moves can be extrapolated, once again, from philosophical works (the New Rhetoric provides useful hints here as well).

Another important issue that needs in depth investigation is the one of personalisation. This is especially important in view of a possible full scale implementation of the arguing agent described here. While general rhetorical strategies are well defined and understood, it is less clear how these strategies should be accommodated when used by, or addressed to, an individual. This equates to explaining how specific mental attitudes makes people more or less susceptible to a given rhetorical strategy, or perspective, or how a given style of dialogue (e.g. a Socratic one) can be more or less effective given different addressees. This problem is also an ontological one: not every behaviour, or action, or state of affairs is appropriate to be, for instance, promoted by means of the same strategy, or from the same perspective on the world.

Methodologically speaking, the major unaddressed issue in this thesis is evaluation. Informal considerations on the suitability of the framework have been provided, grounded on the theoretical works that underpin the framework itself. A more concrete evaluation has been supported by the e-mail experiment conducted to collect naturally occurring dialogues. However, a thorough test for the framework is needed, carried out, if possible, by different researchers, to ensure an unbiased report. And, more importantly, the construction of the "rhetorical agent", and its deployment in a specific domain, with real users, would provide specific insights on the model plausibility and feasibility in implementation.

Despite all these issues left open, we believe the work presented in this thesis has been successful in approaching the problem of creating a formal framework to represent rhetorical argumentation by using insights coming from consolidated theories of argumentation from philosophy. We believe this approach is a winning one: the philosophy or argument can count on centuries of in depth investigations on several aspects of the issue at hand, that cannot and should not be ignored by the artificial intelligence modeller. This view is not shared by all, however. It has been observed (Zukerman, McConachy, & Korb 2000) that works from "hypothetical reasoning", despite providing "theoretical insights to the fields of dialectics", do not “present implementable computational mechanism.” A similar comment is found in (Maybury 1993): “While these ideas are suggestive, they are not formalised precisely enough to form the basis for a computational theory”. Marcu (1996b), we have seen already, is also sceptical as to the possibility of adequately accommodating “conceptual structures that are proposed by research on argumentation” in computational systems. We obviously disagree with such a tendency to dismiss work in argumentation theory as a mere source of vague hints. On the contrary, we think it is precisely the task of the artificial intelligence researcher to look for computational scope, and provide suitable computational models, of the ones from other disciplines, while preserving their theoretical strength. We have tried to do exactly this in our work, and we content ourselves to be not alone in this view: recent events, such as the Symposium on Argument and Computation (Reed & Norman 2000), or the series of workshops on Computational Models of Natural Argument (Carenini, Grasso, & Reed 2002; Reed, Grasso, & Carenini 2003), have proved that times are ripe for argumentation theory and computation to finally live long together and prosper.

I have done.

You have heard me.

The facts are before you.

I ask for your judgement.

[Aristotle - Rhetoric, XIX]

APPENDICES

Appendix A

The New Rhetoric Schemata Formalised

We present here a collection of New Rhetoric (NR) schemata which were classified and formalised in this thesis' work. The set below does not include *all* schemata introduced by Perelman and Olbrechts-Tyteca: in particular, the ones for which the emphasis was on the linguistic facets of the argument, rather than on the ontological properties of the domain, have not been considered in this work.

A.1 Quasi logical

These arguments claim acceptability for being similar to those used in formal reasoning, such as the logical or mathematical. By doing so, "one takes advantage of the prestige of rigorous thought" (NR § 45). They may exploit one of several formal relations or properties.

A.1.1 Subset

These schemata exploit the set relation ($A \subset B$). They may either emphasise the notion of inclusion in a set, or the notion of division of something into parts.

Inclusion

Schema 1 (Parts of the Whole): The part of a set should be given the same evaluation of the set as a whole.
EXAMPLE: Bananas are fruit, so they are good for your health.
ONTOLOGICAL CONSTRAINTS: The element to evaluate is part of a bigger set.
ACCEPTABILITY CONSTRAINTS: The audience evaluates positively/negatively the set.
RELEVANCE CONSTRAINTS: The audience is aware of the set inclusion relationship.
SUFFICIENCY CONSTRAINTS: The element to evaluate should possess all the prototypical characteristics of the set, and the elements in the set have to be homogeneous.

Schema 2 (Whole into Parts): The whole should be given the same evaluation of its parts.
EXAMPLE: Of course you like fruit: don't you like bananas?
ONTOLOGICAL CONSTRAINTS: The element to evaluate is partitioned into subsets.
ACCEPTABILITY CONSTRAINTS: The audience evaluates positively/negatively the part.
RELEVANCE CONSTRAINTS: The audience is aware of the set inclusion relationship.
SUFFICIENCY CONSTRAINTS: The set should possess prototypical characteristics that are inherited by all the parts, and the elements in the set have to be homogeneous.

Division

Schema 3 (Partition): One of the partitions must apply.
EXAMPLE: If you don't like vegetables and pulses, you must get your Vitamin C from fruit!
ONTOLOGICAL CONSTRAINTS: A state of affairs (eating Vitamin C) is completely partitioned into non-overlapping sets.
ACCEPTABILITY CONSTRAINTS: The state of affairs is valuable.
RELEVANCE CONSTRAINTS: The partitions are known to the audience.
SUFFICIENCY CONSTRAINTS: The partitioning is complete, and non overlapping. The state of affairs is necessary.

Schema 4 (Species): Similar to the Partition schema, here the elements of the set are subdivided in types, or species, and one of the species must apply. In this case the division is natural and accepted, so the arguer does not need to enumerate all the parts.
EXAMPLE: Have a glass of freshly squeezed orange juice, if you don't want to eat fruit.
ONTOLOGICAL CONSTRAINTS: There is a "is-a" relationship between the object to value and another object (the specie).
ACCEPTABILITY CONSTRAINTS: The specie has valuable.
RELEVANCE CONSTRAINTS: The object belongs to the specie.
SUFFICIENCY CONSTRAINTS: One of the elements of the specie must apply.

Schema 5 (Dilemma): Two opposite hypotheses are in fact equivalent as choosing any of them make no difference to the conclusion, which is inevitable.
EXAMPLE: You either avoid junk food, thus making your friend tease you for being picky, or eat junk food, thus making your friend tease you for being fat!
ONTOLOGICAL CONSTRAINTS: The two actions/states of affairs are incompatible. The respective conclusions hold.
ACCEPTABILITY CONSTRAINTS: The conclusion is valued/devalued by the audience, together with one of the two actions (the other action will be the focus of the claim).
RELEVANCE CONSTRAINTS: Everything must be in focus (in fact, a short form of this argument would be less effective).
SUFFICIENCY CONSTRAINTS: One of the two actions/states of affairs must take place.

A.1.2 Contradiction

Schema 6 (Incompatibility): A choice must be made between two states of affairs.
EXAMPLE: You cannot eat chips all the time and stay slim.
ONTOLOGICAL CONSTRAINTS: An action or a state (1) leads to a state (2).
ACCEPTABILITY CONSTRAINTS: The audience accepts that (1) leads to (2), and value positively/negatively state not-(2).
RELEVANCE CONSTRAINTS: The audience has in focus that (1) leads to (2).
SUFFICIENCY CONSTRAINTS: One of the two states must be applicable. It should not be possible to compromise between (1) and (2).

A.1.3 Identity

Schema 7 (Definition): Give a definition of the action/state to promote/denigrate.
EXAMPLE: "Being overweight means that your body weight is greater than the norm".
ONTOLOGICAL CONSTRAINTS: Any "definition" for the concept to evaluate has to be in the ontology (e.g. in form of a "is-a" link, or as a more complex series of concepts, or the etymology of the word denoting the concept in the language of use).

ACCEPTABILITY CONSTRAINTS: The audience has a bad/good evaluation of the concept defining the concept to evaluate (e.g. disease)
RELEVANCE CONSTRAINTS: none
SUFFICIENCY CONSTRAINTS: There is no other definition of the concept with the opposite evaluation (e.g. Being overweight also means that you have a curvy figure).

Schema 8 (Identification): Identify the action/state to promote/denigrate.
EXAMPLE: "Frying is the type of cooking that involves the use of hot fat like oil or butter".
ONTOLOGICAL CONSTRAINTS: The schema uses the identification of the concept in the knowledge base, therefore the only constraint is that such concept "exists" in the ontology.
ACCEPTABILITY CONSTRAINTS: The audience has a bad/good evaluation some components of the identification (e.g. "fat is bad")
RELEVANCE CONSTRAINTS: none
SUFFICIENCY CONSTRAINTS: All the components of the identifications have the same evaluation (e.g. "olive oil is good", or "cooking is good").

Schema 9 (Justice): Two elements which are (appears to be) identical require identical treatment.
EXAMPLE: "If frying and boiling are two ways of cooking that can be done at home, they both should be preferred to eating out".
ONTOLOGICAL CONSTRAINTS: The two elements are related by a definition or an identification relation (see schemata above).
ACCEPTABILITY CONSTRAINTS: The audience has a bad/good evaluation of one of the two elements.
RELEVANCE CONSTRAINTS: The audience is aware that the definition/identification of the two elements is identical/similar.
SUFFICIENCY CONSTRAINTS: The identification/definition has to be as much similar as possible (in the example, the fact that boiling does not involve the use of fat makes the argument weak, as the two definitions are not close enough).

Schema 10 (Tautology): X is X
EXAMPLE: health is health!
ONTOLOGICAL CONSTRAINTS: none
ACCEPTABILITY CONSTRAINTS: the audience already values the concept in the same orientation the speaker wants (it's a reinforcement schema, and serves to establish hierarchies of values).
RELEVANCE CONSTRAINTS: none
SUFFICIENCY CONSTRAINTS: none

A.1.4 Symmetry

Schema 11 (Reciprocity): Two situations should be given the same evaluation on the grounds of some symmetrical property that they share (NR § 53).
EXAMPLE Since you like biscuits, you should like apples too, as apples are as crunchy as biscuits.
ONTOLOGICAL CONSTRAINTS: A relation of reciprocity can be defined to hold between two concepts when for each of them an attribute can be identified having same semantics and same value.
ACCEPTABILITY CONSTRAINTS: The audience should share the view that the "reciprocal" concept used in the argument has the desired evaluation, from a certain perspective.
RELEVANCE CONSTRAINTS: The ontological and the acceptability constraints are both in the focus of attention.
SUFFICIENCY CONSTRAINTS: Reinforce the symmetry: other pairs of attributes of the two concepts in question can be identified sharing the same semantics and the same value.

Schema 12 (Complementariness): Two actions, A and B should be given the same evaluation because they are both a necessary condition to bring about a certain effect.

EXAMPLE You cannot concentrate on eating less alone, without also making some exercise.
ONTOLOGICAL CONSTRAINTS: A and B are actions, and together lead to the same state of affairs S.
ACCEPTABILITY CONSTRAINTS: The audience evaluate positively/negatively S and one of the two actions (say A).
RELEVANCE CONSTRAINTS: The audience is aware than A leads to S.
SUFFICIENCY CONSTRAINTS: A and B are in fact both needed (the desired effect should not be obtained by means A alone).

Schema 13 (Contraries): Two states of affairs, A and B should be given the same[1] evaluation because they are one the opposite of the other.
EXAMPLE If you like to eat, you should like to fast.
ONTOLOGICAL CONSTRAINTS: A and B are defined as opposite in the ontology. This depends on the ontology structure, but in general it may happen because one of the slots defining the concepts has opposite values (x and $notx$) in the two definitions (in the example, ingesting or not ingesting food).
ACCEPTABILITY CONSTRAINTS: The audience evaluate positively/negatively one of the two states.
RELEVANCE CONSTRAINTS: The audience is aware of the opposite relation.
SUFFICIENCY CONSTRAINTS: A and B share other properties (e.g. they require the same effort, or can be done at the same places, etc.)

Schema 14 (Inverse): Two actions, A and B should be given the same evaluation because they are one the inverse of the other.
EXAMPLE If you like to eat, you should like to cook.
ONTOLOGICAL CONSTRAINTS: A and B are actions, and a concept c is part of their definitions but with different roles (in the example, "food" is the product of cooking, and the object of eating).
ACCEPTABILITY CONSTRAINTS: The audience evaluate positively/negatively one of the two actions.
RELEVANCE CONSTRAINTS: The audience is aware of the inverse relation.
SUFFICIENCY CONSTRAINTS: A and B share other properties (e.g. they require the same effort, or can be done at the same places, etc.)

A.1.5 Transitivity

This relation is in fact a "meta" relation, as it can be applied to any other quasi-logical schema, and exploits the relation that: if A is in a relation r with B, and B is in r with C, then A is in the relation r with C. Therefore if a quasi-logical schema can be used to value a over b, and the same schema can be used to value b over c, than the transitivity relation applied to the schema allows valuing a over c.

A.1.6 Probability

Schema 15 (By the Probable): A positive/negative evaluation is mitigated by the rarity of occurrence.
EXAMPLE: There is only one chance to win the lottery.
ONTOLOGICAL CONSTRAINTS: The event is being evaluated can be associated to a probability measure.
ACCEPTABILITY CONSTRAINTS: The audience values the event positively/negatively.
RELEVANCE CONSTRAINTS: The probability relation must be in focus.
SUFFICIENCY CONSTRAINTS: The occurrence of the event is only a matter of probability.

Schema 16 (Variability): A more extensive collection of items will not fail to be advantageous.
EXAMPLE: There are so many varieties of fruit: you will find one that you like.

[1] Note that while the same evaluation seems counterintuitive, the schema in fact applies the symmetrical relationship to associate the two concepts. In order to give the opposite evaluation to the two states one can use other schemata, e.g. in the sequential group.

ONTOLOGICAL CONSTRAINTS: The class contains many instances.
ACCEPTABILITY CONSTRAINTS: The audience accepts that there are many instances.
RELEVANCE CONSTRAINTS: The fact that the class contains many instances is in focus.
SUFFICIENCY CONSTRAINTS: The number of instances is significant.

A.1.7 Comparison

Schema 17 (Superlative): Something is portrayed as the most valuable thing. The refuse to compare it with other things alone suggests that the standpoint can be held.
EXAMPLE: Health is the most precious of gifts.
ONTOLOGICAL CONSTRAINTS: none
ACCEPTABILITY CONSTRAINTS: none
RELEVANCE CONSTRAINTS: none
SUFFICIENCY CONSTRAINTS: The claim is true (there are no other things more valuable than health).

Schema 18 (Comparison): An evaluation is forced on an item by claiming it can be compared to something valuable/not valuable (even if it is less/more valued).
EXAMPLE: This thesis is less good than Aristotle's writings (... but it can be compared to!)
ONTOLOGICAL CONSTRAINTS: Two instances of the same class are used.
ACCEPTABILITY CONSTRAINTS: One of the instance is valuable "by commonplace".
RELEVANCE CONSTRAINTS: The two elements are instances of the same class.
SUFFICIENCY CONSTRAINTS: The comparison is justified (e.g. the instances have enough properties in common).

Schema 19 (Sacrifice): The evaluation is done on the basis of how much effort is needed.
EXAMPLE: It takes 3 hours to cook a nice tomato sauce.
ONTOLOGICAL CONSTRAINTS: The ontology contains a notion of effort or cost that can be applied to the element to value/devalue.
ACCEPTABILITY CONSTRAINTS: The cost/effort is perceived as high/low.
RELEVANCE CONSTRAINTS: none.
SUFFICIENCY CONSTRAINTS: All the element with the same effort are equally valuable.

A.2 Based on the structure of reality

A.2.1 Sequential

These are all schemata derived from the causal link $Act \rightarrow Consequence$. They emphasise (value or devalue) one of the two terms of the link.

Concentrating on the *Consequence*

Schema 20 (By Consequence): devalue the effect of a causal link presenting it as a simple consequence (e.g. by showing it is only one of many) rather than a desirable end.
EXAMPLE A tan is just the consequence of a chemical reaction to the exposition to the sun.
ONTOLOGICAL CONSTRAINTS: The schema can only be applied to a state of affairs, which must have a recognised cause or means.
ACCEPTABILITY CONSTRAINTS: The means presented is not valued by the audience.
RELEVANCE CONSTRAINTS: The consequence relation, and the evaluation of the means should be in focus.

SUFFICIENCY CONSTRAINTS: The means is necessary and sufficient to produce the effect. The end should not have other positive consequences (which may become the desirable end).

Schema 21 (By Effects): value the effect of a causal link, proposing it as “the end” (the act is played down as just one way to achieve the effect, which is important. The schema “creates the need” and suggests a way to achieve it).
EXAMPLE The only important thing is to disarm Saddam.
ONTOLOGICAL CONSTRAINTS: The schema can only be applied to a state of affairs, which must have a recognised cause or means.
ACCEPTABILITY CONSTRAINTS: None.
RELEVANCE CONSTRAINTS: The audience is aware of a possible cause/means to achieve the end.
SUFFICIENCY CONSTRAINTS: The cause/means is necessary and sufficient for the end to materialise. The end should not have further negative/positive consequences, from the perspective chosen for the argument.

Schema 22 (Of the Easy): value an end just because it is easy to achieve.
EXAMPLE Eating fruit can simply be done by having a glass of orange juice.
ONTOLOGICAL CONSTRAINTS: The schema can only be applied to an action, which must have a recognised consequence.
ACCEPTABILITY CONSTRAINTS: The act is easy, and valuable.
RELEVANCE CONSTRAINTS: The consequence relation, and the evaluation of the act should be in focus.
SUFFICIENCY CONSTRAINTS: The act is sufficient to achieve the end.

Schema 23 (Of the Difficult): devalue an end because it is hard to achieve.
EXAMPLE Slimming will require hours upon hours of exercise.
ONTOLOGICAL CONSTRAINTS: The schema can only be applied to an action, which must have a recognised consequence.
ACCEPTABILITY CONSTRAINTS: The audience believe the end is valuable, and the act is difficult (e.g it takes long, the actor is not particularly good at it, it is expensive etc.)
RELEVANCE CONSTRAINTS: The consequence relation, the difficulty of the act, and the evaluation of the consequences should be in focus.
SUFFICIENCY CONSTRAINTS: There must not be other means.

Concentrating on the *Act*

Schema 24 (Pragmatic): Aims at evaluating an act or an event in terms of its positive or negative consequences (NR § 62). The consequences can be observed, foreseen or even purely hypothetical.
EXAMPLE: since eating apples will help you slimming, and you would like to slim, then you should eating fruit.
ONTOLOGICAL CONSTRAINTS: The schema can only be applied to an action, which must have a recognised consequence.
ACCEPTABILITY CONSTRAINTS: The audience should value positively/ negatively the consequence of the action from one perspective.
RELEVANCE CONSTRAINTS: The consequence relation, and the evaluation of the consequences should be in focus.
SUFFICIENCY CONSTRAINTS: The action is necessary and sufficient for the consequence to happen. The action should not have negative/positive consequences, from the perspective chosen for the argument.

Schema 25 (By Effectiveness): value/devalue an act because it is the best/worst way to achieve the end.
EXAMPLE Cooking is too much effort for eating.

ONTOLOGICAL CONSTRAINTS: The schema can only be applied to an action, which must have a recognised consequence.
ACCEPTABILITY CONSTRAINTS: The audience believes the end is valuable. The act is cheap/expensive.
RELEVANCE CONSTRAINTS: The consequence relation and the evaluation of the end/act have to be in focus.
SUFFICIENCY CONSTRAINTS: The act achieves the end. Other means to achieve the end are less convenient.

Schema 26 (By Means): devalue an act reversing it into a means
EXAMPLE You cook just because you need to eat.
ONTOLOGICAL CONSTRAINTS: The schema can only be applied to an action, which must have a recognised consequence.
ACCEPTABILITY CONSTRAINTS: The audience believes the end is valuable but from a perspective which is low in the hierarchy.
RELEVANCE CONSTRAINTS: The consequence relation, and the evaluation of the consequences should be in focus.
SUFFICIENCY CONSTRAINTS: The act has no other valuable consequences.

Schema 27 (By Ends): value an act reversing it into an end
EXAMPLE Cooking is a great activity (not just a means to eat).
ONTOLOGICAL CONSTRAINTS: The schema can only be applied to an action, which must have a recognised consequence.
ACCEPTABILITY CONSTRAINTS: The audience either believes the end is not particularly valuable, but needs to be achieved, or believes the end is valuable, but there are other means to achieve it.
RELEVANCE CONSTRAINTS:
SUFFICIENCY CONSTRAINTS:

Schema 28 (Of the Motives): value/devalue an act because of a distant consequence.
EXAMPLE Eating fruit will make you slim, but then your Gucci dress will no longer fit.
ONTOLOGICAL CONSTRAINTS: The schema can only be applied to an action, which must have a recognised consequence (C1), which in turn has another consequence (C2).
ACCEPTABILITY CONSTRAINTS: The audience should value positively/negatively C2, and negatively/positively C1.
RELEVANCE CONSTRAINTS: The two consequence relations, and the evaluation of the consequences should be in focus.
SUFFICIENCY CONSTRAINTS: The act is necessary and sufficient to achieve C1. C1 is necessary and sufficient to achieve C2. Other means to achieve C2 are less convenient.

A.2.2 Coexistence

As opposed to sequential relations, the relations of coexistence relate two terms which are not "on an equal level" (NR § 68), in the sense that one of them is more "basic" than the other, like the essence and its manifestations.

Schema 29 (Persons and their acts): The persons are valued for their "stability", therefore their acts should be consistent.
EXAMPLE: You are a sensible person: you should take care of your health.
ONTOLOGICAL CONSTRAINTS: An act or behaviour or property can be associated to the audience, and the act/behaviour to promote is consistent with such act (e.g. it is more general, or it is consequential, etc.).
ACCEPTABILITY CONSTRAINTS: The audience accepts the association with the act/behaviour/property.

RELEVANCE CONSTRAINTS: The relation between the two situations should be evident to the audience.
SUFFICIENCY CONSTRAINTS: The negation/opposite of the act/behaviour to promote cannot coexist with the audience's act/behaviour (not taking care of one's health is not sensible).

Schema 30 (Authority): The opinion of a well established authority is valued.
EXAMPLE: Doctors suggest to eat 5 fruit&veg a day.
ONTOLOGICAL CONSTRAINTS: The opinion presented (eat 5 fruit&veg a day is good) is held by an individual or a group which is considered an authority with respect to the perspective to promote (health).
ACCEPTABILITY CONSTRAINTS: The audience accepts the authority.
RELEVANCE CONSTRAINTS: none
SUFFICIENCY CONSTRAINTS: The opinion can be seen from the perspective from which the authority is seen ("Doctors suggest to buy a Ferrari" should not be admissible).

Schema 31 (Group/Member): An act is valued/devalued as being usual in a group of valued/devalued people.
EXAMPLE Good boys eat lots of fruit.
ONTOLOGICAL CONSTRAINTS: The act to promote is associated with a group of individual who have it as habit.
ACCEPTABILITY CONSTRAINTS: The group has the desired evaluation.
RELEVANCE CONSTRAINTS: None.
SUFFICIENCY CONSTRAINTS: The habit is associated with each element of the group, and there are no other contradictory behaviours in the group.

Schema 32 (Individuality): An act is valued/devalued as being unusual for the group the audience belong to.
EXAMPLE Be the first of your friends to start eating fruit.
ONTOLOGICAL CONSTRAINTS: The act to promote is not associated with a group of individual.
ACCEPTABILITY CONSTRAINTS: None.
RELEVANCE CONSTRAINTS: None.
SUFFICIENCY CONSTRAINTS: The habit will not cause the person to be excluded from the group (as in "be the first vegetarian to eat meat")

A.3 Establishing the structure of reality

Various figures of speech belong to this category, most notably metaphor and analogy. These were not considered at this stage.

Schema 33 (Example): Creates an evaluation by generalising from a specific case.
EXAMPLE: Fruit is healthy: consider oranges (they contain vitamin C)
ONTOLOGICAL CONSTRAINTS: The ontology contains an element e subclass (or instance) of another element s.
ACCEPTABILITY CONSTRAINTS: none.
RELEVANCE CONSTRAINTS: The relation should be in focus.
SUFFICIENCY CONSTRAINTS: The property of the subclass/instance can be generalised to the whole class.

Schema 34 (Illustration): Reinforce an evaluation by considering a specific case.
EXAMPLE: Fruit is healthy: consider oranges (they contain vitamin C)
ONTOLOGICAL CONSTRAINTS: The ontology contains an element e subclass (or instance) of another element s.
ACCEPTABILITY CONSTRAINTS: The audience values positively the superclass.

RELEVANCE CONSTRAINTS: The relation should be in focus.
SUFFICIENCY CONSTRAINTS: The property of the class can be inherited by the subclass/instance.

Schema 35 (Model/Antimodel): Aims at encouraging/discouraging imitation of a behaviour.
EXAMPLE: It is good practice to use olive oil for cooking: all TV chefs do so.
ONTOLOGICAL CONSTRAINTS: The concept to be valued is in an action. A person, or a category of person, or an abstract "perfect being", can be identified that could have that action as a habit.
ACCEPTABILITY CONSTRAINTS: The audience should share the view that the individual chosen is in fact a model/antimodel, from some perspective.
RELEVANCE CONSTRAINTS: It should be evident to the audience that the individual chosen for imitation is a model/antimodel.
SUFFICIENCY CONSTRAINTS: No other "relevant" habit of the model, where relevant means here that can be evaluated from the perspective chosen, are counterproductive (bad/good).

A.4 Dissociation

This is a very rich and complex class, that has not been analysed in this thesis. Reasoning by dissociation basically involves breaking a classification or association by introducing new perspectives. This is done by contrasting two terms, or philosophical pairs:

$$\frac{\text{Term I}}{\text{Term II}}$$

The argument is that while an evaluation is valid in the reality described by Term I, it is not anymore if considering the reality described by Term II. The prototypical philosophical pairs is the

$$\frac{\text{appearance}}{\text{reality}}$$

"Appearance" corresponds to what is immediately perceivable. But appearances can be misleading, and they may create inconsistencies. "Reality" is by definition coherent, and looking at the world from this point of view can eliminate the incompatibilities. After the dissociation has been pointed out, the second term (reality in this case) can be used to distinguish what can be retained and what has to be discarded of the world when seen from the first term's point of view (appearance). Argumentation by dissociation therefore involved a more elaborated discourse, as the distinctions has to be pointed out. For instance, one can argue that in order to lose weight you have to eat more often and better, as losing weight by starvation is only an apparent solution, while a real diet is a balanced one.

Appendix B

E-mail Dialogue Example

D2-4.0	Hi! What are you having for dinner tonight?
D2-4.1	I'm planning to eat a jacket potato (no butter or marg) with grated Edam cheese and salad. For dessert, a baked apple with sultanas and plain yoghurt.
D2-4.2	It's good you have a salad! How do you normally prepare it?
D2-4.3	I normally use lettuce, red peppers, broccoli, mushrooms, cucumber, tomatoes, sometimes water cress, sometimes tined beans (eg. kidney, chick peas). The dressing is made of olive oil, usually cider vinegar, black pepper, mustard and a touch of artificial sweetener. Ok?
D2-4.4	A very good choice! Is this what you usually have for dinner?
D2-4.5	Perhaps twice a week. Other times I might make pasta with a tomato/vegetable sauce or rice and beans/vegetables or a vegetarian pizza or bread, cheese and salad. I usually eat raw fruit for dessert (an apple or perhaps a banana and yogurt).
D2-4.6	For how many months/years have you regularly had this sort of dinner?
D2-4.7	I've been eating like this for five years or so, although I did eat fish (but no meat) until about two months ago. I changed my diet in 1992 because I wanted to lose some weight and become fitter and healthier.
D2-4.8	How much fruit would you eat per day?
D2-4.9	Typically a banana and a kiwi fruit with breakfast, an apple with lunch and an apple and/or a banana in the evening. I guess that's quite a lot.
D2-4.10	Fruit is never too much. It is recommended to have at least 5 portions of fruit and vegetables per day.
D2-4.11	That's great because today I've already eaten a banana and a kiwi fruit for breakfast plus, at lunch, two strawberries, a plum, a pear and another banana! My excuse for this extravagance is that I went to a lunchtime meeting at which fruit was liberally provided.
D2-4.12	How much bread do you take per day?
D2-4.13	One slice of toast for breakfast (wholemeal, with marg) and maybe four more slices for lunch with cheese.
D2-4.14	Filler foods (bread, potatoes, rice, pasta, cereals) should constitute the main part (40-50%) of a person's meal. Do you think you meet this target?
D2-4.15	yes
D2-4.16	Brilliant. If you consider your current dietary habit, is there anything that you would like to change/improve?

D2-4.17	Yes, I would like to eat a greater variety of foods. As a vegetarian who tries to avoid high fat foods and who doesn't want to spend too much time preparing food, I feel my choice is often quite limited. I also know I serve myself meals that are often too large, especially pasta or rice-based dishes or sandwiches for lunch. I think I should cut down on caffeine (I often drink six or so mugs of tea a day) and drink more water. Another point: I wish I could be more relaxed about what I eat. Because I know high fat food is unhealthy I try to avoid it completely and feel bad/guilty if I do eat it. I'd like to feel relaxed enough to eat the occasional higher fat (and tasty) meal without worrying about it.
D2-4.18	There are so many varieties of vegetables and quick recipes to experiment with! Just have fun with them! I wouldn't be too worried about eating some more starchy foods, as long as you pay attention to how much sodium you take (you may wish to examine the nutrition labelling on your regular choice of bread as many 'manifactured' sliced loaves are high in sodium). You can substitute a mug of tea with one of herbal tea, from time to time. It is what we do for most of the time that makes an impact on our health not the things we do occasionally. A good advice is probably eat what you like and thoroughly enjoy it, just not too often! Some tips can be very useful: for instance, if you like chips, consider that the thicker the chip, the less fat it absorbs during cooking! (home made thick chips have half the fat of French fries).
D2-4.19	Good points, especially about taking the longer view. I never eat chips, by the way, so I never need to worry about how thick they are! I didn't know about the sodium in bread. Although your advice about experiementing with different recipes makes sense, I find it hard to find the time/motivation to do this.
D2-4.20	Do you cook your meals yourself?
D2-4.21	I share the cooking 50:50 with my partner.
D2-4.22	Do you like cooking?
D2-4.23	Not especially. There are three or four basic meals I know how to cook and I don't have much interest in learning anything new (even though I know it's important). Cooking feels to me like a lot of effort for something (ie. eating) that's over quite quickly. Also, I often feel tired at the end of a day's work and don't want to spend too much time in the kitchen.
D2-4.24	You do not cook just because you have to eat! Cooking can be a very relaxing and interesting activity, better than watching TV!
D2-4.25	I know you're right but that still doesn't make it easy to do! I think I'm still in what the psychologists call the 'contemplation stage' when it comes to changing my cooking habits. In fact, my partner's the same so she can't easily encourage me to get stuck in.
D2-4.26	Well, you did make a change in your eating habit, and you have been acting on a regular basis for quite a while, so you are in a "maintenance stage" as far as your diet is concerned. Having fun at cooking could only help you to 'stay there', perhaps decreasing your wish for fatty-tasty things. And maybe YOU can encourage your partner! Do you usually cook together, or do you take turns?
D2-4.27	We take turns.
D2-4.28	Cooking together could be funnier, and more stimulating.
D2-4.29	I'll certainly discuss it with my partner but the main problem with this idea is time.
D2-4.30	You seem to consider cooking as a time expensive, complex task. It hasn't to be like that!
D2-4.31	That could be because I'm a man. I've struggled to reach a point where I reckon I do my fair share of the cooking (and the rest of the housework) but I wasn't brought up to do it. That means that it's still a struggle, not easy or a joy.

D2-4.32	That's why it could be useful seeing things from a different perspective, and being creative. How often do you eat away from home?
D2-4.33	I'm self-employed and work from home so eat most of my meals there. I guess, on average, I eat out three times a week (a mixture of lunches and dinners).
D2-4.34	Do you find that eating away from home make healthier choices more or less difficult?
D2-4.35	Much more difficult. It's very hard to buy low-fat food as sandwiches or meals. I find that even many vegetarian restaurants use far too much cheese or smother salads in high-fat dressings.
D2-4.36	Do you use any ”trick” when you eat out in order to improve things?
D2-4.37	'mmmm. I sometimes buy low-calorie sandwiches (although they often taste like paper) or buy some fruit. Alternatively, I'll just eat less and avoid anything fried.
D2-4.38	It's not always necessary to eat less. For instance, if you go for a pizza, ask for a thick crust one, it's more filling and you will eat proportionally more base and less toppings. Or, try to go to places when you can choose the filling of your sandwich. And ask for a salad and a slice of bread before the meal.
D2-4.39	Good points.
D2-4.40	Some time ago you said you would like to eat a greater variety of foods. Which is a great idea because mixtures of foods complement each other, ensuring that all essential nutrients (including fat!) are provided to the body. Can you list the pros and cons of actually doing it?
D2-4.41	pros: Better health, Tastier food. cons: More expensive, More shopping, More cooking, Less free time.
D2-4.42	I have some difficulties in understanding your cons list: isn't it a matter of buying/cooking something “instead” of something else, rather than “in addition” to something else? The shopping/cooking time would be the same, only you would buy different things every week. I think.
D2-4.43	No, because to eat a greater variety of food I would have to spend time following and learning new recipes as well as searching for new ingredients (and I might not be able to buy all of them in my local supermarket). I don't have a car and it's much esaier for me if I can buy everything I need in the same place. You probably think I'm being 'difficult' about all this but the fact is that I do want to broaden my repertoire and I'm trying to be honest about why I don't. I guess that, in the context of all the demands on my time and attention, this doesn't yet feel important enough for the extra effort and commitment required.
D2-4.44	Again, you seem to label the act of cooking as something complex, for which training is necessary. This is probably true for “gourmet” dishes, but sometimes a new dish is just cooking the same meal but with different things. Eg. substituting cabbage with kale, or aubergines with artichokes. Does it make sense to you? Or do you have something else in mind when you think of eating a greater variety of foods?
D2-4.45	I'd say it feels complicated to make the kinds of changes you suggest (and I'm not talking about gourmet meals) when I don't know an awful lot about the process of cooking with different foods. It wouldn't have occurred to me to make the substitutions you suggest because (I guess) I don't know enough about the composition or inter-relationship of different foods. So if I had learnt a recipe with aubergines I would always make it with aubergines. Does that help?

Appendix C

Informal Competency Questions

C.1 Questions regarding the argumentation ontology

ICQ number:	1
Question:	Is there an argument in favour of a topic T?
Rationale:	Main aim.
Decomposition:	One of the following: 1. the topic S is valuable 2. a New Rhetoric schema can be applied in favour of T

ICQ number:	2
Question:	Is there an argument against a topic T?
Rationale:	Main aim.
Decomposition:	One of the following: 1. the topic T is not valuable 2. a New Rhetoric schema can be applied against T

ICQ number:	3
Question:	Is there a definition argument in favour of a topic T?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	Topic T generalises topic T_s and topic T_s is valuable

ICQ number:	4
Question:	Is there a definition argument against a topic T
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	Topic T generalises topic T_s and topic T_s is not valuable

ICQ number:	5
Question:	Is there an argument by Incompatibility in favour of a state S?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	There exist another state S_1, which is not valuable, and S and S_1 are not compatible.

ICQ number:	6
Question:	Is there an argument by Incompatibility against a state S?
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	There exist another state S_1, which is valuable, and S and S_1 are not compatible.

ICQ number:	7
Question:	Is there an argument by Incompatibility in favour of an action A?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	One of the following: 1. Action A leads to a state S, and there is an argument by Incompatibility in favour of S. 2. There is an action A_1, which is not valuable, and A and A_1 are not compatible

ICQ number:	8
Question:	Is there an argument by Incompatibility against an action A?
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	One of the following: 1. Action A leads to a state S, and there is an argument by Incompatibility against S. 2. There is an action A_1, which is valuable, and A and A_1 are not compatible

ICQ number:	9
Question:	Is there a reciprocity argument in favour of action A when applied to argument P?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	There is an action A_1 which is valuable when applied to argument P

ICQ number:	10
Question:	Is there a reciprocity argument against action A when applied to argument P?
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	There is an action A_1 which is not valuable when applied to argument P

ICQ number:	11
Question:	Is there a transitivity argument in favour of a topic T versus topic T_1?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	Topic T is more valuable than topic T_2 and topic T_2 is more valuable than topic T_1

ICQ number:	12
Question:	Is there a transitivity argument against a topic T versus topic T_1?
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	Topic T is less valuable than topic T_2 and topic T_2 is less valuable than topic T_1

ICQ number:	13
Question:	Is there a part/whole argument in favour of a topic T
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	Topic T is part of T_s and T_s is valuable

ICQ number:	14
Question:	Is there a part/whole argument against topic T
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	Topic T is part of T_s and T_s is not valuable

ICQ number:	15
Question:	Is there an argument by partition in favour of topic T?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	There is a set of topics, $Topics = \{T_1, \ldots, T_n\}$, such that $T \in Topics$ and all elements of $Topics$ have a property P and all elements not in $Topics$ have not property P and all T_i in $Topics, T_1 \neq T$ are not valuable.

ICQ number:	16
Question:	Is there a comparison argument in favour of topic T
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	T has some quantifiable property P, the fact that T has property P is valuable, and there is another topic T_1 which has "less P" than T

ICQ number:	17
Question:	Is there a comparison argument against topic T
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	T has some quantifiable property P, the fact that T has property P is valuable, and there is another topic T_1 which has "more P" than T

ICQ number:	18
Question:	Is there an argument by causality in favour of state S?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	S leads to a state S_1, which is valuable.

ICQ number:	19
Question:	Is there an argument by causality against state S?
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	S leads to a state S_1, which is not valuable.

ICQ number:	20
Question:	Is there a pragmatic argument in favour of action A?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	A leads to a state S_1, which is valuable.

ICQ number:	21
Question:	Is there a pragmatic argument against action A?
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	A leads to a state S_1, which is not valuable.

ICQ number:	22
Question:	Is there an argument by authority in favour of topic T?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	There is some authority for whom T is valuable.

ICQ number:	23
Question:	Is there an argument by authority against topic T?
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	There is some authority for whom T is not valuable.

ICQ number:	24
Question:	Is there an argument by group pressure in favour of topic T?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	The hearer belongs to a group, for whom T is valuable.

ICQ number:	25
Question:	Is there an argument by group pressure against topic T?
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	The hearer belongs to a group, for whom T is not valuable.

ICQ number:	26
Question:	Is there an argument by model in favour of action A?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	There is a model whose habit is A

ICQ number:	27
Question:	Is there an argument by model against action A?
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	There is a model whose habit is not doing A

ICQ number:	28
Question:	Is there an argument by anti-model in favour of action A?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	There is an anti-model whose habit is not doing A

ICQ number:	29
Question:	Is there an argument by anti-model against action A?
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	There is an anti-model whose habit is A

ICQ number:	30
Question:	Is there an argument by example in favour of topic T?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	There exists T_1 which specifies T and T_1 is valuable

ICQ number:	31
Question:	Is there an argument by example against topic T?
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	There exists T_1 which specifies T and T_1 is not valuable

ICQ number:	32
Question:	Is there an argument by dissociation in favour of topic T?
Rationale:	Solve ICQ 1 and implementation of New Rhet schema.
Decomposition:	There exist T_1 and T_2 which specify T and T_1 is valuable while T_2 is not

ICQ number:	33
Question:	Is there an argument by dissociation against topic T?
Rationale:	Solve ICQ 2 and implementation of New Rhet schema.
Decomposition:	There exist T_1 and T_2 which specify T and T_1 is valuable while T_2 is not

ICQ number:	34
Question:	is topic T valuable?
Rationale:	Solve ICQ 1, 6, 8, 13, 16, 18
Decomposition:	Topic T is valuable for the current audience.

ICQ number:	35
Question:	is topic T not valuable?
Rationale:	Solve ICQ 2, 5, 7, 14, 15, 17, 19
Decomposition:	Topic T is not valuable for the current audience.

ICQ number:	36
Question:	is action A valuable when applied to the argument C?
Rationale:	Solve ICQ 9
Decomposition:	The audience attributes a good value to A when applied to argument C with respect to a perspective P.

ICQ number:	37
Question:	is action A not valuable when applied to the argument P?
Rationale:	Solve ICQ 10
Decomposition:	The audience attributes a good value to A when applied to argument C with respect to a perspective P.

ICQ number:	38
Question:	is topic T more valuable than topic T_1?
Rationale:	Solve ICQ 11, 39
Decomposition:	One of the following: 1. Topic T is valuable and T_1 is not valuable from any perspective 2. Topic T is valuable from a more important perspective than T_1 3. The better perspective from which T is valuable is more important than the better perspective from which T_1 is valuable 4. the worst perspective from which T is not valuable is better than the worst perspective from which T_1 is not valuable

ICQ number:	39
Question:	is topic T less valuable than topic T_1?
Rationale:	Solve ICQ 12
Decomposition:	T_1 is more valuable than topic T

ICQ number:	40
Question:	is topic T valuable for people P?
Rationale:	Solve ICQ 22, 24, 34
Decomposition:	People P have the opinion that topic T has a good value with respect to a perspective Per.

ICQ number:	41
Question:	is topic T not valuable for people P?
Rationale:	Solve ICQ 23, 25, 35
Decomposition:	People P have the opinion that topic T has a bad value with respect to a perspective Per.

ICQ number:	42
Question:	Is there a model for action A?
Rationale:	Solve ICQ 26, 27
Decomposition:	There exist a model who has A as a habit.

ICQ number:	43
Question:	Is there an anti-model for action A?
Rationale:	Solve ICQ 26, 27
Decomposition:	There exist a anti-model who has A as a habit.

ICQ number:	44
Question:	Which is the better perspective for topic T?
Rationale:	Solve ICQ 38
Decomposition:	Find a perspective P from which topic T has a good value and such that any other perspective P_1 from which topic T has a good value is less important than P.

ICQ number:	45
Question:	Which is the worst perspective for topic T?
Rationale:	Solve ICQ 38
Decomposition:	Find a perspective P from which topic T has a bad value and such that any other perspective P_1 from which topic T has a bad value is less important than P.

ICQ number:	46
Question:	Is a perspective more important than another?
Rationale:	Solve ICQ: 38, 44, 45
Decomposition:	None (basic ICQ)

C.2 Questions regarding the state-action ontology

ICQ number:	47
Question:	are states S and S_1 incompatible?
Rationale:	Solve ICQ 6, 5, 50
Decomposition:	One of the following: 1. S prevent S_1 from happening, or vice versa 2. S leads to a state S_2 which is incompatible with S_1 3. S_1 leads to a state S_2 which is incompatible with S 4. S leads to S' and S_1 leads to S'_1 and S' and S'_1 are incompatible.

ICQ number:	48
Question:	does action A lead to state S?
Rationale:	Solve ICQ 7, 8
Decomposition:	One of the following: 1. A achieve, favour, maintain or strengthen S 2. A achieve, favour, maintain or strengthen S_1, a state which leads to S

ICQ number:	49
Question:	does state S lead to state S_1?
Rationale:	Solve ICQ 47, 50
Decomposition:	One of the following: 1. S achieve, favour, maintain or strengthen S 2. S achieve, favour, maintain or strengthen S_1, a state which leads to S

ICQ number:	50
Question:	are actions A and A_1 incompatible?
Rationale:	Solve ICQ 8, 7
Decomposition:	A leads to a state S, and A_1 leads to a state S_1 and S and S_1 are incompatible

ICQ number:	51
Question:	Can action A be applied to property argument P?
Rationale:	Solve ICQ 9, 10, 36, 37
Decomposition:	action A has P as object, location, effort or time.

ICQ number:	52
Question:	Find the list of Topics having property P
Rationale:	Solve ICQ 15
Decomposition:	None (basic ICQ)

ICQ number:	53
Question:	Topic T has more of property P than topic T1
Rationale:	Solve ICQ 16, 17
Decomposition:	None (basic ICQ)

ICQ number:	54
Question:	Find all specializations of topic T.
Rationale:	Solve ICQ 30, 31, 32, 33
Decomposition:	None (basic ICQ)

ICQ number:	55
Question:	Find all generalizations of topic T.
Rationale:	Solve ICQ 11, 12
Decomposition:	None (basic ICQ)

ICQ number:	56
Question:	Is a person part of a Group?
Rationale:	Solve ICQ:24, 25
Decomposition:	None (basic ICQ)

C.3 Questions regarding the behavioural ontology

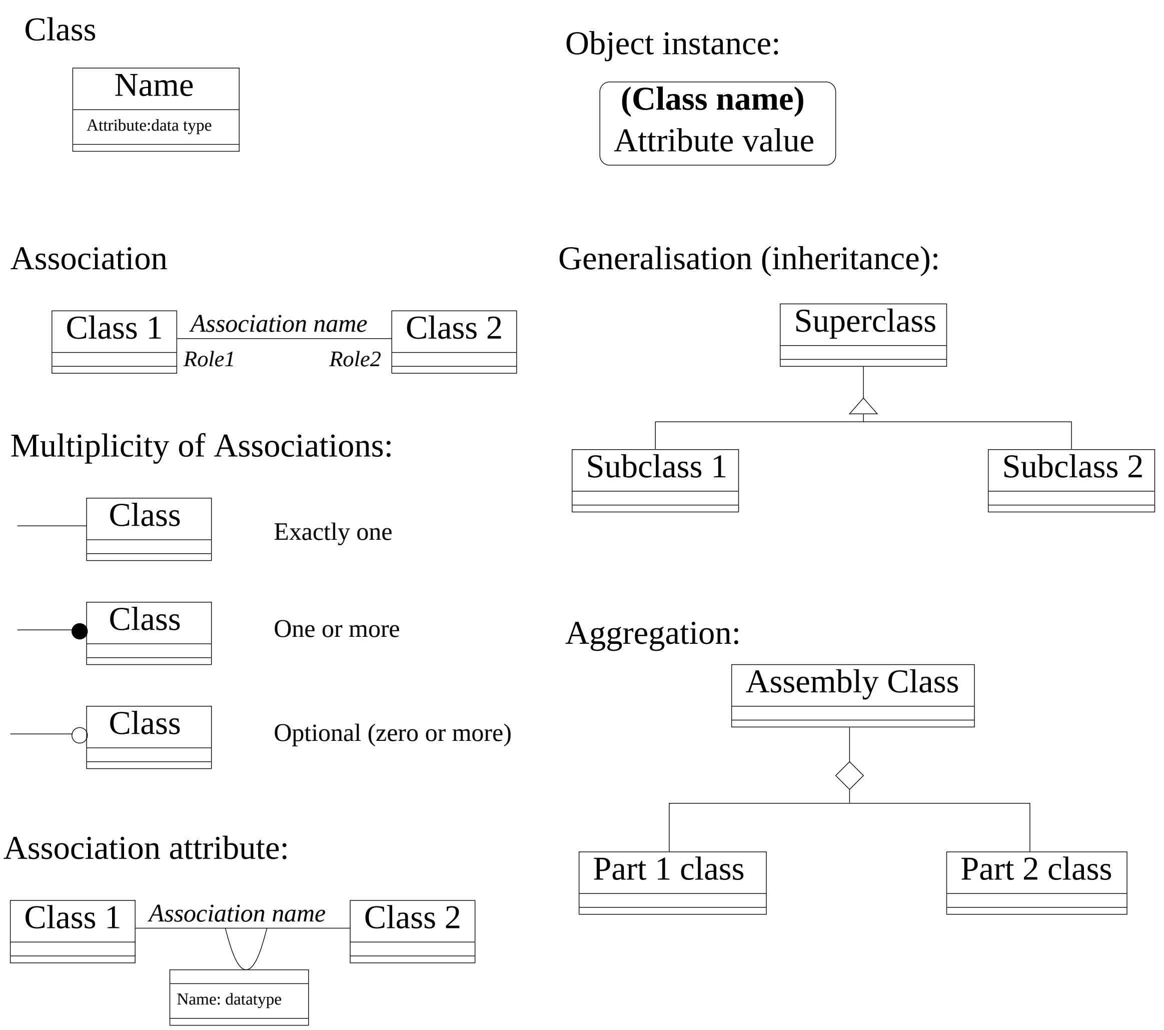
ICQ number:	57
Question:	Which state is a person in towards an action?
Rationale:	Find the stage of change
Decomposition:	None (basic ICQ)

ICQ number:	58
Question:	Does a person feel to be susceptible to a state?
Rationale:	Establish health belief
Decomposition:	None (basic ICQ)

Appendix D

Notation for Object Models

Below it is depicted the Object Model Notation we have used in this thesis, as adapted from the one by Rumbaugh *et al.* (1991).



Appendix E

ATMS

In this Appendix a summary is given of the main principles and techniques of Assumption Based Maintenance Systems, as implemented in the prototype system, and as explained by de Kleer (1986a), de Kleer (1986b), de Kleer (1986c), de Kleer (1988), Forbus & de Kleer (1988), de Kleer (1993).

E.1 General Definitions

The basic object in an ATMS network is a *node*, which is represented as the triple:

$$\langle datum, label, justification \rangle$$

where:

- the *datum* is just an identifying name for the problem solver datum;
- the *label* is a set of *environments*, each representing one context in which the datum holds.
- the *justification* shows how the problem solver derived the node from other nodes (the chain of inferences used);

The *environment* is the key notion here. These are minimal sets of validating assumptions for the datum. In other words, each set is necessary and sufficient to make the datum true in that environment. In a way, the label represents how defeasible the datum is. For instance the node[1]:

$$\langle x, \{\{A, B\}, \{C\}\}, \{(b), (c, d)\}\} \rangle$$

represents the fact that datum x can be derived in two ways, either from b, or from $c \wedge d$. Furthermore, it holds in the environments containing either assumptions $\{A, B\}$ or assumption C.

A *premise*, that is a datum that is always true, is represented by a node containing the empty environment as label:

$$\langle p, \{\{\}\}, \{\ldots\} \rangle$$

to signify that there are no restrictions to the believability of the datum (it is true in every context). Conversely, a *contradiction*, that is a datum that is always false, has an empty label:

$$\langle c, \{\}, \{\ldots\} \rangle$$

[1] In de Kleer's notation, we indicate assumption with upper case letters, and derived data with lower case ones.

Finally, an assumption is represented as a node whose label contains the environment consisting of the assumption itself:

$$\langle A, \{\{A\}\ldots\}, \{(A), \ldots\}\}\rangle$$

for, when created, the assumption is self-justifying. New environments and new justifications can be added while reasoning progresses.

If the problem solver discovered a contradiction, the set of assumptions causing it is added to the justification set of a special node representing falsity:

$$\langle \perp, \{\}, \{\ldots\}\rangle$$

The list of inconsistent environments, that can be thought of as the label of datum $\perp$, together with the list of assumptions, can be used to determine all the possible *extensions*, that is the “possible worlds” that the ATMS describes. For example, from the justifications:

$$\begin{array}{l} A \rightarrow a \\ B \rightarrow b \\ A, B \rightarrow c \\ a, b \rightarrow \perp \end{array}$$

the ATMS generates the two possible worlds: $\{A, a\}$ and $\{B, b\}$.

To achieve this, each time a new inference is made by the problem solver, the justification is passed to the ATMS, which updates the label for the datum corresponding to the inference’s conclusion. A propagation algorithm is used to change labels of affected nodes, if appropriate, so that each label is sound, complete and minimal. This means that: each environment in the label does derive the datum; every environment which derives the datum is a superset of one of the environments in the label; and no environment in the label is superset of any other. These properties are achieved by removing from a label environments that turned out to be inconsistent, and environments that are subsumed by other environments. The latter step is what make the ATMS efficient, as it exploits the property that if a datum is derived from a set of assumptions, it is derived from all the supersets of this set.

Basic operations

The basic operations in an ATMS are:

1. create a node for a problem solver datum; an indexing scheme is used to create a unique node name for each datum;

2. create an assumption;

3. add a justification to a node.

Adding a justification

the ATMS is provided of a new justification by the problem solver of the form:

$$(just_1)\ldots(just_k)\ldots \rightarrow n$$

where

$$just_k = (node_{1k}\ldots node_{ik}\ldots)$$

where every $node_{ik}$ has label j_{ik}.

A complete label for n is

$$\bigcup_k \{x | x \bigcup_i x_i$$

where $x_i \in j_{ik}\}$.

This is calculated as follows:

1. remove inconsistent and subsumed environment (to ensure soundness and minimality)

2. if the new label for the node is the same as the old one, exit

3. if the node is the contradiction $\perp$, each environment is added to a nogood database, and all inconsistent environments (including their supersets) are removed from every node label
   else the updating process recursively updates the labels of all the consequent nodes (i.e. of all the other nodes whose justifications which mention the node whose label changed).

In the algorithm the following data-structures are needed:

Node :

- datum (the problem solver representation of the fact)
- label (set of environments)
- justification (list of the derivation of the datum)
- consequents (list of justifications in which the node is antecedent)
- contradictory (binary slot indicating whether the datum is contradictory)

Environment :

- assumptions (set of assumptions defining it)
- nodes (set of nodes whose label mentions it)
- contradictory (binary slot indicating whether the environment is contradictory)

Justification :

- informant (the problem solver description of the inference)
- consequent (the node justified)
- antecedents (list of nodes upon which the inference depends)

E.2 Labelling algorithm

Notations:

- J = justification in the form $x_1, x_2..x_k \rightarrow n$
- a = (optional) antecedent node
- I = set of environment just added to the label of a node

The algorithm assumes that the node labels are correct before the introduction of the new justification, and proceeds by propagating the incremental changes caused by the new justification. The functions required are as follows:

PROPAGATE (J, a, I)

1. (compute the incremental label update) $L =$**WEAVE**$(a, I, \{x_1 \dots x_k\})$. If L is empty return.
2. (update label and recur) **UPDATE**(L, n)

UPDATE (L, n)

1. (detect nogoods) if $n = \perp$ then call **NOGOOD**(E) on each E in L and return $\{\}$
2. update n's label ensuring minimality:
   (a) delete every environment from L which is a superset of some label environment of n
   (b) delete every environment from the label of n which is a superset of some element of L
   (c) add every remaining environment of L to the label of n
3. propagate the incremental change to n's label to its consequences: for every justification J in which n is mentioned as an antecedent, call **PROPAGATE**(J, n, L)

WEAVE (a, I, X)

1. (termination condition) if X is empty, return I
2. (iterate over the antecedent nodes) let h be the first node of the list X and R the rest
3. (avoid computing the full label) if $h = a$ return **WEAVE**(nil, I, R)
4. (incrementally construct the incremental label): let I' be the set of all environments formed by computing the union of an environment of I and an environment of h's label [Viewed as propositional inference, each iteration in this step resolves a horn clause in which h occurs negatively with another horn clause (i.e. h with its label) to eliminate h].
5. (ensure that I' is minimal and contains no known inconsistency) remove from I' all duplicates, nogoods, as well as any environment subsumed by any other
6. return **WEAVE**(a, I', R)

NOGOOD (E)

1. mark E as nogood
2. remove E and any superset from every node label

Hyperresolution

It is not possible to encode every possible formula as a set of Horn clauses, but it is possible to add a non-Horn justification as a set of horn clauses and positive clauses of assumptions

$$(A_1 \wedge A_2 \dots \wedge A_n)$$

$choose\{A_1, A_2 \dots A_n\}$ represents the positive clause $(A_1 \wedge A_2 \dots \wedge A_n)$.

The algorithm outlined before will however fail to discover any nogood if chooses are present. For example, in $choose\{A, B\}$; $A, C \rightarrow \perp$; $B, C \rightarrow \perp$ the basic algorithm will fail to realise that $\{C\}$ is nogood. An appropriate a hyperresolution rule can be used: given a set of inconsistent assumption α_i $(A, B \rightarrow \perp)$ and a positive clause (a $choose$):

$$\frac{\begin{array}{l} choose\{A_1 \ldots A_k\} \\ nogood\ \alpha_i \text{ where } A_i \in \alpha_i \text{ and } A_{j \neq i} \notin \alpha_i \text{ for all } 1 \leq i, j \leq k \end{array}}{nogood \bigcup_i [\alpha_i - \{A_i\}]}$$

in the example:

$$\frac{\begin{array}{l} choose\{A, B\} \\ \alpha_1 = nogood\{A, C\}; \alpha_2 = nogood\{B, C\} \end{array}}{nogood\{C\}}$$

or in propositional form:

$$\frac{\begin{array}{l} A \wedge B \\ \neg A \wedge \neg C \\ \neg B \wedge \neg C \end{array}}{\neg C}$$

The hyperresolution rules helps in insuring consistency among Nogoods and labels when Chooses are present. In particular this system uses, from the ones described by de Kleer (1986b):

- **Hyperresolution rule 4:**

$$\boxed{\frac{\begin{array}{l} Choose\{A, B\} \\ Nogood\{\{A\} \cup a\} \text{ where } B \notin a \end{array}}{a \rightarrow B}}$$

  This rule states that, as one environment must contain either A or B, if A is part of a Nogood together with a set a of other assumptions (that is an environment cannot contain both A and all the assumptions in a) then every environment that contains a, as cannot contain A, must contain B.

- **Hyperresolution rule 5:**

$$\boxed{\frac{\begin{array}{l} Choose\{A_1 \ldots A_n\} \\ Nogood\{\{A_1\} \cup a_1\} \text{ where } \{A_2, \ldots, A_n\} \not\subset a_1 \\ \ldots \\ Nogood\{\{A_n\} \cup a_n\} \text{ where } \{A_1, \ldots, A_{n-1}\} \not\subset a_n \end{array}}{Nogood\{a_1 \cup \ldots \cup a_n\}}}$$

  The rule states that, as just one of $A_1, \ldots, A_n$ must be present in every environment, and as each A_i cannot be present if the set of assumptions a_i is present, then all the $a_1, \ldots, a_n$ cannot simultaneously hold, because in that case none of the $A_1, \ldots, A_n$ would hold.

In this way every kind of logic clause can be translated and encoded in the ATMS, and de Kleer (1986b) shows different techniques to do this task for a general clause of the type:

$$\{a_1, a_2, \ldots, a_n, A_1, A_2, \ldots, A_l\} \Rightarrow \{c_1, c_2, \ldots, c_m, C_1, C_2, \ldots, C_k\}$$

where the upper case atoms are assumptions and the lower case ones are derived data.

E.3 NATMS: Negated-assumption ATMS: extended algorithm

The extended algorithm proposed by de Kleer has been implemented in order to be able to deal with negated assumptions. The algorithm:

- achieves label consistency without using the hyperresolution rule,
- integrates well with the basic algorithm,
- produces more complete labels,
- allows arbitrary clauses to be encoded more efficiently and parsimoniously.

The negation of the assumption A is a non-assumption node and is referred to as $\neg A$. Every $choose$ is easily encoded as an NATMS justification:

$$choose\{A, B, C\}\text{becomes}\neg A, \neg B, \neg C \rightarrow \perp$$

Extended algorithm

Observation: any negative clause of size k is logically equivalent to any of the k implications with one literal on its right-hand side. for example:

$$\neg A \wedge \neg B \wedge \neg C$$

can be expressed as any of:

$$\begin{array}{l} A \vee B \rightarrow \neg C \\ A \vee C \rightarrow \neg B \\ B \vee C \rightarrow \neg A \end{array}$$

As all label updates can be determined by propagation, it is necessary (and sufficient) to encode all these material implication as justification. The NATMS algorithm has the equivalent to the following rule:

$$\frac{nogood\{A, A_1, \ldots, A_k\}}{A_1, \ldots, A_k \rightarrow \neg A}$$

For example, given initially empty labels, on discovering the nogood:

$$nogood\{A, B, C\}$$

the NATMS produces the following [$< x, L >$ indicates L is the label for x]:

$$\begin{array}{c} < \neg A, \{\{B, C\}\} > \\ < \neg B, \{\{A, C\}\} > \\ < \neg C, \{\{A, B\}\} > \end{array}$$

The algorithm is the same as that outlined before, with the addition of one step to the function **NOGOOD**:

NOGOOD' (E)

1. mark E as nogood
2. remove E and any superset from every node label
3. (handle negated assumptions) For every $A \in E$ for which $\neg A$ appears in some justification, call **UPDATE**$(\{E - \{A\}\}, \neg A)$

Bibliography

AAAI. 1988. *Proceedings of the 7th National Conference on Artificial Intelligence (AAAI88)*, Morgan Kaufmann Publishers.

Alchourrón, C.; Gärdenfors, P.; and Makinson, D. 1985. On the Logic of Theory Change: partial meet contraction functions and their associated revision functions. *Journal of Symbolic Logic* 50:510–530.

Allen, J., and Miller, B. 1991. The RHET System: a sequence of self-guided tutorials. Technical Report 325, University of Rochester, Computer Science Department.

Allen, J.; Schubert, L.; Ferguson, G.; Heeman, P.; Hwang, C.; Kato, T.; Light, M.; Martin, N.; Miller, B.; Poesio, M.; and Traum, D. 1995. The TRAINS Project: A Case Study in Building a Conversational Planning Agent. *Journal of Experimental and Theoretical Artificial Intelligence* 7:7–48.

Ambros Ingerson, J. 1987. IPEM: Integrated Planning, Execution and Monitoring. Master's thesis, University of Essex, Department of Computer Science.

Amgoud, L., and Cayrol, C. 1998. On the Acceptability of Arguments in Preference-based Argumentation. In Cooper, G., and Moral, S., eds., *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*.

André, E., and Rist, T. 2000. Presenting through Performing: on the use of multiple lifelike characters in knowledge-based presentation systems. In *Proceedings of the 2000 International Conference on Intelligent User Interfaces (IUI'00)*, 1–8.

Anscombre, J., and Ducrot, O. 1983. *L'Argumentation dans la Langue*. Bruxelles: P. Mardaga.

Aristotle. about 350 B.C.-1926. *Rhetorica*, (transl. Freese, J.). London: Loeb.

Aristotle. about 350 B.C.-1928. *Topica et Sophistici Elenchi*, (transl. Ross, W. D.). Oxford: Clarendon Press.

Aristotle. about 350 B.C.-1964. *Prior and Posterior Analytics*, (transl. Warrington, J.). London: Dent.

Aristotle. about 350 B.C.-1968. *de Anima*, (transl. Hamlyn, D.). Oxford: Oxford University Press.

Ashley, K. 1991. Reasoning with Cases and Hypotheticals in HYPO. *International Journal of Man-Machine Studies* 34:753–796.

Baker, M. 1999. Argumentation and Constructive Interaction. In Coirier, P., and Andriessen, J., eds., *Foundations of Argumentative Text Processing*, volume 5 of *Studies in Writing*. Amsterdam: University of Amsterdam Press. 179–202.

Ballim, A., and Wilks, Y. 1991. *Artificial Believers*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Baroni, P.; Giacomin, M.; and Guida, G. 2000. Extending Abstract Argumentation Systems Theory. *Artificial Intelligence* 120(2):251–270.

Barrie, K. 1991. Motivational Counselling in Groups. In Davidson, R.; Stephem, R.; and MacEwan, I., eds., *Counselling Problem Drinkers*. London: Tavistock/Routledge. chapter 7.

Barrington, G., and Sonenberg, L. 1997. Following the Path to the Gorillas . . . or the Banana Tree: reasoning with nested beliefs. In *Proceedings of the AI'97 workshop on Theoretical and Practical Foundations of Intelligent Agents and Agent-Oriented Systems*, 1–16.

Barth, E., and Krabbe, E. 1982. *From axiom to dialogue. A philosophical study of logics and argumentation*. Berlin/New York: Walter de Gruyter.

Barwise, J., and Perry, J. 1983. *Situations and Attitudes*. Cambridge, MA: MIT Press.

Bateman, J., and Zock, M. 2002. List of Natural Language Generation Systems. http://www.fb10.uni-bremen.de/anglistik/langpro/NLG-table/NLG-table-root.htm.

Bateman, J.; Magnini, B.; and Fabris, G. 1995. The Generalized Upper Model Knowledge Base: Organization and use. In Mars, N., ed., *Towards Very Large Knowledge Bases: knowledge building and knowledge sharing*. Amsterdam: IOS Press. 60–72.

Bates, J. 1994. The Role of Emotion in Believable Agents. Technical Report CMU-CS-94-136, Carnegie Mellon University, School of Computer Science.

Becker, H., ed. 1974. *The Health Belief Model and Personal Health Behavior*. Thorofare, NJ: C.B. Slack.

Bell, J., and Huang, Z. 1996. Safety logic II: Normative safety. In *Proceedings of the 12th European Conference on Artificial Intelligence*, 293–297. John Wiley and Sons.

Ben-Yami, H. 1997. Against Characterizing Mental States as Propositional Attitudes. *The Philosophical Quarterly* 47(186):84–89.

Bench-Capon, T. 1997. Argument in Artificial Intelligence and Law. *Artificial Intelligence and Law* 5:249–261.

Benerecetti, M.; Bouquet, P.; and Ghidini, C. 2000. Contextual reasoning distilled. *Journal of Experimental and Theoretical Artificial Intelligence* 12(3):279–305.

Berry, D.; Michas, I.; Forster, M.; and Gillie, T. 1997. What Do Patients Want To Know About Their Medicines and What Do Doctors Want To Tell Them?: A Comparative Study. *Psychology and Health* 12:467–480.

Berry, D.; Michas, I.; and de Rosis, F. 1998. Evaluating Explanations About Drug Prescriptions: effects of varying the nature of information about side effects and its relative position in explanations. *Psychology and Health* 13:767–784.

Binsted, K.; Cawsey, A.; and Jones, R. 1995. Generating Personalised Patient Information Using the Medical Record. In Barahona, P.; Stefanelli, M.; and Wyatt, J., eds., *Proceedings of the 5th Conference on Artificial Intelligence in Medicine Europe (AIME95)*, volume 934 of *LNAI*, 29–41. Springer-Verlag.

Blair, J. 1996. The possibility and actuality of visual arguments. *Argumentation and Advocacy* 33:23–39.

Blázquez, M.; Fernández, M.; García-Pinar, J.; and Gómez-Pérez, A. 1998. Building Ontologies at the Knowledge Level using the Ontology Design Environment. In *Proceedings of the 11th Workshop on Knowledge Acquisition, Modeling and Management (KAW'98)*.

Bratman, M. 1987. *Intention, Plans, and Practical Reason*. Cambridge, MA: Harvard University Press.

Brentano, F. 1874-1973. *Psychology from an Empirical Standpoint*, (transl. Rancurello, A.; Terrell, D.; and McAlister, L.). London: Routledge.

Brinberg, D., and Durand, J. 1983. Eating at Fast-Food Restaurants: an Analysis Using Two Behavioral Intention Models. *Journal of Applied Social Psychology* 13(6):459–472.

Brug, J.; Lechner, L.; and de Vries, H. 1995. Psychosocial Determinants of Fruit and Vegetable Consumption. *Appetite* 25:285–296.

Buchanan, B.; Moore, J.; Forsythe, D.; Carenini, G.; Ohlsson, S.; and Banks, G. 1995. An Intelligent Interactive System for Delivering Individualized Information to Patients. *Artifi cial Intelligence in Medicine* 7(2):117–154.

Burton, M., and Brna, P. 1996. Clarissa: an Exploration of Collaboration through Agent-Based Dialogue Games. In Brna, P.; Paiva, A.; and Self, J., eds., *Proceedings of the European Conference on Artifi cial Intelligence in Education*.

Burton, M.; Brna, P.; and Pilkington, R. 2000. Clarissa: a Laboratory for the Modelling of Collaboration. *Journal of Artifi cial Intelligence in Education* 11:79–105.

Cañamero, D., ed. 1998. *Emotional and Intelligent: The Tangled Knot of Cognition. Papers from the 1998 AAAI Fall Symposium*. Menlo Park, CA: AAAI Press.

Campbell, M.; DeVellis, B.; Strecher, V.; Ammermann, A.; DeVellis, R.; and Sandler, R. 1994. Improving Dietary Behaviour: the Efficacy of Tailored Messages in Primary Care Settings. *American Journal of Public Health* 84(5):783–787.

Carberry, S., and Schroeder, L. 2001. Recognising and Conveying Attitude and its Underlying Motivation. In de Rosis, F., ed., *User Modelling 2001: 2nd Workshop on Attitudes, Personality and Emotions in User-Adapted Interaction*.

Carberry, S. 1990. *Plan Recognition in Natural Language Dialogue*. ACL-MIT Press Series in Natural Language Processing. Cambridge, MA: MIT Press.

Carbogim, D.; Robertson, D.; and Lee, J. 2000. Argument-based Applications to Knowledge Engineering. *Knowledge Engineering Review* 15(2):119–149.

Carenini, G., and Moore, J. 2001. An Empirical Study of the Influence of User Tailoring on Evaluative Argument Effectiveness. In Nebel, B., ed., *Proceedings of the 17th International Joint Conference on Artifi cial Intellicenge (IJCAI 2001)*, 1307–1312. Morgan Kaufmann Publishers.

Carenini, G.; Grasso, F.; and Reed, C., eds. 2002. *Proceedings of the ECAI 2002 workshop on Computational Models of Natural Argument*.

Carenini, G.; Mittal, V.; and Moore, J. 1994. Generating Patient Specific Interactive Explanations. In *Proceedings of the 18th Symposium on Computer Applications in Medical Care (SCAMC94)*. McGraw-Hill Inc.

Carenini, G. 2001. GEA: a Complete, Modular System for Generating Evaluative Arguments. In Alexandrov, V.; Dongarra, J.; Juliano, B.; Renner, R.; and K.Tan., eds., *International Conference on Computational Science*, volume 2073 of *LNCS*. Springer-Verlag. Special session on "Computational Models of Natural Argument", F. Grasso and C. Reed, guest eds.

Carletta, J.; Isard, A.; Isard, S.; Kowtko, J.; Doherty Sneddon, G.; and Anderson, A. 1997. The Reliability of a Dialogue Structure Coding Scheme. *Computational Linguistics* 23(1):13–31.

Carletta, J. 1992. *Risk-Taking and Recovery in Task-Oriented Dialogue*. Ph.D. Dissertation, University of Edinburgh, Department of Artificial Intelligence.

Carletta, J. 1996. Assessing Agreement on Classification Tasks: the kappa Statistic. *Computational Linguistics* 22(2):249–254.

Carlson, L. 1983. *Dialogue Games : an Approach to Discourse Analysis*. Synthese Language Library. Dordrecht: D. Reidel.

Carlson, L. 1984. Focus and Dialogue Games: a Game-Theoretical Approach to the Interpretation of Intonal Focusing. In Vaina and Hintikka (1984). 295–334.

Carnap, R. 1947. *Meaning and Necessity*. Chicago: University of Chicago Press.

Castelfranchi, C. 1998. Modelling Social Action for ai Agents. *Artifi cial Intelligence* 103(1-2):157–182.

Castelfranchi, C. 2000. Conflict Ontology. In Müller, H., and Dieng, R., eds., *Computational Conflicts - Conflict modeling for distributed intelligent systems*. Springer. chapter 2, 21–40.

Cattani, A. 1990. *Forme dell'argomentare: il ragionamento fra logica e retorica [Forms of arguing: logical and rhetorical aspects of reasoning]*. Padova: Edizioni GB.

Cawsey, A.; Binsted, K.; and Jones, R. 1995. An On-Line Explanation of the Medical Record to Patients Via an Artificial Intelligence Approach. In Richards, B., ed., *Proceedings of Health Computing 1995*, 269–275. BJHC Limited.

Cawsey, A.; Grasso, F.; and Jones, R. 1999. A Conversational Model for Health Promotion on the World Wide Web. In P.W.Horn et al. (1999), 379–388.

Cawsey, A.; Jones, R.; and Pearson, J. 2000. The evaluation of a personalised information system for patients with cancer. *User Modeling and User-Adapted Interaction* 10(1).

Cawsey, A. 1991. Using plausible inference rules in description planning. In *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics (EACL-91)*, 119–124.

Cawsey, A. 1992. *Explanation and Interaction: the Computer Generation of Explanatory Dialogues*. Cambridge, MA: MIT Press.

Chu-Carroll, J., and Carberry, S. 1994. A Plan-Based Model for Response Generation in Collaborative Task-Oriented Dialogues. In *Proceedings of the 12th National Conference on Artifi cial Intelligence (AAAI94)*, 799–805. AAAI.

Clark, H., and Carlson, T. 1982. Speech Acts and Hearer's Beliefs. In Smith (1982). chapter 1, 1–36.

Clark, H., and Wilkes-Gibbs, D. 1986. Referring as a Collaborative Process. *Cognition* 22:1–39.

Clark, P. 1990. Representing Knowledge as Arguments: Applying Expert System Technology to Judgemental Problem-Solving. In Addis, T., and Muir, R., eds., *Research and Development in Expert Systems*, volume VII. Cambridge University Press. 147–159.

Cohen, P., and Levesque, H. 1990. Intention is Choice with Commitment. *Artificial Intelligence* 42:213–261.

Cohen, P., and Perrault, R. 1979. Elements of a Plan-Based Theory of Speech-Acts. *Cognitive Science* 3:177–212.

Cohen, P. 1978. *On Knowing What to Say: Planning Speech Acts*. Ph.D. Dissertation, University of Toronto. Reproduced as TR 118, University of Toronto, Department of Computer Science.

Cohen, R. 1987. Analyzing the Structure of Argumentative Discourse. *Computational Linguistics* 13(1-2):11–24.

Cole, P., and Morgan, J., eds. 1975. *Speech Acts*, volume 3 of *Syntax and Semantics*. Seminar Press.

Core, M., and Allen, J. 1997. Coding dialogs with the DAMSL annotation scheme. In Traum, D., ed., *AAAI Fall Symposium on Communicative Action in Humans and Machines*.

Coulthard, M. 1985. *An Introduction to Discourse Analysis*. London and New York: Longman, 2 edition.

Coulthard, M., ed. 1992. *Advances in Spoken Discourse Analysis*. London: Routledge.

Crosswhite, J. 2000. Rhetoric and Computation. In Reed and Norman (2000).

Cycorp. 1997. The Upper Cyc Ontology. `http://www.cyc.com/cyc-2-1/cover.html` Release 2.1.

Dalianis, H., and Johannesson, P. 1998. Explaining Conceptual Models - Using Toulmin's argumentation model and RST. In *Proceedings of The Third International workshop on the Language Action Perspective on Communication Modelling (LAP98)*, 131–140.

Damasio, A. 1994. *Descartes' Error*. New York, NY: Putnam & Sons.

Dascal, M. 1979. Conversational Relevance. In Margalit, A., ed., *Meaning and Use*. Reidel: Dordrecht. 72–96.

Davidson, D. 1975. Thought and Talk. In Guttenplan, S., ed., *Mind and Language: Wolfson College Lectures 1974*. Oxford: Oxford University Press. 7–23.

Davis, R.; Shrobe, H.; and Szolovits, P. 1993. What is a Knowledge Representation? *Artificial Intelligence Magazine* 14(1):17–33.

Davis, E. 1990. *Representation of Commonsense Knowledge*. San Mateo, CA: Morgan Kaufmann.

De Carolis, B.; de Rosis, F.; Grasso, F.; Rossiello, A.; Berry, D.; and Gillie, T. 1996. Generating Recipient-Centred Explanations About Drug Prescription. *Artificial Intelligence in Medicine* 8(2):123–145.

de Kleer, J. 1986a. An Assumption-Based Truth Maintenance System. *Artificial Intelligence* 28:127–162.

de Kleer, J. 1986b. Extending the ATMS. *Artificial Intelligence* 28:163–196.

de Kleer, J. 1986c. Problem Solving with the ATMS. *Artificial Intelligence* 28:197–224.

de Kleer, J. 1988. A General Labeling Algorithm for Assumption-Based Truth Maintenance. In *Proceedings of the 7th National Conference on Artificial Intelligence (AAAI88)* (1988), 188–192.

de Kleer, J. 1993. A Perspective on Assumption-Based Truth Maintenance. *Artificial Intelligence* 59:63–67.

de Rosis, F., and Grasso, F. 2000. Affective natural language generation. In Paiva, A., ed., *Affective Interactions*, volume 1814 of *LNAI*. Springer-Verlag.

de Rosis, F.; Grasso, F.; Berry, D.; and Gillie, T. 1995. Mediating Hearer's and Speaker's Views in the Generation of Adaptive Explanations. *Expert Systems With Applications* 8(4):429–443.

de Rosis, F.; Grasso, F.; and Berry, D. 1997. Strengthening Argumentation in Medical Explanations by Text Plan Revision. In Keravnou, E.; Garbay, C.; Baud, R.; and Wyatt, J., eds., *Proceedings of the 6th Conference on Artificial Intelligence in Medicine Europe (AIME97)*, volume 1211 of *LNAI*, 349–360. Springer-Verlag.

de Sousa, R., and Morton, A. 2002. Emotional Truth. In *Proceedings of the Aristotelian Society, Supp.*, volume 76, 247–263.

Dennett, D. 1978. *Brainstorms*. Montgomery, Vt: Bradford.

Department of Health. 1992. The Health of the Nation: a Strategy for Health in England. London: HMSO.

Descartes, R. 1931. *Meditations on First Philosophy*, (transl. Haldane, E., and Ross, G.). Cambridge: Cambridge University Press.

Dowling, W., and Gallier, J. 1984. Linear Time for Testing the Satisfiability of Propositional Horn Formulae. *Journal of Logic Programming* 1(3):267–270.

Doyle, J. 1979. A Truth Maintenance System. *Artificial Intelligence* 12(2):231–272.

Doyle, J. 1992. Reason Maintenance and Belief Revision: Foundation vs. Coherence Theories. In Gärdenfors (1992a). 29–51.

Dubois, D., and Prade, H. 1992. Belief Change and Possibility Theory. In Gärdenfors (1992a). 142–182.

Dung, P. 1995. On the Acceptability of Arguments and its Fundamental Role in Nonmonotonic Reasoning, Logic Programming and n-Person Games. *Artificial Intelligence* 77(2):321–357.

Elhadad, M. 1995. Using Argumentation in Text Generation. *Journal of Pragmatics* 24:189–220.

Elhadad, M., ed. 2000. *Proceedings of the 1st International Conference on Natural Language Generation (INLG-2000)*.

Eliasmith, C. 2002. The myth of the Turing machine: the failings of functionalism and related theses. *Journal of Experimental and Theoretical Artificial Intelligence* 14:1–8.

Elliott, C. 1992. *The Affective Reasoner: A process model of emotions in a multi-agent system*. Ph.D. Dissertation, Northwestern University.

Fagerli, R., and Wandel, M. 1999. Gender Differences in Opinions and Practices with Regard to a "Healthy Diet". *Appetite* 32:171–190.

Fagin, R.; Halpern, J.; Moses, Y.; and Vardi, M. 1996. *Reasoning about Knowledge*. London: MIT Press.

Fahnestock, J. 1999. *Rhetorical Figures in Science*. New York, NY: Oxford University Press.

Fauconnier, G. 1985. *Mental Spaces: Aspects of Meaning Construction in Natural Language*. Cambridge, MA: MIT Press.

Fernández, M. 1999. Overview of Methodologies for Building Ontologies. In Benjamins, V.; Chandrasekan, B.; Gómez-Pérez, A.; Guarino, N.; and Uschold, M., eds., *Proceedings of the IJCAI-99 Workshop on Ontologies and Problem Solving Methods*.

Feteris, E. 2000. A Dialogical Theory of Legal Discussions: Pragma-Dialectical Analysis and Evaluation of Legal Argumentation. *Artificial Intelligence and Law* 8:115–135.

Fodor, J., ed. 1975. *The Language of Thought*. New York, NY: Thomas Y. Crowell.

Fodor, J. 1978. Propositional attitudes. *The Monist* LXI(4):501–523.

Forbus, K., and de Kleer, J. 1988. Focusing the ATMS. In *Proceedings of the 7th National Conference on Artificial Intelligence (AAAI88)* (1988), 193–198.

Forsythe, D. 1995. Using Ethnography in the Design of an Explanation System. *Expert Systems With Applications* 8(4):403–417.

Fox, M., and Long, D. 1995. Hierarchical Planning Using Abstraction. *IEE Proceedings - Control Theory and Applications* 142(3):197–210.

Fox, J., and Parsons, S. 1997. On using Arguments for Reasoning about Actions and Values. In *Proceedings of the AAAI Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning*.

Fox, J., and Parsons, S. 1998. Arguing about Beliefs and Actions. In Hunter, A., and Parsons, S., eds., *Applications of Uncertainty Formalisms*. Springer Verlag.

Freedman, R. 1996. *Interaction of Discourse Planning, Instructional Planning and Dialogue Management in an Interactive Tutoring System*. Ph.D. Dissertation, Computer Science Department, Northwestern University.

Freeman, K., and Farley, A. 1996. A model of Argumentation and its Application to Legal Reasoning. *Artificial Intelligence and Law* 4(3/4):163–197.

Friedman, N., and Halpern, J. 1999. Modeling Beliefs in Dynamic Systems. Part II: Revision and Update. *Journal of Artificial Intelligence Research* 10:117–167.

Fries, E., and Croyle, R. 1993. Stereotypes Associated with a Low-Fat Diet and their Relevance to Nutrition Education. *Journal of the American Dietetic Association* 93(5):551–555.

Frijda, N.; Manstead, A.; and Bem, S., eds. 2000. *Emotions and Beliefs: How Feelings Influence Thoughts*. Paris: Maison des Science de l'Homme and Cambridge University Press.

Frohlich, D., and Luff, P. 1990. Applying the technology of conversation to the technology for conversation. In Luff, P.; Gilbert, N.; and Frohlich, D., eds., *Computers and Conversation*. Academic Press. chapter 9, 187–220.

Galliers, J. 1992. Autonomous Belief Revision and Communication. In Gärdenfors (1992a). 220–246.

Garagnani, M. 1997. Belief modelling for discourse plans. In *Proceedings of the 16th Workshop of the UK Planning and Scheduling SIG*.

Garavelli, B. 1990. *Manuale di Retorica [Handbook of Rhetoric]*. Milano: Bompiani.

Genesereth, M., and Nilsson, N. 1987. *Logical Foundations of Artificial Intelligence*. San Mateo CA: Morgan Kaufmann Publishers.

Gernsbacher, M., and Derry, S., eds. 1998. *Proceedings of the 20th Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Gilbert, M. 1995. Coalescent Argumentation. *Argumentation* 9(5):837–852.

Gilbert, M. 1997. Prolegomenon to a Pragmatics of Emotion. In *Proceedings of the 2nd International Conference of the Ontario Society for the Study of Argumentation (OSSA'97)*.

Giunchiglia, F., and Bouquet, P. 1998. A context based framework for mental representation. In Gernsbacher and Derry (1998), 392–397.

Gärdenfors, P., ed. 1992a. *Belief Revision*, volume 29 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press.

Gärdenfors, P. 1992b. Belief Revision: an Introduction. In *Belief Revision* (1992a). 1–20.

Gärdenfors, P. 1992c. The dynamics of knowledge and belief: Foundational vs. coherence theories. In Bicchieri, C., and dalla Chiara, M. L., eds., *Knowledge, Belief and Strategic Interaction*. Cambridge University Press. 377–396.

Gómez-Pérez, A.; Arpírez, R.; Corcho, O.; Ding, Y.; Fernández-López, M.; Manzano, D.; and Suárez-Figueroa, M. 2002. State of the Art in Ontologies from the Semantic Web perspective. Deliverable D1.1, Esperonto Services (IST-2001-34373).

Gómez-Pérez, A.; Fernández-López, M.; and Corcho, O. 2002. Ontoweb: Technical roadmap. Deliverable D1.1.2, OntoWeb (IST-2000-29243).

Gordon, T. 1995. *The Pleadings Game, An Artificial Intelligence Model of Procedural Justice*. Dordrecht: Kluwer Academic Publishers.

Grassi, E. 1980. *Rhetoric as Philosophy: The Humanist Tradition*. University Park and London: Pennsylvania State University Press.

Greiner, R., and Subramanian, D., eds. 1995. *Proceedings of the AAAI Fall Symposium on Relevance*. Menlo Park, CA: AAAI. Tech. Rept. FS-94-02.

Grice, H. 1957. Meaning. *Philosophicsl Review* 66:377–388.

Grice, H. 1975. Logic and Conversation. In Cole and Morgan (1975). 41–58.

Groarke, L.; Tindale, C.; and Fisher, L. 1997. *Good Reasoning Matters!* Toronto: Oxford University Press.

Groarke, L. 1996. Logic, art and argument. *Informal Logic* 18:105–129.

Groarke, L. 1997. Informal Logic. Stanford Encyclopedia of Philosophy, http://www.plato.stanford.edu.

Grosz, B., and Sidner, C. 1986. Attention, Intentions and the Structure of Discourse. *Computational Linguistics* 12(3):175–204.

Grosz, B.; Joshi, A.; and Weinstein, S. 1995. Centering: a Framework for Modelling Local Coherence of Discourse. Technical Report IRCS 95-01, University of Pennsylvania, Department of Computer Science.

Grove, A. 1988. Two modellings for theory change. *Journal of Philosophical Logic* 17:157–170.

Gruber, T. 1992. Ontolingua: A Mechanism to Support Portable Ontologies. Technical Report KSL-91-66, Stanford University, Knowledge Systems Laboratory.

Gruber, T. 1994. Towards Principles for the Design of Ontologies used for Knowledge Sharing. In Guarino and Poli (1994).

Grueninger, U. 1995. Arterial hypertension: lessons from patient education. *Patient Education and Counseling* 26:37–55.

Gruninger, M., and Fox, M. 1995. Methodology for the Design and Evaluation of Ontologies. In *IJCAI-95 Workshop on Basic Ontological Issues in Knowledge Sharing*.

Guarino, N., and Poli, R., eds. 1994. *Formal Ontology in Conceptual Analysis and Knowledge Representation*. Boston, MA: Kluwer.

Guarino, N., and Welty, C. 2002. Evaluating Ontological Decisions with OntoClean. *Communications of the ACM* 45(2):61–65.

Guarino, N. 1995. Formal Ontology, Conceptual Analysis and Knowledge Representation. *International Journal of Human-Computer Studies* 43(5/6):625–640.

Guarino, N. 1997. Understanding, Building, and Using Ontologies. *International Journal of Human-Computer Studies* 46(2/3):293–310.

Hage, J. 1997. *Reasoning with Rules*. Dordrecht: Kluwer Academic Publishers.

Hage, J. 2000. Dialectical Models in Artificial Intelligence and Law. *Artifi cial Intelligence and Law* 8:137–172.

Hamblin, C. 1970. *Fallacies*. London: Methuen.

Hamblin, C. 1971. Mathematical Models of Dialogue. *Theoria* 37:130–155.

Hintikka, J. 1962. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Ithaca, New York: Cornell University Press.

Hintikka, J. 1984. Rules, Utilities and Strategies in Dialogical Games. In Vaina and Hintikka (1984). 277–294.

Hirst, G.; DiMarco, C.; Hovy, E.; and Parsons, K. 1997. Authoring and Generating Health-Education Documents that are Tailored to the Needs of the Individual Patient. In Jameson, A.; Paris, C.; and Tasso, C., eds., *User Modeling - Proceedings of the 6th International Conference*, 107–118. SpringerWien.

Hovy, E. 1988. *Generating Natural Language under Pragmatic Constraints*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Hovy, E. 1990. Pragmatics and Natural Language Generation. *Artifi cial Intelligence* 43:153–197.

Hovy, E. 1993. Automated Discourse Generation using Discourse Structure Relations. *Artifi cial Intelligence* 63(1-2):341–385.

Huang, Z., and Bell, J. 1996. Safety logic I: Absolute safety. In *Proceedings of Commonsense '96*, 59–66.

Huang, X. 1994. Planning Argumentative Texts. In *Proceedings of the 17th International Conference on Computational Linguistics (COLING94)*.

Hume, D. 1758. *Enquiry Concerning the Human Understanding*. New York: Liberal Arts Press.

Jameson, A.; Kipper, B.; Ndiaye, A.; Schäfer, R.; Simons, J.; Weis, T.; and Zimmermann, C. 1994. Cooperating To Be Noncooperative: the Dialog System PRACMA. In Nebel, B., and Dreschler-Fischer, L., eds., *KI-94: Advances in Artificial Intelligence, Proceedings of the 18th German Conference on Artificial Intelligence*, 106–117. Heidelberg: Springer.

Johnson, R. H., and Blair, J. A. 1994a. *Logical self-defense*. Toronto: McGraw-Hill, 3rd edition.

Johnson, R. H., and Blair, J. A., eds. 1994b. *New Essays in Informal Logic*. Windsor, Ontario: Informal Logic.

Johnson-Laird, P. 1983. *Mental Models: towards a cognitive science of language, inference and consciousness*. Cambridge University Press.

Johnson, C. G., ed. 2001. *Symposium on Emotion, cognition, and affective computing*, AISB'01 Convention.

Jokinen, K.; Maybury, M.; Zock, M.; and Zukerman, I., eds. 1996. *Proceedings of the ECAI-96 Workshop on: Gaps and Bridges: New directions in Planning and NLG*.

Jones, A., and Sergot, M. 1993. On the Characterisation of Law and Computer Systems: The Normative Systems Perspective. In Meyer, J.-J., and Wieringa, R., eds., *Deontic Logic in Computer Science: Normative System Specification*. John Wiley and Sons. chapter 12.

Jones, D.; Bench-Capon, T.; and Visser, P. 1998. Methodologies for Ontology Development. In *Proc. IT&KNOWS Conference of the 15th IFIP World Computer Congress*. Chapman-Hall.

Kamlah, W., and Lorenzen, P. 1984. *Logical Propaedeutic: Pre-school of reasonable discourse*. Lanham, MD: University Press of America.

Kaplan, D., and Montague, R. 1960. A Paradox Regained. *Notre Dame Journal of Formal Logic* 1(3):79–90.

Karacapilidis, N., and Papadias, D. 1998. A Computational Approach for Argumentative Discourse in Multi-Agent Decision Making Environments. *AI Communications* 11(1):21–33.

Karacapilidis, N., and Papadias, D. 2001. Computer Supported Argumentation and Collaborative Decision Making: the hermes System. *Information Systems* 26:259–277.

Karacapilidis, N.; Trousse, B.; and Papadias, D. 1997. Using Case-Based Reasoning for Argumentation with Multiple Viewpoints. In Leake, D., and Plaza, E., eds., *Case-Based Reasoning: Research and Development, Proceedings of the 2nd Intern. Conference on Case-Based Reasoning (ICCBR'97)*, volume 1266 of *LNAI*, 541–552. Springer-Verlag.

Karacapilidis, N. 1996. An Argumentation Based Framework for Defeasible and Qualitative Reasoning. In Jokinen et al. (1996), 37–42.

KBSI. 1994. The IDEF5 Ontology Description Capture Method Overview. Technical report, KBSI, Texas.

Kelly, T. 2002. The Rationality of Belief and some other Propositional Attitudes. *Philosophical Studies* 110:163–196.

Knott, A., and Sanders, T. 1998. The Classification of Coherence Relations and their Linguistic Markers: An exploration of two languages. *Journal of Pragmatics* 30:135–175.

Kobsa, A. 1990. Modeling the User's Conceptual Knowledge in BGP-MS, a User Modeling Shell System. *Computational Intelligence* 6:193–208.

Konolige, K. 1982. A First Order Formalization of Knowledge and Action for a Multi-Agent Planning System. In Hayes, J.; Michie, D.; and Pao, Y.-H., eds., *Machine Intelligence 10*. Chichester, England: Ellis Horwood.

Kowalski, R., and Toni, F. 1996. Abstract Argumentation. *Artifi cial Intelligence and Law* 4(3/4):275–296.

Kowtko, J.; Isard, S.; and Doherty, G. 1992. Conversational Games Within Dialogue. Technical Report HCRC/RP-31, University of Edinburgh, Human Comunication Research Centre.

Kraus, S.; Sycara, K.; and Evenchik, A. 1998. Reaching Agreements through Argumentation: a logical model and implementation. *Artifi cial Intelligence* 104:1–69.

Krause, P.; Ambler, S.; Elvang-Goransson, M.; and Fox, J. 1995. A Logic of Argumentation for Reasoning Under Uncertainty. *Computational Intelligence* 11(1):113–131.

Kripke, S. 1963. Semantical Condiserations on Modal Logics. *Acta Philosophica Fennica* 16:83–94.

Lajoie, S.; Greer, J.; Munsie, S.; Wilkie, T.; Guerrara, C.; and Aleong, P. 1995. Establishing an Argumentation Environment to Foster Scientific Reasoning with Bio-World. In *Proceedings of the International Conference on Computers in Education (ICCE)*, 89–96.

Lakoff, G., and Johnson, M. 1980. *Metaphors We Live By*. Chicago: The University of Chicago Press.

Lambert, L., and Carberry, S. 1991. A Tripartite Plan-based Model of Dialogue. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (ACL91)*, 47–54.

Lascarides, A., and Asher, N. 1993. Temporal Interpretation, Discourse Relations and Common Sense Entailment. *Linguistics and Philosophy* 16(5):437–493.

Lascarides, A., and Asher, N. 1999. Cognitive States, Discourse Structure and the Content of Dialogue. In *Amstelogue: 3rd Workshop on the Semantics and Pragmatics of Dialogue*.

Lee, M., and Wilks, Y. 1997. Eliminating Deceptions and Mistaken Belief to Infer Conversational Implicature. In Jokinen, K.; Sadek, D.; and Traum, D., eds., *IJCAI 1997 Workshop on Collaboration, Cooperation and Conflict in Dialogue Systems*.

Levin, J., and Moore, J. 1977. Dialogue-Games: Metacommunication Structures for Natural Language Interaction. *Cognitive Science* 1:395–420.

Lewis, D. 1969. *Convention*. Cambridge, MA: Harvard University Press.

Lewis, D. 1973. Counterfactual and comparative possibility. *Journal of Philosophical Logic* 2.

Lo Cascio, V. 1991. *Grammatica dell'Argomentare: strategie e strutture [A grammar of Arguing: strategies and structures]*. Firenze: La Nuova Italia.

Lodder, A., and Herczog, A. 1995. DiaLaw: A dialogical framework for modeling legal reasoning. In *Proceedings of the 5th International Conference on Artificial Intelligence and Law*, 146–155. New York, NY: ACM Press.

Logan, B.; Reece, S.; Cawsey, A.; Galliers, J.; and Sparck Jones, K. 1994. Belief Revision and Dialogue Management in Information Retrieval. Technical Report 339, University of Cambridge, Computer Laboratory.

Loui, R. 1993. Process and Policy: Resource-Bounded non-Demonstrative Reasoning. *Computational Intelligence* 2(5).

Loui, R. 1994. Argument and Arbitration Games. In *Working Notes of the Workshop on Computational Dialectics,12th National Conference on Artificial Intelligence (AAAI94)*, 72–83.

Lowe, D. 1985. Co-operative Structuring of Information: the Representation of Reasoning and Debate. *International Journal of Man-Machine Studies* 23:97–111.

Mackenzie, J. D. 1979. Question-Begging in Non-Cumulative Systems. *Journal of Philosophical Logic* 8:117–133.

Mackenzie, J. D. 1990. Four Dialogue Systems. *Studia Logica* XLIX(4):567–583.

Mann, W., and Thompson, S. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text* 8(3):243–281.

Mann, W.; Moore, J.; and Levin, J. 1977. A Comprehension Model for Human Dialogue. In *Proceedings of the 5th International Joint Conference of Artificial Intelligence (IJCAI-77)*, 77–87.

Mann, W. 1988. Dialogue Games: Conventions of Human Interactions. *Argumentation* 2:511–532.

Mann, W. 2002a. Analyzing Dialogues under Dialogue Macrogame Theory. http://www-rcf.usc.edu/ billmann/dialogue/dganalysis.htm.

Mann, W. 2002b. Dialogue analysis for diverse situations. In Bos, J.; Foster, M.; and Matheson, C., eds., *EDILOG'02: Proceedings of the 6th workshop on the semantics and pragmatics of dialogue*, 109–116.

Mann, W. 2002c. Dialogue Macrogame Theory. In Jokinen, K., and McRoy, S., eds., *3rd SIGdial Workshop on Discourse and Dialogue*.

Marcu, D. 1996a. Building up Rhetorical Structure Trees. In *Proceedings of the 13th National Conference on Artificial Intelligence (AAAI96)*, volume 2, 1069–1074. AAAI press / MIT press.

Marcu, D. 1996b. The Conceptual and Linguistic Facets of Persuasive Arguments. In Jokinen et al. (1996), 43–46.

Marcu, D. 1997. *The Rhetorical Parsing, Summarization, and Generation of Natural Language Texts*. Ph.D. Dissertation, University of Toronto, Department of Computer Science.

Marcu, D. 2000a. Extending a formal and computational model of rhetorical structure theory with intentional structures à la grosz and sidner. In *The 18th International Conference on Computational Linguistics COLING'2000*.

Marcu, D. 2000b. *The Theory and Practice of Discourse Parsing and Summarization*. Cambridge, MA: MIT Press.

Marshall, C. 1989. Representing the Structure of Legal Argument. In *Proceedings of the 2nd International Conference on Artificial Intelligence and Law*, 121–127. ACM Press.

Maudet, N., and Moore, D. 1999. Dialogue games for computer supported collaborative argumentation. In Hoadley, C., and Roschelle, J., eds., *Proceedings of the Computer Support for Collaborative Learning (CSCL) 1999 Conference*. Mahwah, NJ: Lawrence Erlbaum Associates.

Maybury, M. 1992. Communicative Acts for Explanation Generation. *International Journal of Man-Machine Studies* 37:135–172.

Maybury, M. 1993. Communicative Acts for Generating Natural Language Arguments. In *Proceedings of the 11th National Conference on Artificial Intelligence (AAAI93)*, 357–364. AAAI.

McBurney, P., and Parsons, S. 2001. Agent ludens: Games for agent dialogues. In Gmytrasiewicz, P., and Parsons, S., eds., *Proceedings of the Third Workshop on Game-Theoretic and Decision-Theoretic Agents (GTDT2001)*, AAAI Spring Symposium. Menlo Park, CA: AAAI Press.

McBurney, P. 2002. *Rational Interaction*. Ph.D. Dissertation, University of Liverpool, Department of Computer Science.

McCoy, K. F. 1989. Generating Context Sensitive Responses to Object-Related Misconceptions. *Artificial Intelligence* 41(2):157–195.

McDermott, D. 1981. Artificial Intelligence Meets Natural Stupidity. In Haugeland, J., ed., *Mind Design*. Cambridge, MA: MIT Press. chapter 5, 143–160.

McKay, T. 2000. Propositional attitude reports. In Zalta, E., ed., *The Stanford Encyclopedia of Philosophy*. Spring 2000 edition. (http://plato.stanford.edu/archives/spr2000/entries/prop-attitude-reports/).

McKeown, K. 1985. *Text Generation: Using Discourse Strategy and Focus Constraints to Generate Natural Language Texts*. Studies in Natural Language Processing. Cambridge University Press.

Miksch, S.; Cheng, K.; and Hayes-Roth, B. 1996. The Patient Advocate: a Cooperative Agent to Support Patient-Centered Needs and Demands. In *AMIA Annual Fall Symposium*.

Miller, G. 1995. WORDNET: A Lexical Database for English. *Communications of ACM* 11:39–41.

Millikan, R. 1986. Thought without Laws: Cognitive Science with Content. *Philosophical Review* 95.

Minsky, M. 1985. *The Society of Mind*. New York, NY: Simon and Schuster.

Montague, R. 1963. Syntactical treatment of modality, with corollaries on reflexion principles and finite axiomatizations. *Acta Philosophica Fennica* 16:153–167.

Moore, J., and Paris, C. 1992. Exploiting User Feedback to Compensate for the Unreliability of User Models. *User Modeling and User-Adapted Interaction* 2(4).

Moore, J., and Paris, C. 1993. Planning Text for Advisory Dialogues: Capturing Intentional and Rhetorical Information. *Computational Linguistics* 19(4):651–695.

Moore, J., and Pollack, M. 1992. A Problem for RST: The Need for Multi-Level Discourse Analysis. *Computational Linguistics* 18(4):537–544.

Moore, R. 1985. A Formal Theory of Knowledge and Action. In Hobbs, J., and Moore, R., eds., *Formal Theories of the Commonsense World*. Norwood, New Jersey: Ablex Publishing Corporation. chapter 9, 319–358.

Morreau, M., and Kraus, S. 1998. Synctactical Treatment of Propositional Attitudes. *Artifi cial Intelligence* 106:161–177.

μ Groupe. 1976. *Rhètorique Générale*. Larousse. (transl. as *A General Rhetoric*, P.B. Burrel and E.M. Slotkin tr., Johns Hopkins University Press, 1981).

Murphy, J. 1974. *Rhetoric in the Middle Ages*. Berkeley-Los Angeles-London: University of California Press.

Nagel, T. 1974. What is it like to be a bat? *The Philosophical Review* LXXXIII:435–450.

Nebel, B. 1990. *Reasoning and Revision in Hybrid Representation Systems*. Number 422 in Lecture Notes in Computer Science. Springer-Verlag.

Neches, R.; Fikes, R.; Finin, T.; Gruber, T.; Patil, R.; Senator, T.; and Swartout, W. 1991. Enabling technology for knowledge sharing. *Artifi cial Intelligence Magazine* 12(3):36–56.

Ortony, A.; Clore, G.; and Collins, A. 1988. *The Cognitive Structure of Emotions*. Cambridge University Press.

Parsons, S., and Jennings, N. 1996. Negotiation Through Argumentation - a Preliminary Report. In *Proceedings of the 2nd International Conference on Multi-Agent Systems (ICMAS-96)*.

Parsons, S.; Sierra, C.; and Jennings, N. 1998. Agents That Reason and Negotiate by Arguing. *Journal of Logic and Computation* 8(3):261–292.

Parsons, S.; Wooldridge, M.; and Amgoud, L. 2002. An Analysis of Formal Inter-Agent Dialogues. In Castelfranchi, C., and Johnson, W., eds., *Proceeding of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'02)*, 394–401. ACM Press.

Parsons, S. 1996. Comparing Normative Argumentation to Other Probabilistic Systems. In *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*.

Paulson, L. 1991. *ML for the working programmer*. Cambridge University Press.

Pearl, J. 1991. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, 2nd printing, revised edition.

Pera, M. 1991. *Scienza e Retorica [Science and Rhetoric]*. Bari: Laterza. (Translated as "The Discourse of Science", The University of Chicago Press, 1994).

Perelman, C., and Olbrechts-Tyteca, L. 1958. *La Nouvelle Rhétorique: Traité de l'argumentation*. Paris: Presses Universitaires de France.

Perelman, C., and Olbrechts-Tyteca, L. 1969. *The New Rhetoric: a treatise on argumentation*. Notre Dame, Indiana: University of Notre Dame Press.

Picard, R. 1996. Does HAL Cry Digital Tears?: Emotion and Computers. In Stork, D., ed., *HAL's Legacy - 2001's Computer as Dream and Reality*. Cambridge, MA: MIT Press. chapter 13, 279–303.

Pilkington, R.; Hartley, J.; Hintze, D.; and Moore, D. 1992. Learning to Argue and Arguing to Learn: An Interface for Computer-Based Dialogue Games. *Journal of Artifi cial Intelligence in Education* 3(3):275–285.

Potts, H., and Wardle, J. 1998. The List of Heuristics for Studying Personality Correlates of Food Choice Behaviour: a Review and Result from Two Samples. *Appetite* 30:79–82.

Prakken, H. 1997. *Logical Tools for Modelling Legal Argument: A Study of Defeasible Reasoning in Law*. Dordrecht: Kluwer Academic Publishers.

Prochaska, J., and Clemente, C. D. 1992. Stages of Change in the Modification of Problem Behavior. In Hersen, M.; Eisler, R.; and Miller, P., eds., *Progress in Behavior Modifi cation*, volume 28. Sycamore, IL: Sycamore Publishing Company.

Putnam, H. 1975. *Mind, Language and Reality*. Cambridge: Cambridge University Press.

P.W.Horn; Shahar, Y.; Lindberg, G.; S.Andreassen; and Wyatt, J., eds. 1999. *Artifi cial Intelligence in Medicine. Proceedings of the Joint European Conference on Artifi cial Intelligence in Medicine and Medical Decision Making, AIMDM'99*, volume 1620 of *LNAI*. Springer-Verlag.

Quine, W. 1966. *The Ways of Paradox and Other Essays*. Cambridge, MA: Harvard University Press.

Rao, A., and Georgeff, M. 1991. Modeling Rational Agents within a BDI-Architecture. In Allen, J.; Fikes, R.; and Sandewall, E., eds., *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning (KR'91)*, 473–484. Morgan Kaufmann Publishers.

Ravenscroft, A. 2000. Designing Argumentation for Conceptual Development. *Computers and Education* 34:241–255.

Read, J. 1993. Coding and classification in health and medical care systems. In Hedley, A., ed., *Proceedings of the 2nd Hong Kong (Asia-Pacifi c) Medical Informatics Conference*, 175–180. Hong Kong: Hong Kong Society of Medical Informatics Ltd.

Rector, A.; Solomon, W.; Nowlan, W.; and Rush, T. 1995. A terminology server for medical language and medical information systems. *Methods of Information in Medicine* 34:147–157.

Reed, C., and Long, D. 1997. Content Ordering in the Generation of Persuasive Discourse. In *Proceedings of the 15th International Joint Conference of Artifi cial Intelligence (IJCAI-97)*. IJCAI.

Reed, C., and Long, D. 1998. Generating the structure of argument. In *Proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics (COLING-ACL'98)*, 1091–1097.

Reed, C., and Norman, T., eds. 2000. *Symposium on Argument and Computation: position papers*. `http://www.csd.abdn.ac.uk/˜tnorman/sac/`.

Reed, C.; Long, D.; Fox, M.; and Garagnani, M. 1996. Persuation as a Form of Inter-Agent Negotiation. In *Working Notes of the 2nd Australian Workshop on Distributed AI*.

Reed, C.; Grasso, F.; and Carenini, G., eds. 2003. *Proceedings of the IJCAI 2003 workshop on Computational Models of Natural Argument*.

Reed, C.; Long, D.; and Fox, M. 1996. An Architecture for Argumentative Dialogue Planning. In Gabbay, D., and Ohlbach, H., eds., *Practical Reasoning: Proceedings of the First International Conference on Formal and Applied Practical Reasoning (FAPR'96)*, number 1085 in Lecture Notes in Computer Science, 555–566. Springer-Verlag.

Reed, C. 1997. Representing and applying knowledge for argumentation in a social context. *AI and Society* 11(3/4):138–154.

Reed, C. 1998a. Dialogue frames in agent communication. In Demazeau, Y., ed., *Proceedings of the 3rd International Conference on Multi-Agent Systems (ICMAS'98)*, 246–253. IEEE Press.

Reed, C. 1998b. *Generating Arguments in Natural Language*. Ph.D. Dissertation, University College London.

Reed, C. 1999. The Role of Saliency in Generating Natural Language Arguments. In Dean, T., ed., *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence (IJCAI'99)*, 876–881. Morgan Kaufmann Publishers.

Reiter, E., and Dale, R. 2000. *Building Natural Language Generation Systems*. Cambridge: Cambridge University Press.

Reiter, E., and Osman, L. 1997. Tailored Patient Information: Some Issues and Questions. In *Proceedings of the ACL-1997 Workshop 'From Research to Commercial Applications: Making NLP Technology Work in Practice'*, 29–34.

Reiter, E.; Cawsey, A.; Osman, L.; and Roff, Y. 1997. Knowledge Acquisition for Content Selection. In *Proceedings of the 6th European Workshop on Natural Language Generation*, 117–126.

Reiter, E.; Robertson, R.; and Osman, L. 1999. Types of Knowledge Required to Personalize Smoking Cessation Letters. In P.W.Horn et al. (1999), 389–399.

Reiter, E.; Robertson, R.; and Osman, L. 2000. Knowledge Acquisition for Natural Language Generation. In Elhadad (2000), 217–224.

Reiter, E.; Robertson, R.; and Osman, L. 2003. Lessons from a failure: Generating tailored smoking cessation letters. *Artificial Intelligence* 144:41–58.

Richard, M. 1990. *Propositional Attitudes: an essay on thoughts and how we ascribe them*. Cambridge Studies in Philosophy. Cambridge University Press.

Richards, T. 1989. *Clausal Form Logic: An Introduction to the Logic of Computer Reasoning*. Addison-Wesley.

R.M. Chisholm. 1957. *Perceiving: A Philosophical Study*. Cornell University Press.

Roe, L.; Hunt, P.; Bradshaw, H.; and Rayner, M. 1997. *Health Promotion Interventions to Promote Healthy Eating in the General Population: a review*. Health Education Authority - The Cromwell Press.

Rosenthal, D. 1991. *The Nature of Mind*. Oxford University Press.

Rumbaugh, J.; Blaha, M.; Premerlani, W.; Eddy, F.; and Lorensen, W. 1991. *Object-Oriented Modeling and Design*. Prentice-Hall.

Russell, S., and Norvig, P. 2003. *Artificial Intelligence: A Modern Approach*. Prentice Hall, 2nd edition.

Russell, B. 1912. *The Problems of Philosophy*. New York, NY: Oxford University Press.

Russell, B. 1921. *The Analysis of Mind*. London: George Allen & Unwin.

Russell, B. 1948. *Human Knowledge: Its Scope and Limits*. London: George Allen & Unwin.

Sacerdoti, E. 1974. Planning in a Hierarchy of Abstraction Spaces. *Artificial Intelligence* 5:115–135.

Sacerdoti, E. 1977. *A Structure for Plans and Behavior*. Amsterdam: North-Holland.

Sacks, H.; Schegloff, E.; and Jefferson, G. 1974. A simplest systematics for the organization of turn taking for conversation. *Language* 50(4):696–735.

Sadalla, E., and Burroughs, J. 1981. Profiles in Eating: Sexy Vegetarians and Other Diet-Based Social Stereotypes. *Psychology Today* 15(10):51–57.

Saeedi, M., and Sillince, J. 1999. Incorporating rhetorical and plausible reasoning in a system for simulating argumentation. *Knowledge Based Systems* 12:113–127.

Sartor, G. 1994. A Formal Model of Legal Argumentation. *Ratio Juris* 7:177–211.

Schiffer, S. 1972. *Meaning*. Oxford University Press.

Searle, J. 1969. *Speech Acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Searle, J. 1975. Indirect Speech Acts. In Cole and Morgan (1975). 59–82.

Shafer, G. 1976. *A Mathematical Theory of Evidence*. Princeton NJ: Princeton University Press.

Shoham, Y., and Cousins, S. 1994. Logics of Mental Attitudes in AI, a Very Preliminary Survey. In Lakemeyer, G., and Nebel, B., eds., *Foundations of Knowledge Representation and Reasoning*, volume 810 of *LNAI*. Springer-Verlag.

Sillince, J. 1994. Multi-agent Conflict Resolution: a Computational Framework for an Intelligent Argumentation Program. *Knowledge-Based Systems* 7(2):75–90.

Sinclair, J., and Coulthard, M. 1975. *Towards an Analysis of Discourse: The English used by teachers and pupils*. Oxford University Press.

Skinner, C.; Siegfried, J.; Kegler, M.; and Strecher, V. 1993. The potential of computers in patient education. *Patient Education and Counseling* 22(1):27–34.

Sloman, A. 1999. Beyond Shallow Models of Emotion. In *i3 Spring Days Workshop on Behaviour Planning for Life-Like Characters and Avatars*.

Smith, N., ed. 1982. *Mutual Knowledge*. Academic Press.

Sowa, J. 2000. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Brooks/Cole.

Sperber, D., and Wilson, D. 1990. Rhetoric and relevance. In Wellbery, D., and Bender, J., eds., *The Ends of Rhetoric: History, Theory, Practice*. Stanford: Stanford University Press. 140–155.

Sperber, D., and Wilson, D. 1995. *Relevance: communication and cognition*. Oxford: Blackwell, 2 edition.

Sperber, D. 1997. Intuitive and reflective beliefs. *Mind and Language* 12(1):67–83.

Stalnaker, R. 1974. Pragmatic presuppositions. In Munitz, M., and Unger, P., eds., *Semantics and Philosophy*. New York, NY: New York University Press. 197–213.

Stefik, M. 1981. Planning and Meta-Planning (MOLGEN: Part 2). *Artifi cial Intelligence* 16:141–170.

Stent, A., and Allen, J. 2000. Annotating Argumentation Acts in Spoken Dialog. Technical Report 740, The University of Rochester, Computer Science Department. (TRAINS Technical Note 00-1).

Studer, R.; Benjamins, V.; and Fensel, D. 1998. Knowledge engineering, principles and methods. *Data and Knowledge Engineering* 25(1-2):161–197.

Swales, J. 1990. *Genre Analysis: English in academic and research settings*. Cambridge University Press.

Sycara, K. 1990. Persuasive Argumentation in Negotiation. *Theory and Decision* 28(3):203–242.

Tamma, V., and Bench-Capon, T. 2002. Attribute meta-properties for formal ontological analysis. In Gómez-Pérez, A., and Benjamins, R., eds., *Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management, EKAW 2002*, volume 2473 of *LNAI*, 301–316. Springer-Verlag.

Taylor, G., and Whitehill, S. 1981. A Belief Representation for Understanding Deception. In P.J.Hayes., ed., *Proceedings of the 7th International Joint Conference of Artificial Intelligence (IJCAI-81)*, 388–393. William Kaufmann.

Taylor, J.; Carletta, J.; and Mellish, C. 1996a. Combining Power with Tractability in Belief Models. Technical Report HCRC/RP-76, University of Edinburgh, Human Comunication Research Centre.

Taylor, J.; Carletta, J.; and Mellish, C. 1996b. Requirements for Belief Models in Cooperative Dialogue. *User Modeling and User-Adapted Interaction* 6(1):23–68.

Taylor, J. 1993. An ATMS-Based Belief Model for Dialogue Simulation. Technical Report HCRC/RP-45, University of Edinburgh, Human Comunication Research Centre.

Taylor, J. 1994. *A Multi-Agent Planner for Modelling Dialogue*. Ph.D. Dissertation, University of Sussex.

Teufel, S.; Carletta, J.; and Moens, M. 1999. An Annotation Scheme for Discourse-Level Argumentation in Research Articles. In *Proceedings of the EACL*.

Thomason, R. 1980. A note on syntactical treatment of modality. *Synthese* 44:391–395.

Thomason, R. 1988. Intra-agent modality and nonmonotonic epistemic logic. In Gilboa, I., ed., *Theoretical Aspects of Reasoning About Knowledge: Proceedings of the Seventh TARK Conference*, 57–69. Morgan Kaufmann.

Thomason, R. 2000. Modeling the beliefs of other agents. In Minker, J., ed., *Logic-Based Artificial Intelligence*. Dordrecht: Kluwer Academic Publishers. 375–473.

Tindale, C. 1994. Contextual relevance in argumentation. In Johnson and Blair (1994b). chapter 6, 67–81.

Toulmin, S. 1958. *The Uses of Argument*. Cambridge University Press.

Traum, D. 1994. *A Computational Theory of Grounding in Natural Language Conversation*. Ph.D. Dissertation, University of Rochester, Computer Science Department.

Turner, M. 1991. *Reading Minds: The Study of English in the Age of Cognitive Science*. Princeton: Princeton University Press.

US Department of Health and Human Services. 1991. Healthy People 2000: National Health Promotion and Disease Prevention Objectives. DHHS publication PHS 91-50212.

Uschold, M., and Gruninger, M. 1996. Ontologies: Principles, Methods and Applications. *Knowledge Engineering Review* 11(2):93–136.

Vaina, L., and Hintikka, J., eds. 1984. *Cognitive Constraints on Communication*. Dordrecht: D. Reidel.

van Eemeren, F., and Grootensdorst, R. 1984. *Speech Acts in Argumentative Discussions. A Theoretical Model for the Analysis of Discussions Directed Towards Solving Conflicts of Opinion*. Dordrecht/Berlin: Foris/Mouton de Gruyter.

van Eemeren, F., and Grootensdorst, R. 1992. *Argumentation, Communication, and Fallacies: A Pragma-Dialectical Perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.

van Eemeren, F., and Houtlosser, P. 2000. Rhetorical Analysis Within a Pragma-Dialectical Framework - The Case of R. J. Reynolds. *Argumentation* 14:293–305.

van Eemeren, F.; Grootensdorst, R.; Henkemans, F.; Blair, J.; Johnson, R.; Krabbe, E.; Plantin, C.; Walton, D.; Willard, C.; Woods, J.; and Zarefsky, D. 1996. *Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Developments*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Velásquez, J., ed. 1999. *Proceedings of the Agents'99 Workshop on Emotion Based Agent Architecture (EBAA'99)*.

von Wright, G. 1951. Deontic logic. *Mind* 60(237):1–15.

Vreeswijk, G. 1993. Defeasible Dialectics: A Controversy-Oriented Approach Towards Defeasible Argumentation. *The Journal of Logic and Computation* 3(3):3–27.

Vreeswijk, G. 1997. Abstract Argumentation Systems. *Artificial Intelligence* 90(1-2):225–279.

Walker, M., and Moore, J. 1997. Empirical Studies in Discourse. *Computational Linguistics* 23(1):1–12.

Walker, M. 1992. Redundancy in Collaborative Dialogue. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING92)*.

Walker, M. 1996. The Effect of Resource Limits and Task Complexity on Collaborative Planning in Dialogue. *Artificial Intelligence* 85(1-2):181–243.

Walton, D., and Krabbe, E. 1995. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. Albany, NY: New York University Press.

Walton, D. 1982. *Topical Relevance in Argumentation*. Cambridge: Cambridge University Press.

Walton, D. 1989. *Informal Logic: a Handbook for Critical Argumentation*. Cambridge University Press.

Walton, D. 1996. *Argument Structure: A Pragmatic Theory*. University of Toronto Press.

Walton, D. 1998. *The New Dialectic: Conversational Contexts of Argument*. Toronto Studies in Philosophy. University of Toronto Press.

Wardle, J.; Steptoe, A.; Bellisle, F.; Davou, B.; Reschke, K.; Lappalainen, R.; and Fredriskon, F. 1997. Healthy Dietary Practices among European Students. *Health Psychology* 16(5):443–450.

Warnick, B., and Kline, S. 1992. The New Rhetoric's Argument Schemes: a Rhetorical View of Practical Reasoning. *Argumentation and Advocacy* 29:1–15.

Washington, R. 1994. *Abstraction Planning in Real Time*. Ph.D. Dissertation, Stanford University, Department of Computer Science.

Weiß, G., ed. 1999. *Multi-Agent Systems*. Cambridge, MA: MIT Press.

Werch, C., and Clemente, C. D. 1994. A Multi-Component Stage Model for Matching Drug Prevention Strategies and Messages to Youth Stage of Use. *Health Education Research* 9(1):37–46.

Wieringa, R., and Meyer, J.-J. 1993. Applications of deontic logic in computer science: A concise overview. In van der Torre, L., and Tan, Y.-H., eds., *Deontic Logic in Computer Science*. Chichester, England: John Wiley & Sons. 17–40.

Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Basil Blackwell.

Woods, J., and Walton, D. 1982. *Argument: the Logic of Fallacies*. Toronto: McGraw-Hill.

Woods, J. 1994. Sunny Prospects for Relevance? In Johnson and Blair (1994b). chapter 7, 82–92.

Wooldridge, M., and Jennings, N. 1995. Intelligent Agents: Theory and Practice. *Knowledge Engineering Review* 10(2):115–152.

Wooldridge, M. 2000. *Reasoning About Rational Agents*. Cambridge, MA: MIT Press.

World Health Organization. 1986. Ottawa Charter for Health Promotion. WHO/HPR/HEP/95.1.

Ziebland, S.; Thorogood, M.; Yudkin, P.; Jones, L.; and Coulter, A. 1998. Lack of Willpower or Lack of Wherewithal? 'Internal' and 'External' Barriers to Changing Diet and Exercise in a Three Year Follow-Up of Participants in a Health Check. *Social Science and Medicine* 46(4-5):461–465.

Zukerman, I., and Litman, D. 2001. Natural language processing and user modeling: Synergies and limitations. *User Modeling and User-Adapted Interaction* 11:129–158.

Zukerman, I.; Korb, K.; and McConachy, R. 1996. Perambulations on the Way to an Architecture for a Nice Argument Generator. In Jokinen et al. (1996), 32–36.

Zukerman, I.; McConachy, R.; and Korb, K. 2000. Using Argumentation Strategies in Automated Argument Generation. In Elhadad (2000), 55–62.