Deep Linguistic Processing with GETARUNS for spoken dialogue understanding

Rodolfo Delmonte, Antonella Bristot, °Vincenzo Pallotta

Department of Language Science
Università "Ca' Foscari"
30123 - VENEZIA
delmont@unive.it
°Department of Computer Science
Webster University, Geneva
Switzerland
pallotta@webster.ch

Abstract

In this paper we will present work carried out to scale up the system for text understanding called GETARUNS, and port it to be used in dialogue understanding. The current goal is that of extracting automatically argumentative information in order to build argumentative structure. The long term goal is using argumentative structure to produce automatic summarization of spoken dialogues. Very much like other deep linguistic processing systems, our system is a generic text/dialogue understanding system that can be used in connection with an ontology – WordNet - and other similar repositories of commonsense knowledge. We will present the adjustments we made in order to cope with transcribed spoken dialogues like those produced in the ICSI Berkeley project. In a final section we present preliminary evaluation of the system on two tasks: the task of automatic argumentative labeling and another frequently addressed task: referential vs. non-referential pronominal detection. Results obtained fair much higher than those reported in similar experiments with machine learning approaches.

1. Introduction

In this paper we will present work carried out to scale up the system for text understanding called GETARUNS, and port it to be used in dialogue understanding. Very much like other deep linguistic processing systems (Allen et al., 2007), our system is a generic text/dialogue understanding system that can be used in connection with an ontology – WordNet - and/or a repository of commonsense knowledge like CONCEPTNET. Word sense disambiguation takes place at the level of semantic interpretation and is represented in the Discourse Model. We will present the adjustments we made in order to cope with transcribed spoken dialogues like those produced in the ICSI Berkeley project. The low level component is organized according to LFG theory; the system also does pronominal binding, quantifier raising and temporal interpretation. The high level component is where the Discourse Model is created from the Logical Form of an utterance. For longer sentences the system switches from the topdown to the bottomup system. In case of failure it will backoff to the partial system which produces a very lean and shallow semantics with no inference rules.

The system presented here has been developed for over two decades with the goal of developing a broad-coverage, domain general natural language understanding system. The underlying grammar, lexicon, the semantics and all intermediate modules are intended to be domain-general and to be easily portable to different applications. As is the case with all rule-based systems, (but see also Allen et al., 2007), we have no need to 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 and 2008).

However, when we started last year to use the system to parse ICSI dialogues, we realized that the semantic representation and the output of the parser were both inadequate. So we started to work at deficiencies that we detected in an empirical manner. This approach made us aware of the peculiarities of spoken dialogue texts such as the ones made available in ICSI Berkeley project. These dialogues are characterized by the need to argument in a exhaustive manner the topics to be debated which are the theme of each multiparty dialogue. The mean length of utterances/turns in each dialogue we parsed was rather long. This makes ICSI

dialogues hard to compute. From a count of number of words x turn, we came up with the following mean figures:
- percent of turns made of one single word: 30%
- percent of turns made of up to three words: 40%
- number of words x turn overall: 7
- number of words x turn after subtracting short utterances: 11

These values correspond to those found for PennTreebank corpus where we can count up to 94K sentences for 1M words – again 11 words per sentence. In analyzing ICSI, 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 spoken dialogues is a particularly hard task which requires at least the following information to be made available:
- adequate treatment of fragments;
- adequate treatment of short turns, in particular one-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 one or more such expressions, or cases when they are found inside the utterance;
- adequate treatment of overlaps;
- adequate treatment of speaker identity for pronominal coreference;

In addition, we decided that every dialogue turn had to receive one polarity label, indicating negativity or positivity, and this is computed by looking into a dictionary of polarity items.

We will address each such topics in what follows.

3. The Spoken Dialogue Additions

We will proceed by addressing each problem presented above in the order with which it is coped with in the system, i.e. as follows:
- overlaps
- dialogue act labeling
- fragment analysis
- disfluency treatment
- pronominal binding special routines
- non-referential linguistic element
- anaphora resolution routines
- current speaker as Subject of Point of View

3.1 The Algorithm for Overlaps

Overlaps are an important component of all spoken dialogue analysis. In all dialogue transcription, overlaps are treated as a separate turn from the one in which they occur, which usually follows it. This is clearly wrong

from a computational point of view. For this reason, when computing overlaps we set as our first goal that of recovering the temporal order. This is done because:
- overlaps may introduce linguistic elements which influence the local context;
- 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 we built looks at time stamps, and every time the following turn begins at a time preceding the ending time of 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 current utterance and continues with the overlapping turn. This may be reiterated in case another overlap follows which again begins before the end of current utterance. Eventually, it returns to the analysis of the current turn with the remaining portion of current utterance.

In Table 1 below we present data related to overlaps for the first 10 dialogues we computed. 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 being regarded as normal and non-disrupting of the current speaker's conversational plan.

	total	Cont	Int	Int/Con	Int/Chng	Int/O
turns	**13158**	-	-	-	-	-
while	**1624**	**1369**	**46**	**87**	**22**	**63**
after	1461	-	-	-	-	-

Table 1. Overlaps and their effects on Planning. We use the following abbreviations: Cont = continue; Int = Interrupt; Int/Con = Interrupt and Continue; Int/Chng = Interrupt and Change; Int/O = Interrupt Others

On a total number of 13158 turns we thus computed 3085 overlaps divided up nicely almost half and half for each of the two classes. Then we proceeded by subdividing WHILE overlaps into 5 subclasses where Continue indicates 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 Continues his/her plan in a following turn; Inter_Change, the current speaker is interrupted and changes his/her plan, by either changing subject topic, or answering the overlapper. Eventually

we had Inter_Other which indicates cases in which dialogue is interrupted by other speakers.

As can be easily noticed, the case constituted by Inter_Change which is the most interesting from a semantic and pragmatic point of view is in fact the less frequent. 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.

3.2 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, acknowledgements expressions and other similar speech acts. In case this procedure has success, no further computation takes place. However, this only applies to utterances shorter than 5 words, and should be made up only of such special words. No other linguistic element should be present apart from non-words, that is words which are only partially produced and have been transcribed with a dash at the end.

- graceful failure procedures for ungrammatical sentences, which might be full-fledged utterances but semantically non interpretable due to the presence of repetitions, false starts and similar disfluency phenomena. Or else they may be just fragments, i.e. partial or incomplete utterances, hence non-interpretable as such; this is done by imposing grammatical constraints of wellformedness in the parser;

- failure procedures for utterances which are constituted just by disfluency items and no linguistically interpretable words. These must be treated as semantically empty utterances and are recognizable by the presence of orthographic signs indicating that the word/s have not been completed and are just incomprehensible; this is done by inspecting the input in search of special orthographic marks and preventing the utterance to be passed down to the partial/deep parser.

On the contrary, we implemented a principled treatment of elliptical utterances and contribute one specific speech act or communicative act. They may express agreement/ disagreement, acknowledgement, assessment, continuers etc. All these items are computed as being complements of abstract verb SAY which is introduced in the analysis, and has as subject, the name of current speaker.

4. Automatic Argumentative Annotation

At first we shall provide a state of the art and then we shall comment in detail our approach.

4.1 Detecting Argumentative structure – issues and theories

As shown by Rosemberg and Silince (1999), tracking argumentative information from meeting discussions is of central importance for building summaries of project memories since, in addition to the "strictly factual, technical information", these memories must also store relevant information about decision-making processes. In a business context, the information derived from meetings is useful for future business processes, as it can explain phenomena and past decisions and can support future actions by mining and assessment (Pallotta et al., 2004).

In a section below we will describe in detail how discourse processing takes place. Here we want to highlight the main features of this process. This first level of processing is based on the shallow dialogue model proposed in (Armstrong, 2003), of which it is a modified version. This model provides a simple operational structure of dialogues based on three categories:

• *a dialog* is a non-empty set of episodes; a new episode is identified by a topic/speaker shift.

• *an episode* is a non empty set of turns; turns are individuated at prosodic level – more on turns below.

• *a turn* is a non-empty sequence of clauses/utterances and their boundary is a long pause.

In addition to the shallow dialogue model, we consider the adoption of a deeper structured representation based on argumentation theory. We assume that meeting dialogues are better viewed from the Collaborative Decision Making (CDM) perspective. In CDM, a meeting is defined as a multi-party (multi-agent) decision making process: a collaborative process, where agents follow a series of communicative actions in order to establish a common ground on the dimension of the problem. The main four dimensions of CDM process are:

• an overall task issue;

• a set of alternative proposals;

• a set of arguments in favor or against each proposals;

• a collection of choice criteria (perspectives and preferences) settled upon the participants;

• a decision (or evaluation) function that combines criteria to judge the alternatives.

This definition focuses on the processes, which take place during meetings and how these processes contribute to the accomplishment of a joint goal. In order to capture the above dimensions, we then adopted and extended a suitable argumentative model of discussions, namely the IBIS model proposed by (Kunz and Rittel, 1970). The IBIS model provides us with an abstract description of the discussion's rationale by

outlining the important points discussed, the conflicts arisen and, hopefully solved, and the decisions that have been made. The IBIS model abstracts from the dynamics of the discussion, which needs to be modeled as well in order to extract the IBIS structures from meeting events. Relevant meeting events are special types of Dialogue Acts that have an argumentative force. This type of Dialogue Acts called Argumentative Acts, are backward-looking acts with forward-looking expectations (Goffman 1981).

Within the Adjacency Pairs model (Schegloff & Sacks 1973), the importance of tracking agreement and disagreement in discussions has been recognized also in (Galley et al., 2004; Hillard, Ostendorf, and Shriberg, 2003). Although these methods have the great advantage of being automatic, they only partially help in reconstructing the argumentative information we need in order to answer real user queries. This model has been adopted by (Niekrasz et al. 2005) for the real-time reconstruction of an argumentative structure by overhearing discussions in design meetings. Finally, (Rienks and Verbree 2006) propose the Twente Annotation Schema that is based on fewer categories but more relation types being inspired by the Rhetorical Structure Theory (Mann and Thompson 1988).

The argumentative structure defines the different patterns of argumentation used by participants in the dialogue, as well as their organization and synchronization in the discussion. The limits of sequential analysis of conversation (Schegloff & 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 is considered. In fact, when analyzing dialogues, adjacency pairs are not enough to represent the hierarchical structure of the discussion. To that end we need to add a relation that links non adjacent pairs. 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 visualization 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, 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.

When using the IBIS mark-up labels, a meeting is decomposed into several stages such as issues, proposals, and positions, each stage being possibly related to specific aggregations of elementary dialogue acts. Moreover, argumentative interactions may be viewed as specific parts of the discussion where several dialogue acts are combined to build such an interaction; as for instance, a disagreement could be seen as an aggregation of several acts of reject and accept of the same proposal. From this perspective, we elaborated an argumentative coding scheme, the Meeting Description Schema (Pallotta et al. 2004), which takes into account the different stages (or episodes) defined by the IBIS model and extend the concept of adjacency pairs to relate these episodes to each other and to the corresponding argumentative function.

In MDS, the argumentative structure of a meeting is composed of a set of topic discussion episodes (a discussion about a specific topic). In each discussing topic, there exists a set of issue discussion 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 either express their positions and standpoints 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 (for example a rejection to a proposed alternative in a discussing issue) or questions and answers.

4.2 Our Approach

Automatic Argumentative Annotation, is carried out by a special module activated at the very end of the computation of the each dialogue. This module takes as input the complete semantic representation produced by the system recorded in Prolog facts in the Discourse Model (hence DM). The elements of semantic representation we use are the following ones:

- all facts in Situation Semantics contained in the Discourse Model, which include individuals, sets, classes, cardinality, properties related to entities by means of their semantic indices;

- facts related to spatiotemporal 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 (for an evaluation of system's performance see Delmonte et al. 2007);
- dialogue acts labels associated to each utterance or turn following ICSI classification;
- overlaps information computed at utterance level;
- topic labels associated to semantic indices of each entity marked as topic of discourse;
- all utterances with their indices as they have been automatically split by the system.

To produce Argumentative annotation, the system uses the following 21 Discourse Relations labels:

statement, narration, adverse, result, cause, motivation, explanation, question, hypothesis, elaboration, permission, inception, circumstance, obligation, evaluation, agreement, contrast, evidence, hypoth, setting, prohibition

These are then mapped onto five general argumentative labels. In addition we use the label DISFLUENCY for all those turns that contain fragments which are non-sentences and are semantically non interpretable.

ACCEPT,
REJECT/DISAGREE,
PROPOSE/SUGGEST,
EXPLAIN/JUSTIFY,
REQUEST
DISFLUENCY

The algorithm works in the following manner:
1. It recovers Dialogue Acts for each dialogue turn as they have been assigned by the system. These labels coincide with ICSI labels (BKC, ACK, FGB, FHD, RHQ, - that is Floor Grabber, Floor Holder, Backchannel, Acknowledge, RhetoricQuestion - with the addition of NEGation, ASSent, MTVation, PRPosal, GRTeeing, CNLusion;
2. It recovers Overlaps as they have been marked during the analysis;
3. It produces an Opinion label which we call Polarity, 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, where Hot Spots is simply a set of turns in sequence where the interlocutors overlap each other

frequently. Episodes on the contrary are a set of turns in which a single speaker "arguments" his/her topics which may occasionally be interrupted by overlaps or by short continuers, backchannel or other similar phenomena by other speakers without however grabbing the floor;
5. Then the main predicate that assigns argumentative labels is called by a recursive routine:

  i. at first it tries exceptions – which are strongly pragmatically marked - 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.;

  ii. then Short utterances are checked. In case they end up with a question mark they are labeled as Questions. Else, the Dialogue Act label is considered. Negations are also computed here;

  iii. now the main call is activated. In order to start matching the rules, the semantic information is recovered for the current turn, in a recursive clause by clause manner;

  iv. when semantic information has been recovered the rules are fired. There are some 33 rules which take as input the following vector of features:

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

where Output is the output label chosen by the rule; DiscDom may be Factive or NonFactive, Suggestion or Proposal; Relevance may be foreground or background; DomPointView may be objective or subjective. Rules are applied by matching input labels in a Finite State Automaton manner. However sometimes conditions and constraints are made to apply. For instance, ***analyzecontext(NoCl)***, checks to verify whether the current speaker holds the floor in the 2 preceding or following clauses.

  v. the rules produce a set of argumentative labels, one for each clause. The system then chooses the label to associate to 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.

Here below we report a portion of the General Summary extracted from Dialogue 1. Eventually we are able to evaluate the degree of collaboration vs. competitiveness of each participant in the conversation and make a general statement like this one produced automatically by means of canned sentences:

GENERAL INFORMATION ON PARTICIPANTS
The participants to the meeting are 6.
Participants less actively involved are Adam and Andreas who only intervened respectively for 9 and 78 turns.

LEVEL OF INTERACTIVITY IN THE DISCUSSION
The speaker that has held the majority of turns is Don with a total of 549 turns, followed by
Morgan with a total of 512 turns, followed by
Jane with a total of 292.
The speaker that has undergone the majority of overlaps is Morgan followed by Don.
The speaker that has done the majority of overlaps is Morgan followed by Jane.
Morgan is the participant that has been most competitive.
Andreas only intervened after turn no. 1091.

DISCUSSION TOPICS
The main topics have been introduced by the second most important speaker of the meeting, Jane.
The most frequent entities in the whole dialogue partly coincide with the best topics, and are the following, in decreasing order:
level, format, stuff, tag, utterance, guess, frame, file, type, representation, phone, annotations, sentence, information, x_m_l, point, p_file, start, segment, equals, 'ATLAS', prosodic, mean, link, database, data, change, structure, diff, a_p_i_, tool, sort, sequence, program, pitch, external, end, channel, boundaries, work, versions, translate, timeline, text, speaker, overlap, file_format, value, transcripts, store, prosody, phrase, perl, need, lattice, idea, feature, useful, turn, structured, separate, segmentation, search, output, node, meeting, library, language, input, help, handle, example, codes, bunch, alignment, NIST, ICSI.

ARGUMENTATIVE CONTENT
The following participants Andreas, Dave, Don, Jane, Morgan expressed their dissent 44 times.
However Andreas, Dave and Morgan expressed dissent in a consistently smaller percentage.
The following participants Adam, Andreas, Dave, Don, Jane, Morgan asked questions 53 times.
The remaining 1239 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 320 times.

The system has been used to parse the first 10 dialogues of the ICSI corpus for a total number of 98523 words and 13803 turns. This has been done to "train" the system: what happened was that, for the first 5 dialogues, we had to take care of failures. We also had to tune all the modules and procedures carefully. In particular, the module for argumentative automatic classification was incrementally improved in order to cover all conventional ways to express Agreement. For

this reason, we then chose two random additional dialogues to test this second task.

4.3 Experimental Results

We had one skilled linguist to provide a turn level annotation for argumentative labels: we don't have any agreement measure in this case, even though we expect the annotation to be in line with current experiments on the same subject (Pallotta et al. 2007). In the following table we report data related to the experiment of automatic annotation of argumentative categories. On a total of 2304 turns, 2251 have received an argumentative automatic classification, with a Recall of 97.53%. As can be gathered from the following table 2., the F-score is fairly high compared to current results reported in the literature on the same topic which are all below 80%.

	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

Table 2. Overall count of argumentative labels

We computed Precision as the ratio between Correct Argumentative Labels/Found Argumentative Labels, which corresponds to 81.26%. The F-score is 88.65%.

5. The Anaphora Resolution Module

The problem represented by pronominal expressions in dialogues needs to be addressed fully and not by means of ad hoc solutions. This requires a full-fledged system for anaphora resolution. One such system is shown in Fig. 1 below, where we highlight the architecture and main processes undergoing at the anaphora level. First of all, the subdivision of the system into two levels: Clause level – intrasentential pronominal phenomena – where all pronominal expressions contained in modifiers, adjuncts or complement clauses receive their antecedent locally. Possessive pronouns, pronouns contained in relative clauses and complement clauses choose preferentially their antecedents from list of higher level referring expressions. Not so for those pronouns contained in matrix clauses. In particular the ones in subject position are to be coreferred in the discourse. This requires the system to be equipped with a History List of all referring expressions to be used when needed. In the system, three levels are indicated:

Clause level, i.e. simple sentences; Utterance level, i.e. complex sentences; Discourse level, i.e. intersententially.

Our system computes semantic structures in a sentence by sentence fashion and any information useful to carry out anaphoric processes is made available to the following stretch of dialogue.
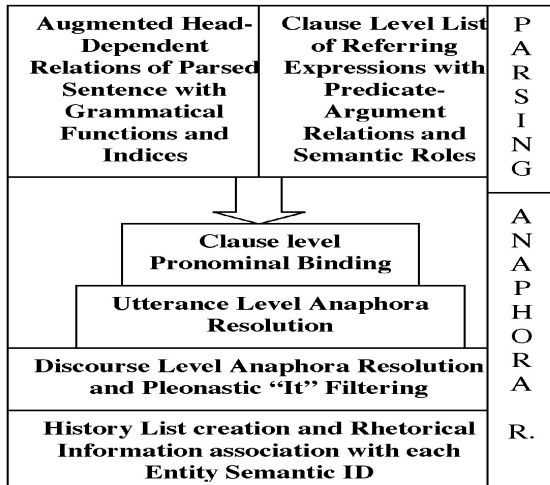
Augmented Head-Dependent Relations of Parsed Sentence with Grammatical Functions and Indices	Clause Level List of Referring Expressions with Predicate-Argument Relations and Semantic Roles	P A R S I N G
Clause level Pronominal Binding		A N A P H O R A R.
Utterance Level Anaphora Resolution		
Discourse Level Anaphora Resolution and Pleonastic "It" Filtering		
History List creation and Rhetorical Information association with each Entity Semantic ID		

Figure 1. Anaphoric Processes in GETARUNS

5.1 The Experiments

We set up a number of experiments in order to test the new version of the system. However we will concentrate only on one of them, that is detecting referential from nonreferential uses of personal pronouns YOU, WE and the pronoun IT.

Here below is a table containing total values for pronouns WE/YOU/IT in all the 10 dialogues analysed.

	Referent	Generic	Total	Found
WE	1186	706	1892	1356
YOU	1045	742	1787	1132
IT	1593	1008	2601	1627
Total	3824	2456	6280	4115

Table 3. Overall count of pronominal expressions

We had two skilled linguists to annotate pronominal WE/IT/YOU properties as either referential/nonreferential. Their agreement on this task was very high with a kappa-score of 0.71. Results for the experiment are as follows,

	Recall	Precision	F-Score
WE	71.67%	81.2%	76.14%
YOU	63.34%	89.3%	74.11%
IT	62.52%	84.6%	72.19%

Table 4. Results for pronominal expressions

6. Conclusions and Future Work

We have presented work carried out to extend and adapt a system for text understanding in order to make it fit for dialogue understanding. We proposed a set of expansions to cope with typical dialogue related problems, such as presence of non-sentential fragments, elliptical fragments interpretable as speech acts, massive presence of generic non-referential pronominal expressions, etc. We implemented a number of additional components: an algorithm that takes care of overlaps and uses that information to split current utterances and temporally realign the conversational flow. A module that computes Argumentative automatic classification labels out of a small set, on top of discourse relations and other semantic markers determined by the semantic component of the system.

The system has been evaluated for two of its most important components, the newly implemented pronominal binding module and the argumentative classification module. Results are very encouraging. However, we note that in that task, labels which may cause great uncertainty and are highly ambiguous, have been lumped together to facilitate the classification task.

Of course we intend to complete the analysis of all dialogues contained in the ICSI corpus and refine our algorithms. Then we would like to use the system with a totally different scenario, as for instance the Switchboard two parties dialogues and see whether the "training" carried out on the basis of multiparty dialogues may be fruitfully applied to such reduced conversational framework. In particular we still need to work at the level of DECISION labeling, which is something that we intend to do at Episode level. We also need 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.

7. References

Allen, J., M. Dzikovska, M. Manshadi, and M. Swift, (2007), Deep linguistic processing for spoken dialogue systems. In ACL 2007 Workshop on Deep Linguistic Processing, pp. 49–56.

Armstrong, S. et al., (2003), Natural language queries on natural language data: a database of meeting dialogues. In Proceedings of NLDB'2003 conference, Burg/Cottbus, Germany.

Bergsma, Shane , Dekang Lin and Randy Goebel, (2008), Distributional Identification of Non-Referential Pronouns, In ACL-HLT 2008, Columbus, Ohio, June 16-18, 2008, pp. 10-18.

Bresnan, J., (2000), Lexical-Functional Syntax, Blackwell.

Bunt, H. (1979), Conversational principles in question-answer dialogues. In D. Krallmann, editor, Zur Theory der Frage, pp. 119–141. Narr Verlag, Tübingen.

Clark A., and Andrei Popescu-Belis, (2004), Multi-level Dialogue Act Tags. In Proceedings of SIGDIAL'04, pp.163–170. Cambridge, MA, USA.

Delmonte, R. A. Bristot, M. A. Piccolino Boniforti and S. Tonelli, (2006), Another Evaluation of Anaphora Resolution Algorithms and a Comparison with GETARUNS' Knowledge Rich Approach, ROMAND 2006, 11th EACL, Trento, Association for Computational Linguistics, 3-10.

Delmonte R., (2007), *Computational Linguistic Text Processing – Logical Form, Semantic Interpretation, Discourse Relations and Question Answering*, Nova Science Publishers, New York.

Delmonte R., (2008), Semantic and Pragmatic Computing with GETARUNS, in Bos & Delmonte (eds.), STEP, College Pub. London, pp. 287-298.

Fellbaum, Christiane, (ed.), (1998), *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge MA.

Galley, Michel , Kathleen McKeown, Eric Fosler-Lussier and Hongyan Jing, (2003), Discourse Segmentation of Multi-Party Conversation. In Proceedings of ACL 2003, pp. 562–569, Sapporo, Japan.

Goffman E., (1981), Forms of Talk. Philadelphia: University of Pennsylvania Press.

Gupta, Surabhi, Matthew Purver and Dan Jurafsky. (2007), Disambiguating Between Generic and Referential "You" in Dialog. Proceedings of ACL 2007 short papers, Prague, Czech Republic.

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.

Javanovich N. and R. op den Akker, (2004), Towards Adressee Identification in Multi-party dialogues, in *Proceedings of the 5th Sigdial Workshop on Discourse and Dialogue*, ACL, Pennsylvania, 2004, pp. 89-92.

Kunz W. and Rittel H. W. J., (1970), Issues as elements of information systems. Technical Report WP-131, Berkeley: University of California.

Lo Cascio V., (1991), Grammatica dell'Argomentare: strategie e strutture. Firenze: La Nuova Italia.

Mann, W.C and S.A Thompson, (1988), Rhetorical Structure Theory: Towards a Functional Theory Text Organization. Text, 8(3):243–281.

Müller, Christoph. (2007), Resolving *It*, *This*, and *That* in Unrestricted Multi-Party Dialog. In: Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics, Prague, Czech Republic, pp. 816-823.

Müller, Christoph, (2006), Automatic Detection on Nonreferential *It* In Spoken Multi-Party Dialog. In: Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics, Trento, Italy, pp. 49-56.

Niekrasz J., Purver M., Dowding J. and Peters S., (2005), Ontology-Based Discourse Understanding for a Persistent Meeting Assistant. In: Proceedings of the AAAI Spring Symposium Persistent Assistants: Living and Working with AI. Stanford.

Pallotta, Vincenzo, Hatem Ghorbel, Afzal Ballim, Agnes Lisowska and Stéphane Marchand-Maillet, (2004), Towards meeting information systems: Meeting knowledge management. In Proceedings of ICEIS 2005, pp. 464–469, Porto, Portugal.

Rosemberg Duska and John A.A. Silince, (1999), Common ground in computer-supported collaborative argumentation. In Proceedings of the CLSCL99, Stanford, CA, USA.

Schegloff E. and Sacks H., (1973), Opening up closings. Semiotica 8: 289-327.

Strube, M. and C. Müller, (2003), A Machine Learning Approach to Pronoun Resolution in Spoken Dialogue, in *Proceedings of the 41st Annual Meeting of the ACL,* Sapporo, Japan, pp. 168–175.