

## Automatic argumentative analysis for interaction mining

Vincenzo Pallotta<sup>a\*</sup> and Rodolfo Delmonte<sup>b</sup>

<sup>a</sup>*Interanalytics, Rue des Savoises, 19, Geneva 1205, Switzerland;* <sup>b</sup>*Department of Computational Linguistics, University of Venice, Venice, Italy*

(Received 19 April 2011; final version received 16 July 2011)

Interaction mining is about discovering and extracting insightful information from digital conversations, namely those human–human information exchanges mediated by digital network technology. We present in this article a computational model of natural arguments and its implementation for the automatic argumentative analysis of digital conversations, which allows us to produce relevant information to build interaction business analytics applications overcoming the limitations of standard text mining and information retrieval technology. Applications include advanced visualisations and abstractive summaries.

**Keywords:** computational models of natural language argument; argument interpretation; online dialogue Applications of Argumentation Technology; social networks Applications of Argumentation Technology; computational linguistics Interdisciplinary links with Computational Argument; interaction Interdisciplinary links with Computational Argument

### 1. Introduction

Interaction mining is an emerging research topic focused on methods and techniques for extracting useful information from interactions between people over digital media. The forms of interaction can be very different as the means for interaction offered by digital technology are many and differentiated. A common denominator is that in order to be analysed, interactions must be recorded. For instance, logs from users' visits of websites have been extensively exploited for web analytics and user profiling. Logs of users' visits are not, however, interactions between users, but rather interactions between users and the content provided by web authors. Moreover, the user initiates the interaction and its length is very limited. We would normally not consider this a human–human interaction, rather a human–computer interaction.

Historically, the Internet was built as a collaboration tool. We might all remember USENET discussion groups, Internet Relay Chat (IRC) and above all email. Then, the web turned the Internet into a multi-media publishing tool. Collaboration came back with the advent of Web 2.0 where user-generated content became mainstream and new tools for interactions between people emerged as social networks. Additionally, the increasing speed of network connections led to moving from text-based, asynchronous and two-party interaction towards more multi-media, synchronous and multi-party interaction.

While there are several types of interactions with digital networked devices, textual (i.e. language-based) interaction is still the most widespread interaction modality. It is a modality that is also easy to record and process, for instance, with linguistic tools. Other types of interactions are emerging but, when possible, they are typically converted into text for their analysis (e.g. speech over telephone is usually transcribed for speech analytics). In the future, we expect to see recording and analysis of other interaction modalities such as gestures or facial expressions.

---

\*Corresponding author. Email: [vincenzo.pallotta@interanalytics.ch](mailto:vincenzo.pallotta@interanalytics.ch)

While the ideas we present here are general enough to be applicable to any interaction modality, we consider in this article interactions in the form of *conversations*. Conversations are interactions with a purpose. The participants in a conversation aim at reaching individual or joint goals. Conversations can show either collaborative or competitive behaviours.

Conversations have been central to the Internet since its inception. As we have already mentioned, one of the first Internet applications, IRC, was of conversational nature. Later, conversational systems have proliferated in various forms as the Internet has become more collaborative. In many situations, new modes of information sharing were based on conversation, for example, blogs, social media, remote conferencing and wikies.

Communication through computer-mediated (digital) conversations is very effective because it releases Internet users from the commitment to engage in the full process of content publishing, which was the original model of the web. Conversations are situated within an interactional context and users can avoid using a narrative style in their communications. This effectiveness also leads to the fact that individual contributions to the conversation are often impossible to be viewed as isolated from their interactional context. This causes some difficulties in retrieving relevant information from this type of data. For instance, in online reputation management,<sup>1</sup> tracking the users' attitude towards a topic in public forums may become very challenging. A standard approach that simply considers the polarity of isolated contributions could lead to misinterpretations of data in those cases where users negatively comment on other users' comments. In such a case, the attitude cannot be simply taken as negative to the main topic, but it needs to be understood in its own context, namely as a reply to a (possibly negative) comment to the main topic.

The above is only one of the possible issues that a standard approach to the analysis of Internet conversations might generate, namely when contributions are taken into account as isolated and not in their context. Another problem is the absence of certain keywords that are likely to be used for retrieving passages of conversations and which are not uttered by the participants. For instance, if one wants to retrieve passages where someone disagreed with someone else in a forum, there is a very little chance that term "disagree" will actually occur in the conversation. In these cases, it is essential to enrich the conversational content with metadata so that a retrieval system can find relevant information behind the content terms. It is an essential requirement in search technology to be able to retrieve all and only the relevant content to queries from web documents, and the same holds for conversational content.

In this article, we address the problem of accurate *Interaction Mining*, and we also provide a viable solution by introducing a novel approach based on the extraction of *argumentative structure* of conversations. This approach is based on the reasonable and validated assumption that when engaged in conversations, users follow a flexible but well-defined discourse structure (Lo Cascio 1991; Anscombe and Ducrot 1997; van Eemeren and Grootendorst 2004). This structure is very often an argumentative structure where participants contribute with specific dialogue acts having a precise argumentative force. This model might not be applicable to every interaction, especially those with no clear purpose. However, we believe that a great deal of interactions are purposeful and that users almost always pursue the goal of either seeking agreement or consensus from other participants on their opinions or trying to convey their arguments in favour or against other participant's opinions.

Considering that essential pragmatic aspects of real conversations are argumentative, we believe that extracting their underlying argumentative structure would greatly benefit interaction mining by means of understanding the purpose underlying the conversations.

The article is organised as follows. In Section 2, we explore the realm of digital conversations with a special focus on Internet conversations. In Section 3, we discuss insightful examples showing the need of new tools, which provide more appropriate ways for accessing conversational content.

In Section 4, we present our approach for the analysis of conversations based on argumentative analysis. In Section 5, we present several aspects of interaction mining and business analytics aimed at providing users with an intuitive global view of digital conversations. Finally, in Section 6, we introduce a tool for abstractive summarisation of digital conversations, which produces high-quality memos based on automatically extracted thematic and argumentative structure. In Section 7, we provide concluding remarks and propose a roadmap for future research.

## 2. Digital conversations

*Digital Conversational Systems* (DCSs) are systems that support the creation of digital records of conversations.<sup>2</sup> DCSs can be based on either stand-alone or interconnected computer systems (e.g. networked through Internet). We review different types of DCSs by outlining what are the essential retrieval requirements for each type.

We distinguish two main categories of DCSs: synchronous and asynchronous. Synchronous DCSs are those in which the exchange between participants occurs in real time and no history of the conversation (before one joins in) is maintained by the system. When the user joins an already started conversation, there is no means for retrieving what has been said before joining, and when the user leaves the conversation, what follows is no longer accessible to the user. In this category, we find Instant Messaging and VoIP systems, call-recording systems and meeting-recording systems. Asynchronous systems, in contrast, allow users to keep track of the whole conversations even when they are not directly participating in them. In this category, we find the most advanced Internet collaborative technologies such as discussion forums, blogs, microblogs, wikis and social media.

An important type of synchronous conversation, which we would consider as an important domain for interaction mining, is *contact centre conversations*. In contact centres, speech analytics has emerged as an important component for quality and workforce management. Contact centre calls are usually recorded and stored for further analysis. Speech analytics is intended to complement standard call centre analytics based on PBX<sup>3</sup> metrics (Schultz 2003, p. 116) such as average call time and number of calls handled per time unit. While these are important metrics to consider, they alone cannot provide the necessary insight to understand why things went wrong (or right) in agent–customer interactions. Currently, speech analytics solutions only partially address this issue by relying on standard indexing techniques (mostly search, even if adapted to the speech domain, e.g. phonetic or key-phrase search). As we will show in the next sections, this approach suffers from the limitations that are intrinsic when applying standard information retrieval tools to conversational data.

Asynchronous DCSs have a more consistent structure because initial topics (roots of arguments) can be easily identified. However, they might have a mixed narrative and argumentative structure, thus potentially making the analysis more difficult. For example, this is the case of a blog or a social network where there is an initial post or video and a sequence of comments attached to it. Microblogs are a special category of blogs where the size of posts is kept intentionally small and where usually the time and location contexts are highly relevant. Microblogs such as Twitter are increasingly gaining importance in our digital life at different levels of usage. Twitter posts (also known as Tweets) can be used for many purposes. For instance, a recent study conducted by UX Alliance (Sanchez, Degaki, and Oliveira 2010) has conducted a user study on how 10 of the 50 World's Most Admired Companies, as listed by Fortune Magazine in 2010, use Twitter as a platform for help desk (with, of course, different levels of quality and usability).

Another type of large-scale repository of digital conversation is that of discussion forums such as USENET newsgroups,<sup>4</sup> Disqus,<sup>5</sup> Reddit<sup>6</sup> or those attached to social media such as IMDB users discussions around movies or more technology-oriented SlashDot. Figure 1 shows a graph

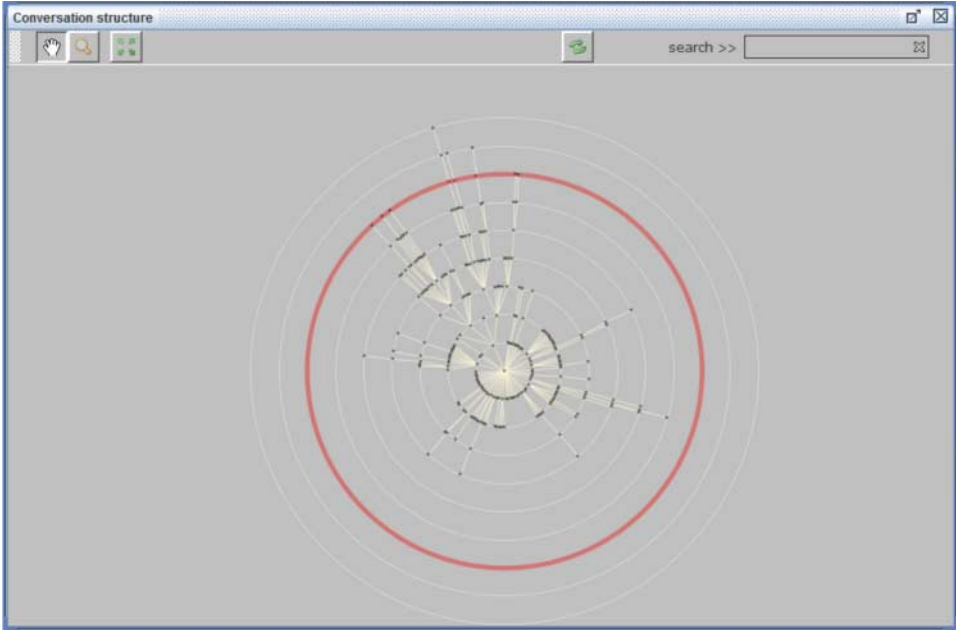


Figure 1. SlashDot discussion thread for a single selected topic.

of the thread structure of a single topic with a high number of participants and contributions of a SlashDot discussion. This picture, obtained with the WET tool from [barcellona.media.org](http://barcellona.media.org), gives us an idea of how complex and nested a web discussion might be. Getting insights from this type of data by reading all contributions could be prohibitive and time consuming. Moreover, a single contribution randomly selected from a discussion thread is hardly understandable alone. The context, that is, the previous contributions, is essential to get an understanding of the purpose of the contribution. Additionally, several studies have shown that blogs follow a specific (i.e. argumentative) conversational structure (Mabry 1997; De Moor and Efimova 2004) and that features can be exploited to detect if comments of blogs are disputative (Mishne and Glance 2006).

From the retrieval perspective, relevant information can easily be missed if the context is not adequately taken into account. In the following sections, we will show that even more advanced content-based approaches based on semantics may fail in specific situations.

Besides blogs and discussion forums, we consider another important source of conversational content, namely the oldest and most common means of Internet conversation platform: email. Email is intrinsically conversational. The average length of email bodies (attachments excluded) used in personal communications is decreasing over time.<sup>7</sup> This might be caused by the fact that email is no longer considered only as a substitute of surface mail or corporate messaging systems. Retrieval of email has improved in recent years but only towards the direction of searching for the occurrence of key terms in the messages' content. While extremely useful, we believe that this type of retrieval approach has severe limitations when dealing with conversations. For instance, one would be interested in retrieving messages where someone has been assigned a task or where a decision made was communicated. Of course, it is very unlikely that terms such "assign", "task" or "decision" will be found in the body of messages, while it is more reasonable to assume that these topics would be signalled by the presence of sentences such as "please, provide me with the report of last two weeks of activity"; similarly, decisions would have to be inferred by looking at how people have agreed on different proposals. For this last type of query, a retrieval system would

need to look at an entire conversation thread in order to figure out which proposal was accepted by the majority of people and thus considered as the decision made.

### 3. Interaction mining of digital conversations

The term interaction (or conversation, speech) mining (or analytics) has been recently used in the context of analysis of contact centre conversations (Takeuchi, Subramaniam, Nasukawa, and Roy 2009; Verint Inc. 2011). We can classify the work done in this area by the type of analysis task performed on the conversational data:

- spotting of events from conversations;
- classification of call types and
- clustering of similar calls.

While these data are undoubtedly important, we believe that better insights can be obtained from contact centre data. The insights we are looking for are those that are, for instance, brought to the attention in the work of Rafaeli, Ziklik and Doucet (2007), where it has been shown that there is a statistically significant correlation between agent's customer-oriented behaviours (COBs) and customer's satisfaction. The selected COBs indicators are as follows: (a) anticipating customer requests, (b) offering explanations/justifications, (c) educating customers, (d) providing emotional support and (e) offering personalised information. We believe that key-phrase spotting and phonetic search could only minimally help in automatically detecting such COBs. Some of these COBs are easily mapped into categories, which, as will be shown, can be detected with our interaction mining solution (e.g. (a) and (b)).<sup>8</sup>

#### 3.1. *Limitations of content-based interaction mining*

We argue that interaction mining performed with systems based on content analysis alone might have severe limitations. We will show in which circumstances such systems can fail in retrieving relevant information from conversations.

First of all, searches can be done within a single conversation and across several (related) conversations. This micro–macro dichotomy is essential in selecting the appropriate indexing and search algorithm. In the first case, one is interested in looking for (and understanding) the relationships between individual participants' contributions (e.g. if a conflict between two people has been resolved), while in the second case, one might be more interested in discovering recurring patterns over several conversations (e.g. the level of sentiment with respect to a given theme).

Second, it is important to distinguish different types of information retrieval from conversational data, the basic one being that of thematic content retrieval. If we restrict ourselves in looking for term occurrences, then we might consider as a baseline a system based on the classical Term Frequency–Inverse Document Frequency (TF–IDF) indexing model and queries made of a bag-of-words (Salton and Buckley 1988). The performance of this method on conversational data is even difficult to quantitatively assess because it would depend on the type of queries asked. Reasonably, we might just assume that if terms from the query are contained in a turn, the turn will be eventually retrieved. With this assumption, we can make our assessment on a qualitative basis, trying to understand the precise circumstances where this approach does not work.

Typically, problems arise if the searched terms are made of related terms that do not occur in the conversations. If the related terms are variants of the occurring terms, then query expansion and reformulation techniques (Navigli and Velardi 2003) might improve the results, but only to

a certain extent. If the query terms happen to describe aspects of the conversation that cannot be directly translated into content words, then the problem becomes more serious. Moreover, if one is looking at the emerging relationships between participants in the conversation such as cooperativeness or social bonds, then, apparently, the standard Information Retrieval (IR) approach is no longer applicable.

In order to help the reader to intuitively grasp the shortcomings of the standard IR approach, we try to answer the following question from a small (made up) conversation, which we report below. The question we would like to answer is as follows:

What were the arguments brought by Mary against the new computer John proposed to buy?

The conversation we consider is the following, and it is made of three turns:

1. John: *We need to buy a new computer, maybe an Apple.*
2. Mary: *Why an Apple? I think IBM computers are more robust.*
3. John: *Well, Apple computers are robust too and cheaper than IBM.*

If we select the (underlined) content words and entities from the query

*What were the arguments brought by Mary against the new computer John proposed to buy?*

and try to match them against the above conversation, we might end up in finding the first turn, which does not contain any clue to the correct answer. If we instead transform the conversation into a narrative, we may obtain the following description:

There was a discussion between John and Mary. John raised the issue of buying a new computer and proposed to buy an Apple. Mary disagreed by noticing that IBM computers are more robust. John replied by saying that Apple computers were cheaper and robust too.

In that case, if we try the query again, we are still not able to find the answer. However, with some reformulation, we might infer that “bring argument against” would translate into “disagree” and we will possibly find the right sentence. Without such a reformulation, a standard IR approach fails to find the correct answer, namely that “IBM computers are more robust”.

A conversation can be queried only if reformulated as a narrative that describes what happened during the conversation. This narrative must contain those words that are likely to occur as query terms. Hence, question answering on conversations can be reframed as an entailment problem where the user makes a statement and wants to check its truth in the conversation (Harabagiu and Hickl 2006).

An experiment of using standard IR for interaction mining has been carried out in the context of “browser evaluation test” (BET) campaign of the Augmented Multimodal Interaction (AMI) project (Wellner, Flynn, Tucker, and Whittaker 2005). In this experiment, a set of “interesting” statements about recorded meetings had been previously created and used as a set of test queries. To be validated – that is, check if they are true or not in the conversation – a portion of the conversation must be retrieved. The task can be performed by humans through a meeting browser and by a computer algorithm. Humans score an average of 70–80% with several browsers and different conditions in retrieving the passage and validating the statement, as reported by Popescu-Belis, Baudrion, Flynn and Wellner, (2008). An algorithm based on standard IR presented by Anh Le and Popescu (2009) shows a performance far below that of humans and scores only 60% accuracy, although immensely faster. The authors believe that this first attempt can be largely improved with smarter IR techniques. However, we argue that this score might only be improved to a small extent as it has been shown in Pallotta, Seretan and Ailomaa (2007) that for a large portion (52%) the BET test queries cannot be answered by standard IR methods. This estimate was based on the judgement of selected experts who considered what type of information was required in

order to answer the BET queries. Notably, the queries that were judged as the difficult ones were exactly those that required substantial knowledge about the dynamic of the conversation and its outcomes.

### 3.2. *Argumentative models of conversations*

Conversations are a pervasive phenomenon in our digital society. We need to consider appropriate models for mining conversational data from different perspectives beyond classical thematic (content-based) indexing approaches. We advocate that recognising the *argumentative structure* of digital conversations can help in improving the effectiveness of standard IR techniques in simple cases and even overcome their limitations in complex cases.

Although argumentative modelling of conversations has been addressed in several works, unfortunately very few provided a computational approach. Rather than focusing on argument recognition, those who provide a computational modelling of arguments of conversations focus more on argument generation (Dessalles 2007), support computer-mediated collaboration systems (Gordon and Karakapillidis 1997; Buckingham Shum et al. 2006) or provide better capabilities to dialogue systems (Larsson 2002) and dialogue systems that engage in argumentation (Chu-Carroll and Carberry 1996; Grasso, Cawsey and Jones 2000; Mazzotta, de Rosis and Carofiglio 2007).

For the purpose of recognition of argumentative structure from conversations, there are basically two types of approaches. The first is based on rhetorical relations between two or more dialogue units (e.g. speech acts or turns), whereas the second assumes that dialogue units can have their own argumentative force as a function of its content. In the first model, which we call “relational”, argumentation always occurs between two units, while in the second model, “flat”, argumentative force becomes a feature of a unit itself. The first approach leads to richer structures, which can also be recursive (i.e. develop a tree-based structure), but which can be very hard to compute. The second simpler approach is more suitable for the classification of argumentative actions based on local features of the single unit. Linguistic analysis can then be used to extract the relevant features needed by the classifier. However, the flat model can be extended with relations to reach the same expressivity of the relational model. For instance, one can introduce standard rhetorical relations such as “elaboration” in order to link to related units with different argumentative forces (e.g. a claim and a justification).

Based on the relational model, Stent and Allen (2000) proposed an extension of Rhetorical Structure Theory (RST) (Mann and Thompson, 1987) to deal with argumentative dialogue. They introduced argumentative relations as adjacency pairs (e.g. propose–accept and assert–reject) in addition to standard discourse relations. They provided an annotation schema and a tool that was used to annotate the Monroe corpus. In a similar vein, Rienks, Heylen and van der Weijden (2005) have proposed an argumentation schema based on relations. Unlike in Stent and Allen (2000), their units are not just text chunks but defined dialogue units, that is, dialogue acts from the Augmented Multiparty Interaction (AMI) tagset (Verbree, Rienks and Heylen 2006). Relations are based on a custom set inspired by the Issue Based Information Systems (IBIS) model (Kunz and Rittel 1970).

Example of the flat model that we mention here is the Larsson Issue-Under-Negotiation (IUN) model (Larsson 2002). This model, based on the Information State Update Dialogue Model, provides a new argumentative perspective to the Ginzburg notion of *Question-Under-Discussion* (Ginzburg 1996). It applies to the class of negotiation dialogues where argumentative threads are seen as pertaining to particular issues, modelled as questions on the IUN stack. The IUN model has new moves such as the following: *Introduce* that introduces new issues; *Proposal* that introduces possible alternative answers thereto; and *Acceptances* or *Rejections* that remove those alternatives. This model is very similar in spirit to the IBIS model and is also grounded in a fully

fledged dialogue theory. This model has been adopted in the CALO Meeting Assistant System (Tür et al. 2010), which has as a goal the real-time reconstruction of an argumentative structure by overhearing discussions in design meetings.

Other flat models, which are simpler, are the “agreement–disagreement” model (Hillard, Ostendorf and Shriberg 2003; Galley, McKeown, Hirschberg and Shriberg 2004), the “decision points” model (Hsueh and Moore 2007, 2009; Fernández, Frampton, Ehlen, Purver and Peters 2008) and the “action items” model (Niekrasz, Purver, Dowding and Peters 2005; Purver et al. 2007).

### 3.3. *Our approach*

In our approach, we adopt an enhanced flat argumentative model that is part of a more comprehensive framework described in Pallotta (2006). The argumentative structure that we adopt defines the different patterns of argumentation used by participants in the dialogue, as well as their organisation and synchronisation in the discussion.

A conversation (or discussion) is decomposed into several *argumentative episodes* such as *issues*, *proposals*, *elaborations* and *positions*, with each episode being possibly related to specific aggregations of elementary dialogue acts. We adopted an argumentative coding scheme, the *Meeting Description Schema* (MDS), introduced in Pallotta, Ghorbel, Ballim, Lisowska and Marchand-Maillet (2004) and strongly influenced by the IBIS model where the argumentative structure of a discussion is composed of a set of topic discussion episodes (a discussion about a specific topic). In each topic discussion, there exists a set of issue episodes. An issue is generally a local problem in a larger topic to be discussed and solved. Participants propose alternatives, solutions, opinions, ideas, etc. in order to achieve a satisfactory decision. Meanwhile, participants express their positions and standpoints either through acts of accepting or rejecting proposals or by asking questions related to the current proposals. Hence, for each issue, there is a corresponding set of proposal episodes (solutions, alternatives, ideas, etc.) that are linked to a certain number of related position episodes (e.g. a rejection to a proposed alternative in a discussed issue) or questions and answers.

The limits of sequential analysis of conversation (Schegloff and Sacks 1973) have been already pointed out by Goffman (1981), who proposed to extend the notion of adjacency pair with that of chains of *interaction rounds*. As for other related work, we also see similarities of our approach with the *argumentation dependency grammar* proposed by Lo Cascio (1991), although in his work only argumentative structure of monologues was considered.

Practically, when analysing the dialogue, adjacency pairs are not enough to represent the hierarchical structure of the discussion: consider, for instance, an answer that replies to two different questions in the discussion. In this case, we need to add a relation that links the answer to both the questions. We call this relation “replies\_to”. The “replies\_to” links a (re)action to one or more previous (possibly in time) actions and induces an argumentative chain structure on the dialogue, which is local to each action and which enables the visualisation of its context. For instance, the context of the action of “accepting a clarification” will be a chain of linked actions, namely the action of the clarification, that of the proposal that is clarified and the action of raising an issue for which the proposal was made. Argumentative actions can overlap in time, as, for instance, in those cases where the acceptance of a justification is uttered in the form of “backchannel” during the presentation of the justification.

Argumentative actions such as REQUEST, ACCEPT and REJECT might correspond to basic dialogue acts (Clark and Popescu-Belis 2004). In this case, we have refined the concept of dialogue act and adjacency pairs by specifying the role of dialogue acts in constructing the argumentative structure of the discussion through the “replies\_to” relation. We also realised that there is an



invariant structure of discussion actions, which can be obtained by a more general schema by simply varying an internal parameter specifying the outcome of the action (in our case, some of them are IBIS categories such as “issue” and “alternative”). The general argumentative structure of a discussion of MDS is provided by the following set of rules<sup>9</sup>:

RAISE(issue) ←—  
     PROPOSE(solution/idea/alternative/opinion).

PROPOSE(solution/idea/alternative/opinion) ←—  
     REQUEST(explanation/justification) |  
     ACCEPT(solution/idea/alternative/opinion) |  
     REJECT(solution/idea/alternative/opinion).

REQUEST(explanation/justification) ←—  
     PROVIDE(explanation/justification).

PROVIDE(explanation/justification) ←—  
     ACCEPT(explanation/justification) |  
     REJECT(explanation/justification) |  
     CHALLENGE(explanation/justification).

CHALLENGE(explanation/justification) ←—  
     RAISE(issue).

ACCEPT(solution/idea/alternative/opinion) ←—  
     PROVIDE(explanation/justification).

REJECT(solution/idea/alternative/opinion) ←—  
     PROVIDE(explanation/justification).

The MDS argumentative structure reflects the Kunz and Rittel’s IBIS model in terms of its process structure. The only structural constraints are the restrictions imposed on the backward-looking relation “replies\_to”, which are graphically represented above by the backward arrow “←—”. The recognition of an argumentative action of type B is conditioned by the presence of previously occurring action of the type A as required in the rule  $A \leftarrow B$ . For instance, the presence of a CHALLENGE action “presupposes” the existence of a previously occurring PROVIDE action to which it replies. This model can also be viewed (by reversing the arrows) as a *structure of expectations* in discussions. For instance, if a RAISE(issue) occurs, we might expect that a PROPOSE(idea) will follow afterward, even if not immediately. This structure has been used as a heuristic for automatically building the thread structure in the A3 algorithm described in Section 2.3.

As remarked in Dascal (1992), “conversations cannot be described in terms of conditions of well-formedness”. Conversations can be modelled by structures of expectations or in terms of “conversational demand” (Dascal 1977). Such expectations do not impose rigid constraints on the structure of the conversation, since they are only presumptive and thus defeasible. Defeating

expectations of conversational demand is comparable to flouting Gricean conversational maxims (Grice 1975), entailing, instead of ill-formedness, a non-standard meaning. As an example of this phenomenon, consider the following dialogue excerpt<sup>10</sup> where the symbol “=>” means “addressing”.

B : *We have <this> problem.*

A(=> B) : *I think we should do <this> and <this>*

C : *I don't understand <this>. Don't you think we should talk about it?*

<long pause>

B(=> A) : [pointing to A] *Well, I think you're right!*

In this example, C is raising an issue, which is ignored by all participants, for example, no proposals are made. After a pause, the discussion is resumed by B who accepts A's initial proposal. The expectation of having proposals made immediately after the raising of an issue is defeated here; therefore, an implicit meaning (action) can be inferred: rejecting the discussion of an issue considered as irrelevant by all the participants.

### 3.4. Effectiveness of argumentative models for interaction mining

To better understand the impact of argumentative analysis in interaction mining, in this section, we will provide a real example of how argumentative indexing can solve outstanding problems in indexing and retrieval of conversational content. We illustrate the power of the MDS argumentative model presented in Section 3.3 by contrasting the limitation of classical term-based indexing for retrieving relevant content of a conversation. Consider the following conversation excerpt from the International Computer Science Institute (ICSI) meetings corpus (Janin et al. 2003):

1702.95 David: *so – so my question is should we go ahead and get na- - nine identical head mounted crown mikes?*

1708.89 John: *not before having one come here and have some people try it out.*

1714.09 John: *because there's no point in doing that if it's not going to be any better.*

1712.69 David: *okay.*

1716.85 John: *so why don't we get one of these with the crown with a different headset?*

For a query such as “*Why was the proposal on microphones rejected?*”, a classical indexing schema would retrieve the first turn from David by matching the relevant query term “microphone”. There is no presence of other query terms such as “reject” or “proposal”. Moreover, it is not possible to map the Why-question onto some query term (e.g. reason, motivation, justification and explanation). This makes it impossible to adequately answer the query without any additional metadata that highlight the argumentative role of the participants' contributions in the conversation.

In Figure 2, we show the argumentative structure of the conversation excerpt in MDS that allows us to correctly answer the query by selecting the third turn. In fact, the Why-question is mapped to a query term that is found as an argumentative index, “justification”, for that turn. Of course, finding justification is not enough, and the retrieval algorithm needs to check whether that justification has been provided as a *rejection* of a *proposal* (or alternative) made to an issue on the topic of microphones. This can be done by navigating back through the argumentative chain up to the “issue” episode whose content thematically matches the “microphone” query term.

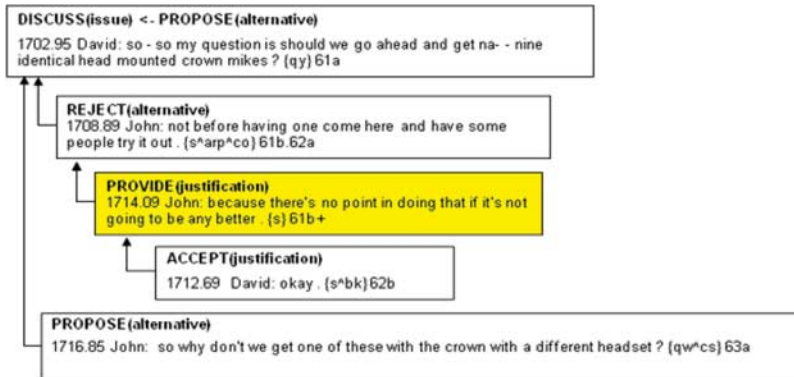


Figure 2. Argumentative structure of a conversation in MDS.

#### 4. A3: automatic argumentative analysis

The interaction mining system that we introduce here is part of a larger system built over the past two decades, whose goal is to perform broad-coverage, domain-general Natural Language Understanding (see Section 4.1). The underlying grammar, lexicon, the semantics and all intermediate modules are intended to be domain general and to be easily portable to different application domains. As is the case with all rule-based systems, we need not collect and annotate corpora for specific subtasks because the system already has good performance in all current parsing- and semantic-related tasks (see Delmonte et al. 2006; Delmonte 2007, 2008).

Conversational data challenge, however, the performance of the system. Our experience with ICSI dialogues (Janin et al. 2003) showed that the semantic representation and the output of the parser were not fully adequate on these data. The system has been improved by solving deficiencies detected in an empirical manner.

This approach made us aware of the peculiarities of spoken dialogue texts such as the ones made available in the ICSI project (Janin et al. 2003) and of the way of implementing solutions in such a complex system. The distribution of turns according to their length expressed in words is as follows:

- per cent of turns made of one single word: 30%,
- per cent of turns made of up to three words: 40%,
- number of words/turn overall: 7 and
- number of words/turn after subtracting short utterances: 11.

These values correspond to those found for PennTreebank corpus where we can count up to 94K sentences for 1 million words – again 11 words per sentence. In analysing ICSI dialogues, we found turns with as much as 54 words depending on the topic under discussion and on the people on the floor.

Computing semantic representations for conversations is a particularly hard task which – when compared with written text processing – requires the following additional information to be made available:

- adequate treatment of fragments;
- adequate treatment of short turns, in particular, one-/two-word turns;
- adequate treatment of first-person singular and plural pronominal expressions;
- adequate treatment of disfluencies, thus including cases of turns made up of just such expressions, or cases when they are found inside the utterance;

- adequate treatment of overlaps and
- adequate treatment of speaker identity for pronominal coreference.

In addition, in our system, every dialogue turn receives one polarity label, indicating negativity or positivity, and this is computed by looking into a dictionary of polarity items. This is subsequently used to decide on argumentative automatic classification.

#### 4.1. *The GETARUNS system*

GETARUNS, the system for text understanding developed at the University of Venice (Delmonte 2007; 2009), is organised as a pipeline that includes two versions of the system: what we call the Partial and the Deep GETARUNS. We will first present the Deep version, which is equipped with three main modules: a lower module for parsing, where sentence strategies are implemented; a middle module for semantic interpretation and discourse model construction which is cast into Situation Semantics; and a higher module where reasoning takes place.

The system is based on the Lexical Functional Grammar (LFG) theoretical framework (Bresnan 2000) and has a highly interconnected modular structure. It is divided up into a pipeline of sequential but independent modules which realise the subdivision of a parsing scheme as proposed in LFG theory where a c-structure is built before the f-structure can be projected by unification into a DAG (Direct Acyclic Graph). In this sense, we try to apply in a given sequence, phrase-structure rules as they are ordered in the grammar: whenever a syntactic constituent is successfully built, it is checked for semantic consistency. In case the governing predicate expects obligatory arguments to be lexically realised, they will be searched and checked for uniqueness and coherence, as the LFG grammaticality principles require.

Syntactic and semantic information is accessed and used as soon as possible: in particular, both categorial information and subcategorisation information attached to predicates in the lexicon are extracted as soon as the main predicate is processed, be it adjective, noun or verb, and are used to subsequently restrict the number of possible structures to be built. Adjuncts are computed by semantic compatibility tests on the basis of selectional restrictions of main predicates and adjuncts heads.

The output of grammatical modules is then fed to the anaphoric resolution module, which activates an algorithm for anaphoric binding. Antecedents for pronouns are ranked according to grammatical function, semantic role, inherent features and their position at f-structure. Eventually, this information is added into the original f-structure graph and then passed on to the Discourse Module (hence DM).

GETARUNS has a highly sophisticated linguistically based semantic module which is used to build up the DM. Semantic processing is strongly modularised and distributed among a number of different modules which take care of spatio-temporal reasoning, discourse-level anaphora resolution and other subsidiary processes such as Topic Hierarchy which cooperate to find the most probable antecedent of co-referring and co-specifying referential expressions when creating semantic individuals. These are then asserted in the DM, which is then the sole knowledge representation used to solve nominal co-reference, before proceeding to access external knowledge in ontologies.

The system uses two anaphora resolution modules, which work in sequence: they constitute independent modules and allow no backtracking. The first one is fired whenever a free sentence external pronoun is spotted; the second one takes the results of the first module and checks for nominal anaphora. They both have access to all data structures and pass the resolved pair, anaphor-antecedent, to the following modules.

Semantic mapping is performed in two steps: at first, a logical form is produced, which is a structural mapping from DAGs onto unscoped well-formed formulas. These are then turned into

situational semantics informational units, infons, which may become facts or “sits”. Each unit has a relation, a list of arguments which, in our case, receive their semantic roles from lower processing – a polarity, a temporal and a spatial location index. The clause-level interpretation procedure interprets clauses on the basis of lexical properties of the governing verb. This is often not available in short turns and in fragments. So in many cases, fragments are built into a sentence by inserting a dummy verb which varies from dummy BE or dummy SAY depending on speech act present.

## 4.2. Conversation analysis with GETARUNS

We will proceed by addressing each aspect of conversations in the order with which it is dealt by the system.

### 4.2.1. Overlaps

Overlaps are an important component of all spoken dialogue analysis (Delmonte 2003). In all dialogue transcriptions, overlaps are treated as a separate turn from the one in which they occur, which usually follows it. This is clearly problematic from a computational point of view. Therefore, when computing overlaps, we set as our first goal that of recovering the temporal order in which speakers interact.<sup>11</sup> This is done because

- overlaps may introduce linguistic elements which influence the local context and
- eventually, they may determine the interpretation of the current utterance.

For these reasons, they cannot be moved to a separate turn because they must be semantically interpreted where they temporally belong. In addition, overlaps are very frequent. The algorithm that we built looks at time stamps of turns, and every time the following turn begins at a time preceding the ending time of the current turn, it enters a special recursive procedure. It looks for internal interruption in the current turn and splits the utterance where the interruption occurs. Then, it parses the split initial portion of the current utterance and continues with the overlapping turn. This may be reiterated in case another overlap follows, which again begins before the end of the current utterance. Eventually, the procedure returns to the analysis of the current turn with the remaining portion of the current utterance.

In Table 1, we present data related to overlaps identified in 10 dialogues we processed.

We classified overlaps into two types – WHILE and AFTER – according to whether they take place inside the turn of the current speaker or at the end. The second case is regarded as normal and non-disrupting of the current speaker’s conversational plan. Thus, on a total number of 13,158 turns, we identified 3085 overlaps divided nearly equally in these two classes.

Table 1. Overlaps and their effects on speaker’s conversational plan.

	Total	Continue	Interrupt	inter_cont	inter_change	inter_other
Turns	13,158	–	–	–	–	–
While	1624	1369	46	87	22	63
After	1461	–	–	–	–	–

We proceeded by further subdividing WHILE overlaps into five subclasses, briefly described below:

- *Continue*: the current speaker continues talking;
- *Interrupt*: the current speaker is interrupted and there is no continuation;

- *Inter\_Cont*: the current speaker is interrupted but then he/she continues his/her plan in a following turn;
- *Inter\_Change*: the current speaker is interrupted and changes his/her plan by either changing the subject topic or answering the overlapping speaker and
- *Inter\_Other*: other speakers interrupt the dialogue as well.

As can be observed, the case constituted by *Inter\_Change*, which is the most interesting one from a semantic and pragmatic point of view, is in fact the less frequent one. We assume, however, that this may be determined by other factors attaining to the type of conversation being entertained by the participants, as well as by the nature of the topics discussed, and eventually by the personalities of the interlocutors.

#### 4.2.2. *The treatment of fragments and short turns*

Fragments and short turns are filtered by a lexical lookup procedure that searches for specific linguistic elements, which are part of a list of backchannels, acknowledgement expressions and other similar speech acts. In case this procedure has success, then no further computation takes place. However, this only applies to utterances shorter than five words and should be made up only of such special words. No other linguistic elements should be present apart from the non-words, that is, words that are only partially produced and have been transcribed with a dash at the end. In such a situation, two failure procedures are activated:

- Graceful failure procedures for ungrammatical sentences, which might be full-fledged utterances but semantically uninterpretable due to the presence of repetitions, false starts and similar disfluency phenomena. In addition, they may be just fragments, that is, partial or incomplete utterances, hence non-interpretable. They are ruled out by imposing grammatical constraints of well-formedness in the parser.
- Failure procedures for utterances that are constituted just by disfluency items and no linguistically interpretable words. These must be treated as semantically empty utterances and are recognisable by the presence of orthographic signs indicating that the words have not been completed and are just incomprehensible; they are ruled out by inspecting the input in search of special orthographic marks and preventing the utterance to be passed down to the partial/deep parser.

A different processing applies to elliptical utterances, which contribute one specific speech act or communicative act. These utterances may express agreement/disagreement, acknowledgements, assessments, continuers, etc. All these items are processed as being complements of the abstract verb SAY which is introduced in the analysis and has as a subject the name of the current speaker.

### 4.3. *The A3 algorithm*

*Automatic Argumentative Analysis (A3)* is carried out by a special module activated after processing each dialogue. This module takes as input the complete semantic representation produced by the system recorded in Prolog facts in the Discourse Model. The elements of the semantic representation that we use are the following:

- all facts in Situation Semantics contained in the Discourse Model, which include individuals, sets, classes, cardinality and properties related to entities by means of their semantic indices;
  - facts related to spatio-temporal locations of events with logical operators and semantic indices;
  - vectors of informational structure containing semantic information at propositional level, computed for each clause;

- vectors of discourse structure with discourse relations computed for each clause from informational structure and previous discourse state<sup>12</sup>;
- dialogue act labels (if available) associated with each utterance or turn following the ICSI classification (Shriberg, Dhillon, Bhagat, Ang and Carvey 2004);
- overlap information computed at utterance level;
- topic labels associated with semantic indices of each entity marked as the topic of discourse and
- all utterances with their indices as they have been automatically split by the system.

To produce argumentative annotation, the system uses the following 20 discourse relation labels: *statement, narration, adverse, result, cause, motivation, explanation, question, hypothesis, elaboration, permission, inception, circumstance, obligation, evaluation, agreement, contrast, evidence, setting and prohibition.*

Discourse relation labels come partly from RST (Mann and Thompson 1987) and partly from other theories, as, for instance, those reported by Hobbs (1993) and Dahlgren (1988). More details are available in Chapters on Discourse Relations in Delmonte (2007). Discourse relations are automatically extracted by GETARUN. An evaluation of its performance is provided in Delmonte, Harabagiu and Nicolae (2007a).

These relations are then mapped onto five MDS argumentative labels:

ACCEPT, REJECT/DISAGREE, PROPOSE/SUGGEST,  
EXPLAIN/JUSTIFY and REQUEST.

In addition, we use the label DISFLUENCY for all those turns that contain fragments, which are non-sentences and are semantically uninterpretable.

The algorithm proceeds in the following manner:

- (1) It recovers (when available) Dialogue Acts for each dialogue turn as they have been assigned by the system. These labels coincide with the ICSI labels FGB (Floor Grabber), FHD (Floor Holder), BKC (Backchannel), ACK (Acknowledge) and RHQ (Rhetoric-Question), with the addition of NEG (Negation), AST (Assent), MTV (Motivation), PRP (Proposal), GRT (Greeting) and CNL (Conclusion).
- (2) It recovers Overlaps as they have been marked during the analysis.
- (3) It produces an Opinion label, which we call Polarity and which can take one of two values, Positive or Negative, according to whether the sentence contains positive or negative linguistic descriptions.
- (4) It produces a list of Hot Spots and builds up Episodes. Hot Spots are a set of turns in sequence where the interlocutors overlap each other frequently. Episodes are a set of turns in which a single speaker presents his/her arguments on a topic and which may occasionally be interrupted by overlaps or by short continuers, backchannel or other similar phenomena but without other speakers however grabbing the floor.
- (5) Finally, the main predicate that assigns argumentative labels is called:
  - (a) At first, it tries exceptions on the basis of the actual words contained in the turn. These exceptions may be constituted by Greetings, specific Speech Acts, Conventional utterances pronounced in specific situations like Thanking, etc.
  - (b) Then, short utterances are checked. In case, they end up with a question mark they are labelled as *question*. Otherwise, the Dialogue Act label is considered. Negations are also computed here.
  - (c) Now, the main call is activated. In order to start matching the rules, the semantic information is recovered for the current turn, clause by clause.

- (d) When semantic information has been recovered, the rules are triggered.

There are about 33 Prolog rules, which take as input the following vector of features:

**assignargument(NoCI, [Pol,DialAct], DiscDom, DiscRel, Relev, DomPointView, Output),**

where Output is the output argumentative label chosen by the rule; Pol is the polarity detected by the sentiment analysis module; DiscRel is one of the discourse relations mentioned above; DiscDom may be *factive* or *non-factive*, *suggestion* or *proposal*; Relev may be *foreground* or *background*; DomPointView may be *objective* or *subjective*.

Rules are applied by matching input labels by means of Finite State Automata. However, sometimes conditions and constraints are applied, for instance, to check whether the current speaker holds the floor in the two preceding or following clauses.

The rules produce a set of argumentative labels, one for each clause. The system then chooses the label to associate with the turn utterance from a hierarchy of argumentative labels graded for *pragmatic relevance*, which establishes that, for instance, QUESTION is more relevant than NEGATION, which is more relevant than RAISE\_ISSUE, etc.

An example of a rule (taken from a larger rule set) for recognising the PROPOSE category is the following Prolog rule:

**assignargument([suspension,Dialact], SimplDiscDom, evaluation,  
foreground, subjective, PoV, propose):-  
  \+ on(bkc,Dialact), !.**

The rule looks at feature vectors (i.e. the first six parameters) and returns the PROPOSE argumentative category as the last argument. The rule's body checks if it is not a backchannel.

The A3 algorithm is able to build a structure like that shown in Figure 2 because it also computes the "replies\_to" back-link between turns. A heuristic algorithm based on the locality of the adjacency pair computes the back-link. For instance, if a turn is labelled as SUGGEST, the algorithm will look for the most recent turn labelled as RAISE\_ISSUE.

In the training stage, the system has been used to process the first 10 dialogues of the ICSI corpus (Janin et al. 2003) containing a total number of 98,523 words and 13,803 turns. In this stage, we tuned all the modules and procedures. We needed to improve, in particular, the module for argumentative automatic classification in order to cover all conventional ways to express agreement. In the test stage, we randomly chose two different dialogues to assess the performance of the A3 algorithm.

#### 4.4. Experimental results

The classification results obtained with the algorithm described above have been compared against human annotations. A skilled linguist provided a turn-level annotation using argumentative labels for the same dialogues used for testing our algorithm. The comparison of the two annotations is a complex task, which still has to be completed. Although we cannot report agreement statistics yet, we expect the annotations to be in line with similar experiments on the same subject (Pallotta et al. 2007).

In Table 2, we report data related to the experiment of automatic annotation of argumentative categories. On a total of 2304 turns, 2247 have received an argumentative automatic classification, with a recall of 97.53%. As can be seen from the table, the A3 algorithm achieves an F-score that is significantly higher than the previous results reported in the literature on the same topic, which are all below 80% (Rienks and Verbree 2006; Hakkani-Tür 2009).



Table 2. Performance results for the argumentative analysis algorithm.

	Correct	Incorrect	Total found
Accept	662	16	678
Reject	64	18	82
Propose	321	74	395
Request	180	1	181
Explain	580	312	892
Disfluency	19		19
Total	1826	421	2247

The overall precision is computed as the ratio between the correct argumentative labels and the found argumentative labels and corresponds to 81.26%. The F-score represents the harmonic mean of precision and recall and is 88.65%.

### 5. Interaction business analytics

Interaction business analytics (IBA) is a new paradigm for analysis and visualisation of interactions over the Internet, introduced in Pallotta, Vrieling and Delmonte (2011). It focuses on aggregating and visualising the argumentative features that constitute the results of an argumentative analysis for several analytical purposes. In this section, we address the problem of argumentative visualisation by reviewing previous work and presenting the representations that we propose.

#### 5.1. Conversation graphs

Currently, there are no standard metadata schemas for structuring and indexing conversations and for visualising their structure (Pallotta et al. 2004; Verbree 2006). Previous work on visualising argumentation has mainly been driven by the need for tools to improve the quality of discussions in real-time meetings (Fujita, Nishimoto, Sumi, Kunifuji and Mase 1998; Bachler, Shum, Roure, Michaelides and Page 2003; Rienks et al. 2005; Buckingham Shum et al. 2006; Michaelides et al. 2006). Some research has also addressed the use of such visualisations for browsing past conversations, and end-user evaluations have been positive with respect to impact in retrieval effectiveness (Rienks and Verbree 2006; Ailomaa and Rajman 2009).

As a means to represent the argumentative structure of meeting recordings, *conversation graphs* were introduced in Ailomaa (2009) to help users search digital conversations in conjunction with thematic indexing (see Figure 3). Conversation graphs are diagrams that summarise what topics were discussed, how long they were discussed, which participants were involved in the discussion and what type of arguments they contributed. Conversation graphs are built directly by looking at

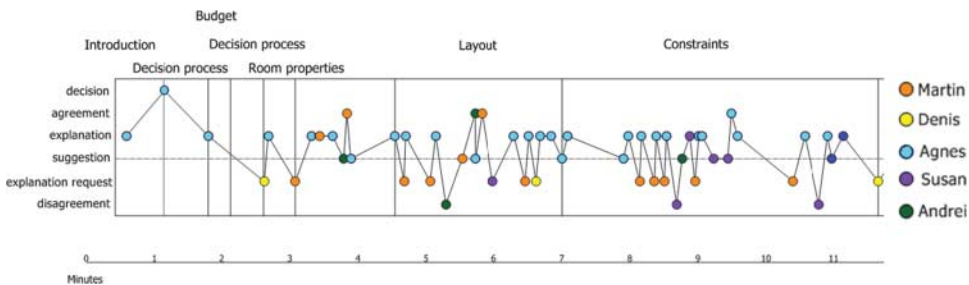


Figure 3. Conversation graph from Ailomaa (2009).

Table 3. Mapping table for argumentative categories to levels of cooperativeness.

Argumentative categories	Level of cooperativeness
Accept explanation	5
Suggest	4
Propose	3
Provide opinion	2
Provide explanation or justification	1
Question	-1
Raise issue	-2
Request explanation or justification	-3
Provide negative opinion	-4
Disagree	-5
Reject explanation or justification	-5

the MDS labels assigned to conversation's turns, either manually or automatically through the A3 algorithm.

An important criterion in the design of conversation graphs is that the visualisation of the argumentative structure has to be intuitive so that users do not need to spend effort on learning the argumentative categories before being able to use them for searching in digital conversations. For this purpose, the graph representation introduces the notion of "positive" and "negative" contributions in discussions. Peaks along the time axis represent positive contributions, such as agreements and decisions, whereas valleys represent negative contributions, such as disagreements. Suggestions are neutral in polarity and are positioned in the middle.

Instead of directly showing argumentative categories in the conversation graph, it is possible to map them into a "cooperation" scale as shown in Table 3. This makes a conversation graph more intuitive and easy to understand as shown in Figure 9.

This way of displaying the argumentative content should make it intuitive to search for the relevant meeting episodes by specifying argumentative search criteria rather than simple content-based criteria. First results of user studies have shown that conversation graphs are promising tools for both querying and browsing indexed digital conversations (Ailomaa 2009; Ailomaa and Rajman 2009).

## 5.2. Aggregated view

If further aggregation is made, conversation graphs can be turned into tools to study the collaborative behaviour of people in groups. Figure 4 shows the distribution of cooperative behaviour of each participant to a conversation for the most important 10 topics discussed through a four-dimensional representation.<sup>13</sup> For each topic in the  $x$ -axis and for each participant in the  $y$ -axis, a coloured histogram is displayed, which represents the level of cooperativeness from the scale in Table 3 as the colour and the degree of involvement in the topic (i.e. the number of turns) as the height. In this particular example, it is easy to recognise that Andreas and Dave display a strong challenging behaviour for topics, respectively, "format" and "word".

As an additional example of the power of argumentative analysis and visualisation, we have analysed an interview made to Google Wave team engineers and published on the web.<sup>14</sup> We analysed the interview transcription with the A3 system and provided several aggregated views of the extracted information, as shown in Figure 5. As expected, the most common argumentative action for the Interviewer is REQUEST EXPLANATION.

For sentiment and subjectivity, Figure 6 shows that the most frequent category is QUESTION.

Another type of argumentative visualisation that focuses on the most prominent contributors is illustrated in Figure 7 showing the reply (or social) structure of the conversation where the

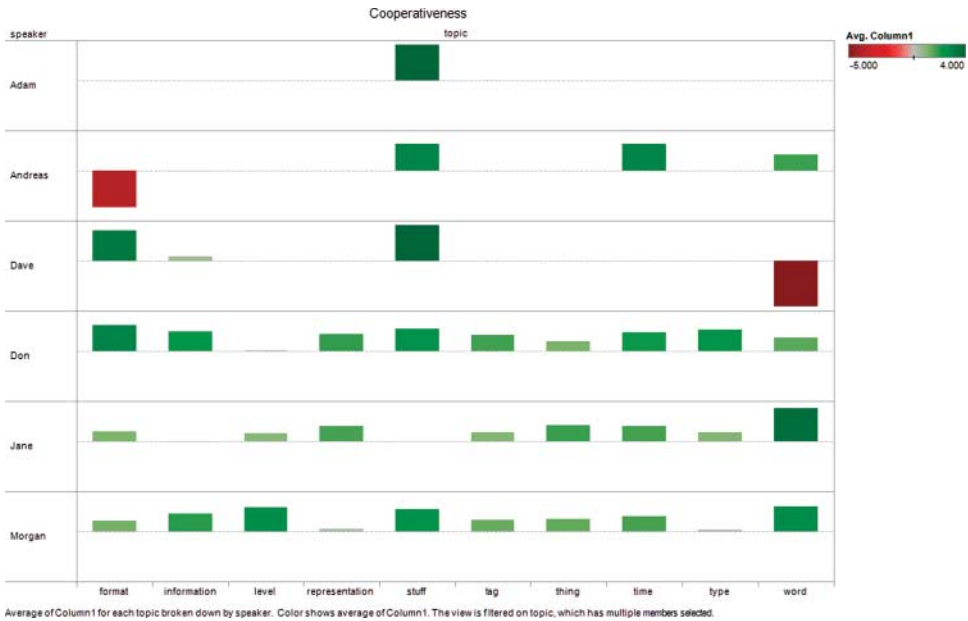


Figure 4. Cooperative behaviour on topics.

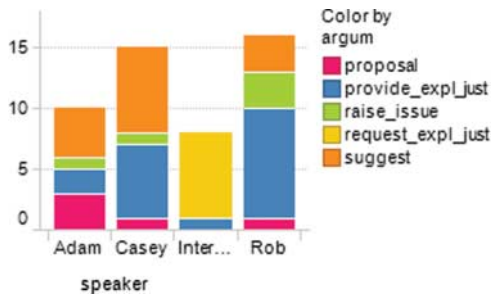


Figure 5. Distribution of argumentative categories per speaker.

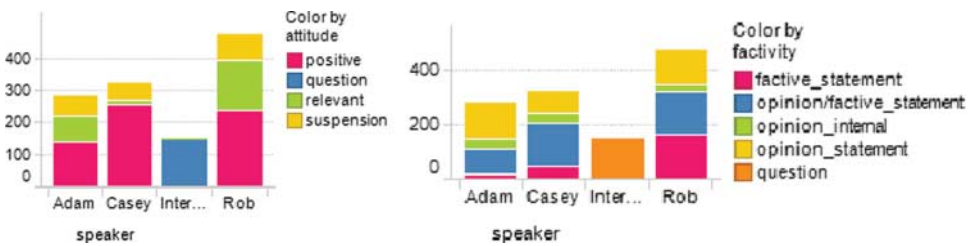


Figure 6. Distribution of sentiment and subjectivity categories per speaker.



Figure 7. Social network analysis of the conversation.

thickness of nodes represents the degree of involvement in the conversation. This representation shows that in the Google Wave team interview, there are, of course, interactions between the interviewer and interviewees, but also among interviewees themselves. The “nil” node represents the start of the conversation where the interviewer does not reply to anybody.

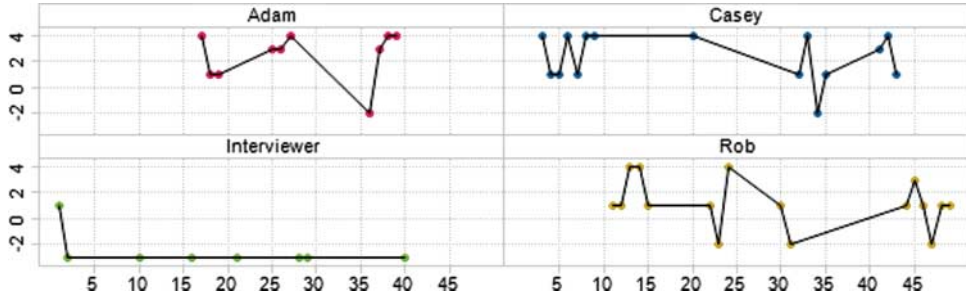


Figure 8. Individual conversation graphs.

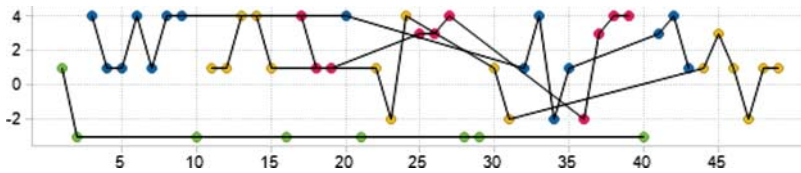


Figure 9. Merged conversation graph.



Figure 10. Dashboard for IBA applications.

For this particular example, we come back to the conversation graphs for the Google Wave team interview. In Figures 8 and 9, respectively, we show the individual conversation graphs and the merged one, where the  $x$ -axis represents the time expressed in turn's numbers and the  $y$ -axis displays the cooperativeness levels obtained from the mapping of Table 3. As can be expected, the interviewer shows the least cooperative behaviour as he challenges the interviewees. At the beginning, the interviewees' reaction is rather polite (or cooperative), while it "degenerates" into "flames" towards the end.

The previous examples clearly show that argumentative analysis can provide very powerful insights into conversational behaviours. Applied to social media and contact centres, this may result into a powerful tool for understanding essential aspects about conversational issues.

We also did an exercise to imagine how an IBA dashboard would look like. In Figure 10, we provide a mock-up of a web-based IBA dashboard, which displays the conversation content on the left, as well as the video (if any) of the recorded conversation. In the central part of the dashboard, we show gauges of participants' cooperativeness, a spider diagram for the participants' attitudes on the overall conversation, a social bond network diagram and the conversation graph. The idea is that the aggregated views can all be clicked in order to drill through and get to finer grained information (i.e. individual contributions).

## 6. Abstractive summarisation of conversations

In this section, we showcase an abstractive summarisation tool for digital conversations based on the A3 system, which produces high-quality memos based on both automatically extracted thematic and argumentative structures. Summaries represent a complementary way to provide users with simplified access to conversational content.

As pointed out by Maybury (2007), analysing and summarising conversations or dialogues are very challenging tasks. Most of existing summarisation techniques are tailored for the narrative genre and can hardly be adapted to the dialogue genre. Moreover, the techniques that are applicable at a large scale are based on extractive summarisation (Zechner 2002; Murray, Renals and Carletta 2005; Garg, Favre, Reidhammer and Hakkani-Tür 2009).

We argue that extractive techniques are inadequate for summarising conversations. In fact, this method that selects the most relevant sentences of a document for inclusion in the summary has severe limitations due to the intrinsic characteristics of the source data: conversations are not as coherent as ordinary narrative text (such as news or scientific articles) and obtaining a coherent text from conversations is practically impossible using the extractive approach. Moreover, any system that performs extractive summarisation must be evaluated against human-annotated test data sets and, as pointed out in Buist, Kraaij and Raaijmakers (2005), inter-annotator agreement is very low for this type of task, which makes test data nearly useless for evaluation. Intuitively, selecting salient content from conversations is a truly difficult task and subjective selection of excerpts leads to fairly different results.

### 6.1. Related work in abstractive summarisation of conversations

Abstractive summarisation is a more appropriate technique for conversational content. The problem of abstractive summarisation of conversations is apparently much harder to solve than its extractive version. It requires almost the full understanding of the source data. Moreover, as also recognised by Murray, Carenini and Ng (2010), in order to perform the generation of an abstractive summary of meetings, it is necessary to classify participants' contributions according to their informativeness and function in the conversation so that an appropriate ontology of the meeting could be adequately populated.

Abstractive summarisation of narrative texts is typically based on sentence compression and/or paraphrase (Mani and Maybury 1999). Liu and Liu (2009) applied this technique to the problem of abstractive summarisation. This work represents an important step in filling the gap between extractive and abstractive summarisation. However, they recognise the need of a sophisticated extraction and generation algorithm to achieve full abstractive summarisation. In fact, their approach is not appropriate for generating abstracts of conversations where turns are already highly compressed.

We conclude our review of related work by looking at Kleinbauer, Becker and Becker (2007), where a language-generation algorithm produces indicative abstracts of meeting transcriptions. The existing annotations of the AMI corpus<sup>15</sup> data are used to create a content representation that is subsequently fed to a sentence planner. The semantic representation produced is then rendered into a surface form with a surface generator. This work is in spirit very similar to ours with the notable exception that their semantic representation is not as rich as ours. Basically, only topics, very general meeting stages (opening, debating and closing) and temporal sequencing of topics discussed are used to generate the abstract.

## 6.2. A3 memos

We implemented an abstractive summarisation system that first generates a threaded description of the conversation's dynamics based on both thematic content and argumentative structure that we are able to automatically extract as described in the previous sections. Subsequently, the memo can be compressed by collapsing the thread portions that are considered to be too detailed to get an insight into the conversation such as requests–provisions of explanations.

The output of our system consists of several sections, describing the conversational settings, participants' number and names, statistics about the interactivity of participants (e.g. the degree of competitiveness), the topics discussed and the arguments. The arguments are grouped into episodes bound together by thematic cohesion. All this information is extracted by relying on the results of the argumentative algorithm described in Section 4.

Figure 11 displays a sample of a memo, which has been generated with the A3 system from the analysis of the ICSI meetings. The meeting data considered here consist of plain turns and speaker with no additional annotations other than the speaker identification.

The quality of abstractive summaries of conversations highly depends on how the conversation is analysed. Being able to build a narrative description of the discussion process allowed us to generate understandable memos.

In the above example, the conversation has been only turned into a narrative and no compression was done. The generated memo outlines the argumentative structure and can be used to generate a more succinct abstract where many details of the conversation are hidden. For instance, explanations and other types of elaborations around proposals or issues can be omitted. Moreover, the details of agreements and disagreements (e.g. the reasons) can be omitted. Finally, overlaps – recognised as backchannels – can also be removed.

A coarse-grained summary can be obtained by further summarising episodes. For instance, in the summary shown in Figure 11, episode 1 is further summarised in the following manner:

Don, Jane and Morgan discussed about “single timeline”. Jane raised an issue: “I remember seeing an example of this”. Don asked Jane to clarify and Morgan accepted Jane's explanation. Morgan agreed and provided an explanation. Don disagreed.

From the above abstract, it is possible to understand what type of interaction took place between the three participants without having to look at the details. From this abstract, one can figure out that the topic “single timeline” was particularly hot, that no agreement was reached at that point and that Jane's opinion was shared by Morgan, but not by Don.

**MEETING MEMO**

**GENERAL INFORMATION ON PARTICIPANTS**

The participants to the meeting are 7.

Participants less actively involved are Ami and Don who only intervened respectively for 38 and 68 turns.

**LEVEL OF INTERACTIVITY IN THE DISCUSSION**

The speaker that has held the majority of turns is Adam with a total of 722 turns, followed by Fey with a total of 561.

The speaker that has undergone the majority of overlaps is Adam followed by Jane.

The speaker that has done the majority of overlaps is Jane followed by Fey.

Jane is the participant that has been most competitive.

**DISCUSSION TOPICS**

The discussion was centred on the following topics: **schemas, action, things and domain.**

The main topics have been introduced by the most important speaker of the meeting.

The participant who introduced the main topics in the meeting is: **Adam.**

The most frequent entities in the whole dialogue partly coincide with the best topics, and are the following: **action, schema, things, 'source-path-goal', person, spg, roles, bakery, intention, specific, case, categories, information, idea.**

**ARGUMENTATIVE CONTENT**

The following participants: **Andreas, Dave, Don, Jane, Morgan** expressed their dissent 52 times. However Dave, Andreas and Morgan expressed dissent in a consistently smaller percentage.

The following participants: **Adam, Andreas, Dave, Don, Jane, Morgan** asked questions 55 times.

The remaining 1210 turns expressed positive content by proposing, explaining or raising issues. However Adam, Dave and Andreas suggested and raised new issues in a consistently smaller percentage.

The following participants: **Adam, Andreas, Dave, Don, Jane, Morgan** expressed acceptance 213 times.

***EPISODE ISSUE No. 1***

*In this episode we have the following argumentative exchanges between **Don, Jane and Morgan**:*

*Don raises the following **issue***

[um so - so it definitely had that as a concept .]

*then he, **overlapped** by Morgan, continues*

[so th ... it has a single timeline .]

*Morgan **accepts** the previous **explanation***

[uhhuh, .]

Figure 11. Automatically generated memo from the A3 system.

*Then Don elaborates the following **explanation***

[and then you can have lots of different sections.]

*then he continues*

[each of which have IDs attached to it .]

[and then you can refer from other sections to those IDs .]

[if you want to .]

[so that you start with - with a timeline tag .]

[timeline .]

[and\_then, you, have, a, bunch, of, times, .]

[I don't ... I don't remember exactly what my notation was .]

*Jane, overlapped by Don, raises the following **issue***

[oh; i, remember seeing an example of this .]

*Jane, overlapped by Dave, **suggests** the following explanation*

[actually we - we use a generalization of the - the sphere format .]

[yeah so there is something like that .]

[but it's um probably not as sophist... .]

*Don asks the following **question***

[well what does H.T.K. do for features ?]

*then he, overlapped by Jane, continues*

[or does it even have a concept of features ?]

*Jane provides the following **answer***

[they h... it has its own ... .]

*then she , overlapped by Morgan, continues*

[I mean Entropic has their own feature format that's called like s... s.d. or some s... s.f. or something like that.]

*Morgan **accepts** the previous **explanation***

[yeah.]

*Don, overlapped by Morgan, **provides** the following **explanation***

[you know we'll do something where we - some kind of data reduction where the prosodic features are sort ... uh either at the word level or at the segment level .]

*Morgan **agrees***

[right, .]

*then he continues*

[or - or something like that .]

[they're not going to be at the phone level and they're n- ... not going to be at the frame level when we get done with sort of giving them simpler shapes and things .]

[and so; the main thing is just being able ... .]

[well I guess the two goals .]

*Don **disagrees** with the previous **explanation***

[um one that Chuck mentioned is starting out with something that we don't have to start over that we don't have to throw away if other people want to extend it for other kinds of questions .]

*Morgan **agrees***

[right, .]

Figure 11. Continued



We are aware that this is only a first step towards accurate and succinct abstractive summaries of conversations. Work needs to be done at the level of DECISION labelling and to improve the discrimination of really argumentative from pragmatically irrelevant utterance, a choice that in some cases is hard to make on an automatic basis. Indeed, the DECISION label can be inferred either by looking at linguistic markers or by interpreting existing argumentative labels, that is, agreed proposals can be assumed to be also decided.

We also need to design an evaluation framework for abstractive summaries of conversations because the current metrics used for evaluating extractive summaries are not well adapted (e.g. ROUGE). We definitely agree with Murray et al. (2008) that the evaluation of abstractive summaries must be extrinsic, that is, based on human judgement or indirectly observed as its helpfulness in performing a given task.

### 7. Conclusions

In this article, we have presented the core language technology for analysing digital conversations and producing from their analysis intuitive visualisations and high-quality summaries for interaction mining.

We addressed the issue of capturing the conversational dynamics through the adoption of argumentation theory as the underlying model for making pragmatic analysis of conversations. We also made the case for the importance of such a type of analysis showing how the shortcomings of classical information retrieval techniques can be overcome by adopting our novel approach.

Our A3 system represents a viable solution for argumentative analysis of conversations and we provided an evaluation of its promising performance. Output of the A3 system can be considered as the cornerstone of IBA, which we have showcased through several examples of possible aggregations and visualisations. Among other possible applications of the A3 system, we presented a novel summarisation tool capable of automatically generating narrative descriptions of conversational data at different levels of compression.

We are currently working on new applications for social media analytics and quality monitoring of contact centre conversations. We also expect to start an evaluation campaign for assessing the

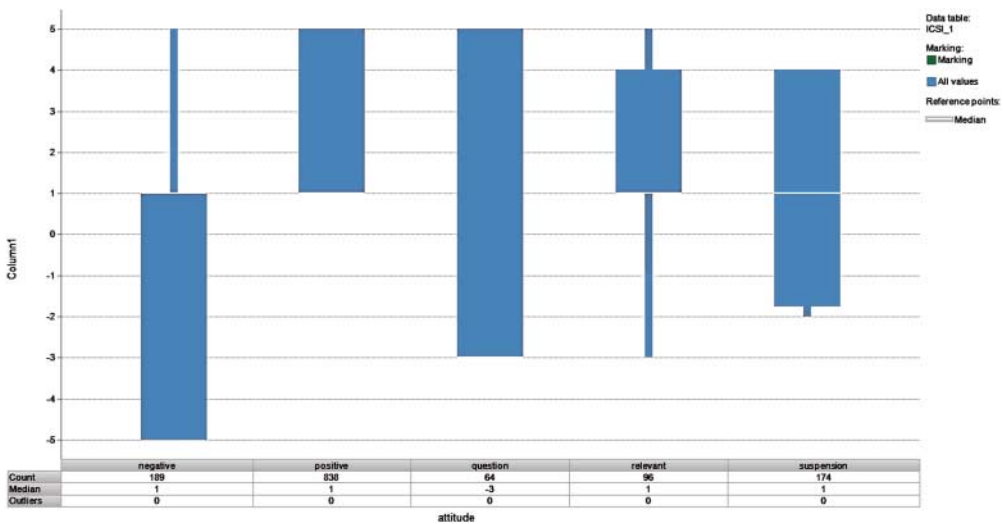


Figure 12. Correlation analysis of cooperative behaviour and sentiment.

quality of the abstractive summaries. We also intend to investigate how conversational graphs scale in terms of the length of conversation and the number of participants. We also plan to study the correlation between argumentative analysis and other types of analyses of interactions. For instance, we realised that there is a significant correlation between cooperative behaviour and sentiment from our data as shown in Figure 12. Possibly, argumentative analysis can be used as a tool for determining sentiment more accurately.

As a roadmap for future research and development in the emerging field of interaction mining, we suggest moving away from the standard content-based approaches so widespread in text mining technology in favour of more pragmatically oriented methods capable of taking the context and the interaction in due account. For their nature, conversations need full pragmatic understanding. The purpose of contributions in conversations cannot be recognised with semantic-only techniques. It requires the understanding of the context where the interaction takes place as well as recognising the purpose of individual contributions. Moreover, we believe that standard visualisation tools might not be adequate for visualising data mined from interactions and work needs to be done in order to design new types of visualisations.

## Notes

1. [en.wikipedia.org/wiki/Online\\_reputation\\_management](http://en.wikipedia.org/wiki/Online_reputation_management)
2. The term Digital Conversation is also used in marketing to denote new ways of advertising based on dialogue between producers and consumers. See [en.wikipedia.org/wiki/Digital\\_conversation](http://en.wikipedia.org/wiki/Digital_conversation)
3. Private Branch eXchange.
4. [http://en.wikipedia.org/wiki/Usenet\\_newsgroup](http://en.wikipedia.org/wiki/Usenet_newsgroup)
5. <http://disqus.com/>
6. <http://www.reddit.com/>
7. This judgement is based on two email usage reports such as UC Berkeley's "How much information?" [http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/\(59kb\)](http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/(59kb)) and the more recent Ferris Research study on email usage: [http://www.ferris.com/research-library/industry-statistics/\(10kb\)](http://www.ferris.com/research-library/industry-statistics/(10kb)).
8. The remaining COBs might be more difficult to spot, but we see it as a challenge that we might consider reasonable as future work and which would be seen as out of reach with current speech analytics approaches.
9. This is not a formalism, but just a way to provide intuitions about the MDS. The symbol "∣" is meant to denote disjunction.
10. The reported dialogue actually occurred during a real meeting witnessed by the authors. However, the dialogue has been neither taped nor transcribed in its original form, but reconstructed by the authors in its essential features.
11. Overlaps are already computed in the transcription phase with time stamps, in this way they can be located precisely in the overlapped utterance. More interesting is the notion of "overlapped utterance" that underlies our approach. The identification of overlaps in the speech wave is outside the scope of this article, but the interested reader can refer to Delmonte, Bristot and Tonelli (2007b); Delmonte (2005); Delmonte, Bristot, Piccolino Boniforti, and Tonelli (2005); Delmonte (2003).
12. For an evaluation of a system's performance, see Delmonte et al. (2007a).
13. The chart in Figure 4 has been created with the TIBCO Spotfire BI suite.
14. <https://sites.google.com/a/pressatgoogle.com/googlewave/home/meet-the-google-wave-team#roundtable>
15. <http://corpus.amiproject.org/>

## References

- Ailomaa, M. (2009), 'Answering Questions About Archived, Annotated Meetings', PhD thesis N° 4512, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

- Ailomaa, M., and Rajman, M. (2009), 'Enhancing Natural Language Search in Meeting Data with Visual Meeting Overviews', in *Proceedings of the 10th Annual Conference of the NZ ACM Special Interest Group on Human-Computer Interaction (CHINZ 2009)*, Auckland, New Zealand, 6-7 July.
- Anh Le, Q., and Popescu-Belis, A. (2009), 'Automatic vs. human question answering over multimedia meeting recordings', in *Proceedings of INTERSPEECH 2009, 10th Annual Conference of the International Speech Communication Association (ISCA)*, Brighton, UK, 6-10 September, pp. 624-627.
- Anscombe, J.-C., and Ducrot, O. (1997), *L'argumentation dans la langue*, Sprimont Mardaga Pierre.
- Bachler, M.S., Shum, S.J.B., Roure, D.C.D., Michaelides, D.T., and Page, K.R. (2003), 'Ontological Mediation of Meeting Structure: Argumentation, Annotation, and Navigation', in *Proceedings of the 1st International Workshop on Hypermedia and the Semantic Web (HTSW2003)*, 26-30 August 2003, Nottingham, UK.
- Bresnan, J. (2000), *Lexical-functional syntax*, Hoboken, NJ: Wiley Blackwell.
- Buist, A.H., Kraaij, W., and Raaijmakers, S. (2005), 'Automatic Summarization of Meeting Data: A feasibility Study', in *Proceedings of the 15th CLIN Conference*.
- Buckingham Shum, S., Slack, R., Daw, M., Juby, B., Rowley, A., Bachler, M., Mancini, C., Michaelides, D., Procter, R., De Roure, D., Chown, T., and Hewitt, T. (2006), 'Memetic: An Infrastructure for Meeting Memory', in *Proceedings of 7th International Conference on the Design of Cooperative Systems*, Carry-le-Rouet, France, 9-12 May.
- Chu-Carroll, J., and Carberry, S. (1996), 'Conflict Detection and Resolution in Collaborative Planning', in *Intelligent Agents II Agent Theories, Architectures, and Languages*, eds. M. Wooldridge, J. Müller and M. Tambe, LNCS 1037, Berlin: Springer Verlag, pp. 111-126.
- Clark, A., and Popescu-Belis, A. (2004), 'Multi-level Dialogue Act Tags', in *Proceedings of SIGDIAL'04 (5th SIGdial Workshop on Discourse and Dialogue)*, Cambridge, MA, USA, pp. 163-170.
- Dahlgren, K. (1988), *Naive Semantics for Natural Language Understanding*, Boston: Kluwer.
- Dascal, M. (1977), 'Conversational Relevance', *Journal of Pragmatics*, 1, 309-328.
- Dascal, M. (1992), 'On the Pragmatic Structure of Conversation', in *Searle on conversation*, eds. J. Searle et al., Amsterdam: Benjamins.
- Delmonte, R. (2003), 'Parsing Spontaneous Speech', in *Proceedings of Special Track on Robust Methods in Processing of Natural Language Dialogues at EUROSPEECH 2003*, eds. V. Pallotta, A. Popescu-Belis and M. Rajman, Genève: ESCA, pp. 1-6.
- Delmonte, R. (2005), 'Parsing Overlaps', in *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen*, eds. B. Fisseni, H.C. Schmitz, B. Schroeder, and P. Wagner, Sprache, Sprechen und Computer, Frankfurt am Main, pp. 497-512.
- Delmonte, R. (2006), 'Building domain ontologies from text analysis: an application for question answering', in *Natural Language Understanding and Cognitive Science, Proceedings of the 3rd International Workshop on Natural Language Understanding and Cognitive Science, NLUCS 2006, in conjunction with ICEIS 2006*, Paphos, Cyprus, May 2006, ed. B. Sharp, INSTICC Press, pp. 3-16.
- Delmonte, R. (2007), *Computational Linguistic Text Processing – Logical Form, Semantic Interpretation, Discourse Relations and Question Answering*, New York: Nova Science Publishers.
- Delmonte, R. (2008), 'Semantic and pragmatic computing with GETARUNS', in *Semantics in Text Processing. STEP 2008 Conference Proceedings, Research in Computational Semantics*, eds. J. Bos and R. Delmonte, College Publications, pp. 287-298.
- Delmonte, R. (2009), *Computational Linguistic Text Processing – Lexicon, Grammar, Parsing and Anaphora Resolution*, New York: Nova Science Publishers.
- Delmonte, R., Harabagiu, S., and Nicolae, G. (2007a), 'A Linguistically based Approach to Detect Causality Relations in Unrestricted Text', in *Proceedings of the Sixth Mexican International Conference on Artificial Intelligence (MICAI 2007)*, eds. A. Gelbukh and A.F.K. Morales, pp. 173-184.
- Delmonte, R., Bristot, A., and Tonelli, S. (2007b), 'Overlaps in AVIP/IPAR, the Italian Treebank of Spontaneous Speech', in *Proceedings of SRSL7 – Semantic Representation of Spoken Language*, eds. M. Alcantara Pla and T. Declerk, -Salamanca: CAEPIA, pp. 29-38.
- Delmonte, R., Bristot, A., Piccolino Boniforti, M.A., and Tonelli, S. (2005), 'Modeling Conversational Styles in Italian by means of Overlaps', Padova: AISV, CNR, pp. 11-30.

- De Moor, A., and Efimova, L. (2004), 'An Argumentation Analysis of Weblog Conversations', in *Proceedings of the 9th International Working Conference on the Language-Action Perspective on Communication Modeling (LAP)*, eds. M. Aakhus and M. Lind, New Brunswick, NJ: Rutgers University, The State University of New Jersey.
- Dessalles, J.-L. (2007), 'A Computational Model of Argumentation in Everyday Conversation: A Problem-centred Approach', Technical Report ParisTech-ENST 2007–D–017, November 2007.
- van Eemeren, F.H., and Grootendorst, R. (2004), *A Systematic Theory of Argumentation: The Pragmadiological Approach*, Cambridge, UK: University Press.
- Fernández, R., Frampton, M., Ehlen, P., Purver, M., and Peters, P. (2008), 'Modelling and Detecting Decisions in Multi-Party Dialogue', in *Proceedings of the 9th SIGdial Workshop on Discourse and Dialogue*, Columbus, OH, June 2008.
- Fujita, K., Nishimoto, K., Sumi, Y., Kunifuji, S., and Mase, K. (1998), 'Meeting Support by Visualizing Discussion Structure and Semantics', in *Proceedings of the 2nd International Conference on Knowledge-Based Intelligent Electronic Systems (KES '98)*, April 21–23, Adelaide, Australia, Vol. 1, pp. 417–422.
- Galley, M., McKeown, K., Hirschberg, J., and Shriberg, E. (2004), 'Identifying Agreement and Disagreement in Conversational Speech: Use of Bayesian Networks to Model Pragmatic Dependencies', in *Proceedings of 42nd Meeting of the ACL*, 21–26 July, Barcelona.
- Garg, N., Favre, B., Reidhammer, K., and Hakkani-Tür, D. (2009), 'ClusterRank: A Graph Based Method for Meeting Summarization', in *Proceedings of Interspeech 2009*, Brighton, UK.
- Ginzburg, J. (1996), 'Interrogatives: Questions, Facts, and Dialogue', in *The Handbook of Contemporary Semantic Theory*, ed. L. Shalom, Oxford: Blackwell, pp. 385–422.
- Goffman, E. (1981), *Forms of Talk*, Philadelphia, PA: University of Pennsylvania Press.
- Gordon, T., and Karacapilidis, N. (1997), 'The Zenon Argumentation Framework', in *Proceedings of the Sixth International Conference on AI and Law*, ACM Press, pp. 10–18.
- Grasso, F., Cawsey, A., and Jones, R. (2000), 'Dialectical Argumentation to Solve Conflicts in Advice Giving: A Case Study in the Promotion of Healthy Nutrition', *International Journal of Human-Computer Studies*, 53(6), 1077–1115.
- Grice, P. (1975), 'Logic and Conversation', in *Syntax and Semantics*, eds. P. Cole and J. Morgan (Vol. 3), New York: Academic Press.
- Hakkani-Tür, D. (2009), 'Towards Automatic Argument Diagramming of Multiparty Meetings', in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Taipei, Taiwan, April.
- Harabagiu, S., and Hickl, A. (2006), 'Methods for Using Textual Entailment in Open-domain Question Answering', in *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual meeting of the Association for Computational Linguistics*.
- Hillard, D., Ostendorf, M., and Shriberg, E. (2003), 'Detection of Agreement vs. Disagreement in Meetings: Training with Unlabeled Data', in *Proceedings of HLT-NAACL 2003*.
- Hobbs, J.R. (1993), 'Intention, Information, and Structure in Discourse: A First Draft', Technical Report Artificial Intelligence Center, SRI International.
- Hsueh, P., and Moore, J.D. (2006), 'Automatic Topic Segmentation and Labeling in Multiparty Dialogs', in *Proceedings of Spoken Language Technology Workshop*, IEEE Volume December 2006, pp. 98–101.
- Hsueh, P., and Moore, J.D. (2007), 'Automatic Decision Detection in Meeting Speech', in *Proceedings of MLMI 2007*, pp. 168–179.
- Hsueh, P., and Moore, J.D. (2009), 'What Decisions Have You Made: Automatic Decision Detection in Conversational Speech', in *Proceedings of HLT-NAACL*, 2009.
- Janin, A., Baron, D., Edwards, J., Ellis, D., Gelbart, D., Morgan, N., Peskin, B., Pfau, T., Shriberg, E., Stolcke, A., and Wooters, C. (2003), 'The ICSI Meeting Corpus', in *Proceedings of IEEE/ICASSP 2003*, 6–10 April 2003, Hong Kong, Vol. 1, pp. 364–367.
- Kleinbauer, T., Becker, S., and Becker, T. (2007), 'Combining Multiple Information Layers for the Automatic Generation of Indicative Meeting Abstracts', in *Proceedings of the 11th European Workshop on Natural Language Generation (ENLG07)*, June 17th–20th, 2007, Dagstuhl, Germany.

- Kunz, W., and Rittel, H.W.J. (1970), 'Issues as Elements of Information Systems', Technical Report WP-131, Berkeley: University of California.
- Larsson, S. (2002), 'Issue-based Dialogue Management', Ph.D. thesis, University of Gothenburg, Sweden.
- Liu, F., and Liu, Y. (2009), 'From Extractive to Abstractive Meeting Summaries: Can It Be Done by Sentence Compression?', in *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, Suntec, Singapore, 4 August 2009, pp. 261–264.
- Lo Cascio, V. (1991), *Grammatica dell'Argomentare: strategie e strutture*. Firenze: La Nuova Italia.
- Mabry, E.A. (1997), 'Framing Flames: The Structure of Argumentative Messages on the Net', *Journal of Computer-Mediated Communication*, 2(4). doi: 10.1111/j.1083-6101.1997.tb00193.x.
- Mani, I., and Maybury, M. (eds.) (1999), *Advances on Automatic Text Summarization*, Boston, MA: MIT Press.
- Mann, W., and Thompson, S. (1987), 'Rhetorical Structure Theory: A Theory of Text Organisation', in *The Structure of Discourse*, ed. L. Polanyi, Norwood, NJ: Ablex.
- Maybury, M. (2007), 'Keynote on Searching Conversational Speech', in *Proceedings Workshop on Searching Spontaneous Conversational Speech* part of the ACM SIGIR '07 Conference, 27 July 2007, Amsterdam.
- Mazzotta, I., de Rosis, F., and Carofiglio, V. (2007), 'A User-adapted Persuasion System in the Healthy Eating Domain', *IEEE Intelligent Systems*, 22(6), 42–51.
- Michaelides, D., Buckingham Shum, S., Juby, B., Mancini, C., Slack, R., Bachler, M., Procter, R., Daw, M., Rowley, A., Chown, T., De Roure, D., and Hewitt, T. (2006), 'Memetic: Semantic Meeting Memory', in *Proceedings of the 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE 2006)*, 26–28 June 2006, Manchester, UK, pp. 382–387.
- Mishne, G., and Glance, N. (2006), 'Leave a Reply: An Analysis of Weblog Comments', in *Proceedings of Third Annual Workshop on the Blogging Ecosystem*.
- Murray, G., Carenini, G., and Ng, R. (2010), 'Interpretation and Transformation for Abstracting Conversations', in *Proceedings of Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL (HLT-NACL)*, Los Angeles, CA, June 2010, pp. 894–902.
- Murray, G., Renals, S., and Carletta, J. (2005), 'Extractive Summarization of Meeting Recordings', in *Proceedings of the 9th European Conference on Speech Communication and Technology*, (EUROSPEECH 05), pp. 593–596.
- Murray, G., Kleinbauer, T., Poller, P., Renals, S., Becker, T., and Kilgour, J. (2008), 'Extrinsic Summarization Evaluation: A Decision Audit Task', in *Proceedings of MLMI 2008, Special Session on User Requirements and Evaluation*, Utrecht, Netherlands.
- Navigli, R., and Velardi, P. (2003), 'An Analysis of Ontology-based Query Expansion Strategies', in *Proceedings of Workshop on Adaptive Text Extraction and Mining (ATEM 2003)*, in the 14th European Conference on Machine Learning (ECML 2003), Cavtat-Dubrovnik, Croatia, 22–26 September 2003, pp. 42–49.
- Niekrasz, J., Purver, M., Dowding, J., and Peters, S. (2005), 'Ontology-based Discourse Understanding for a Persistent Meeting Assistant', in *Proceedings of the AAIL Spring Symposium Persistent Assistants: Living and Working with AI*, Stanford.
- Pallotta, V. (2006), 'Framing Arguments', in *Proceedings of the International Conference on Argumentation ISSA*, June 2006, Amsterdam, Netherlands.
- Pallotta, V., Seretan, V., and Ailomaa, M. (2007), 'User Requirements Analysis for Meeting Information Retrieval Based on Query Elicitation', in *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL 2007)*, Prague, pp. 1008–1015.
- Pallotta, V., Vrieling, L., and Delmonte, R. (2011), 'Interaction Mining: Making Business Sense of Customers Conversations through Semantic and Pragmatic Analysis', in *Business Intelligence Applications and the Web: Models, Systems and Technologies*, eds. M. Zorrilla, J.-N. Mazón, Ó. Ferrández, I. Garrigós, F. Daniel, and J. Trujillo, IGI Global, pp. 122–146.
- Pallotta, V., Ghorbel, H., Ballim, A., Lisowska, A., and Marchand-Maillet, S. (2004), 'Towards Meeting Information Systems: Meeting Knowledge Management', in *Proceedings of ICEIS 2005*, Porto, Portugal, pp. 464–469.

- Popescu-Belis, A., Baudrion, P., Flynn, M., and Wellner, P. (2008), 'Towards an Objective Test for Meeting Browsers: The BET4TQB Pilot Experiment', in *Machine Learning for Multimodal Interaction IV*, ser. LNCS 4892, eds. A. Popescu-Belis, H. Bourlard, and S. Renals, Berlin: Springer, pp. 108–119.
- Purver, M., Dowding, J., Niekrasz, J., Ehlen, P., Noorbaloochi, S., and Peters, S. (2007), 'Detecting and Summarizing Action Items in Multi-party Dialogue', in *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, Antwerp, Belgium, September 2007.
- Rafaëli, A., Ziklik, L., and Doucet, L. (2007), 'The Impact of Call Center Employees' Customer Orientation Behaviors on Customer Satisfaction', *Journal of Service Research*, 10(3), pp. 239–255.
- Rienks, R., and Verbree, D. (2006), 'About the Usefulness and Learnability of Argument-diagrams from Real Discussions', in *Proceedings of the 3rd International Machine Learning for Multimodal Interaction Workshop (MLMI 2006)*, 1–4 May 2006, Bethesda (MD), USA.
- Rienks, R., Heylen, D., and van der Weijden, E. (2005), 'Argument Diagramming of Meeting Conversations', in *Proceedings of the Multimodal Multiparty Meeting Processing Workshop at the 7th International Conference on Multimodal Interfaces (ICMI 2005)*, eds. A. Vinciarelli and J.-M. Odobez, 3–7 October 2005, Trento, Italy, pp. 85–92.
- Salton, G., and Buckley, C. (1988), 'Term-weighting Approaches in Automatic Text Retrieval', *Information Processing & Management*, 24(5), 513–523.
- Sanchez, M., Degaki, D., and Oliveira, F. (2010), 'Around the World in 140 Characters: How Companies Manage Twitter in 17 Different Countries', Technical Report from Mercedes Sanchez Usabilidad and UXalliance.
- Schegloff, E., and Sacks, H. (1973), 'Opening up Closings', *Semiotica*, 8, 289–327.
- Schultz, G. (2003), *Chapter 4. The Customer Care and Contact Center Handbook*. Milwaukee, WI: ASQ Quality Press.
- Shriberg, E., Dhillon, R., Bhagat, S., Ang, J., and Carvey, H. (2004), 'The ICSI Meeting Recorder Dialog Act (MRDA) Corpus', in *Proceedings of 5th SIGdial Workshop on Discourse and Dialogue*, eds. M. Strube and C. Sidner, April 30–May 1, Cambridge, MA, pp. 97–100.
- Stent, A.J., and Allen, J.F. (2000), 'Annotating Argumentation Acts in Spoken Dialog', Technical Report 740 University of Rochester, December 2000.
- Takeuchi, H., Subramaniam, L.V., Nasukawa, T., and Roy, S. (2009), 'Getting Insights from the Voices of Customers: Conversation Mining at a Contact Center', *Information Science*, 179(11), 1584–1591.
- Tur, G., Stolcke, A., Voss, L., Peters, S., Hakkani-Tür, D., Dowding, J., Favre, B., Fernández, R., Frampton, M., Frandsen, M., Frederickson, C., Graciarena, M., Kintzing, D., Leveque, K., Mason, S., Niekrasz, J., Purver, M., Riedhammer, K., Shriberg, E., Tien, J., Vergyri, D., and Yang, F. (2010), 'The CALO Meeting Assistant System', *IEEE Transactions on Audio, Speech and Language Processing*, 18(6), 1601–1611.
- Verbree, A. (2006), 'On the Structuring of Discussion Transcripts Based on Utterances Automatically Classified', Master's thesis, University of Twente, The Netherlands.
- Verbree, A.T., Rienks, R.J., and Heylen, D.K.J. (2006), 'Dialogue-act Tagging Using Smart Feature Selection: Results on Multiple Corpora', in *Proceedings of the First International IEEE Workshop on Spoken Language Technology SLT 2006*, 10–13 December 2006, Palm Beach, Aruba, IEEE Computer Society, pp. 70–73.
- Verint Inc. (2011), 'Customer Interaction Analytics™. A 360° Approach for Analyzing the Voice of the Customer across Multiple Communications Channels', White paper available at [www.verint.com](http://www.verint.com).
- Wellner, P., Flynn, M., Tucker, S., and Whittaker, S. (2005), 'A Meeting Browser Evaluation Test', in *Proceedings of CHI 2005*, Portland, OR, pp. 2021–2024.
- Zechner, K. (2002), 'Automatic Summarization of Open-domain Multiparty Dialogues in Diverse Genres', *Computational Linguistics*, 28(4), pp. 447–485.