Annotating Spoken Dialogs: from Speech Segments to Dialog Acts and Frame Semantics

Marco Dinarelli, Silvia Quarteroni, Sara Tonelli, Alessandro Moschitti, Giuseppe Riccardi*
University of Trento
38050 Povo - Trento, Italy
{dinarelli,silviaq,moschitti,riccardi}@disi.unitn.it, satonelli@fbk.eu

Abstract

We are interested in extracting semantic structures from spoken utterances generated within conversational systems. Current Spoken Language Understanding systems rely either on hand-written semantic grammars or on flat attribute-value sequence labeling. While the former approach is known to be limited in coverage and robustness, the latter lacks detailed relations amongst attribute-value pairs. In this paper, we describe and analyze the human annotation process of rich semantic structures in order to train semantic statistical parsers. We have annotated spoken conversations from both a human-machine and a human-human spoken dialog corpus. Given a sentence of the transcribed corpora, domain concepts and other linguistic features are annotated, ranging from e.g. part-of-speech tagging and constituent chunking, to more advanced annotations, such as syntactic, dialog act and predicate argument structure. In particular, the two latter annotation layers appear to be promising for the design of complex dialog systems. Statistics and mutual information estimates amongst such features are reported and compared across corpora.

1 Introduction

Spoken language understanding (SLU) addresses the problem of extracting and annotating the meaning structure from spoken utterances in the context of human dialogs (De Mori et al., 2008). In spoken dialog systems (SDS) most used models of SLU are based on the identification of slots (entities) within one or more frames (frame-slot semantics) that is defined by the application. While this model is simple and clearly insufficient to cope with interpretation and reasoning, it has supported the first generation of spoken dialog systems. Such dialog systems are thus limited by the ability to parse semantic features such as predicates and to perform logical computation in the context of a specific dialog act (Bechet et al., 2004). This limitation is reflected in the type of human-machine interactions which are mostly directed at querying the user for specific slots (e.g. "What is the departure city?") or implementing simple dialog acts (e.g. confirmation). We believe that an important step in overcoming such limitation relies on the study of models of human-human dialogs at different levels of representation: lexical, syntactic, semantic and discourse.

In this paper, we present our results in addressing the above issues in the context of the LUNA research project for next-generation spoken dialog interfaces (De Mori et al., 2008). We propose models for different levels of annotation of the LUNA spoken dialog corpus, including attribute-value, predicate argument structures and dialog acts. We describe the tools and the adaptation of off-the-shelf resources to carry out annotation of the predicate argument structures (PAS) of spoken utterances. We present a quantitative analysis of such semantic structures for both human-machine and human-human conversations.

To the best of our knowledge this is the first (human-machine and human-human) SDS corpus denoting a multilayer approach to the annotation of lexical, semantic and dialog features, which allows us to investigate statistical relations between the layers such as shallow semantic and discourse features used by humans or machines. In the following sections we describe the corpus, as well as a quantitative analysis and statistical correlations between annotation layers.

*This work was partially funded by the European Commission projects LUNA (contract 33549) and ADAMACH (contract 022593).

Proceedings of EACL 2009 Workshop on Semantic Representation of Spoken Language - SRSL 2009, pages 34–41,
Athens, Greece, 30 March 2009.

2 Annotation model

Our corpus is planned to contain 1000 equally partitioned Human-Human (HH) and Human-Machine (HM) dialogs. These are recorded by the customer care and technical support center of an Italian company. While HH dialogs refer to real conversations of users engaged in a problem solving task in the domain of software/hardware troubleshooting, HM dialogs are acquired with a Wizard of Oz approach (WOZ). The human agent (wizard) reacts to user's spontaneous spoken requests following one of ten possible dialog scenarios inspired by the services provided by the company.

The above data is organized in transcriptions and annotations of speech based on a new multi-level protocol studied specifically within the project, i.e. the annotation levels of words, turns[1], attribute-value pairs, dialog acts, predicate argument structures. The annotation at word level is made with part-of-speech and morphosyntactic information following the recommendations of EAGLES corpora annotation (Leech and Wilson, 2006). The attribute-value annotation uses a predefined domain ontology to specify concepts and their relations. Dialog acts are used to annotate intention in an utterance and can be useful to find relations between different utterances as the next section will show. For predicate structure annotation, we followed the FrameNet model (Baker et al., 1998) (see Section 2.2).

2.1 Dialog Act annotation

Dialog act annotation is the task of identifying the function or goal of a given utterance (Sinclair and Coulthard, 1975): thus, it provides a complementary information to the identification of domain concepts in the utterance, and a domain-independent dialog act scheme can be applied. For our corpus, we used a dialog act taxonomy which follows initiatives such as DAMSL (Core and Allen, 1997), TRAINS (Traum, 1996) and DIT++ (Bunt, 2005). Although the level of granularity and coverage varies across such taxonomies, a careful analysis leads to identifying three main groups of dialog acts:

1. *Core* acts, which represent the fundamental actions performed in the dialog, e.g. requesting and providing information, or executing a task. These include initiatives (often called forward-looking acts) and responses (backward-looking acts);
2. *Conventional/Discourse management* acts, which maintain dialog cohesion and delimit specific phases, such as opening, continuation, closing, and apologizing;
3. *Feedback/Grounding* acts,used to elicit and provide feedback in order to establish or restore a common ground in the conversation.

[1]A turn is defined as the interval when a speaker is active, between two pauses in his/her speech flow.

Our taxonomy, following the same three-fold partition, is summarized in Table 1.

Table 1: Dialog act taxonomy

Core dialog acts	
Info-request	Speaker wants information from addressee
Action-request	Speaker wants addressee to perform an action
Yes-answer	Affirmative answer
No-answer	Negative answer
Answer	Other kinds of answer
Offer	Speaker offers or commits to perform an action
ReportOnAction	Speaker notifies an action is being/has been performed
Inform	Speaker provides addressee with information not explicitly required (via an Info-request)
Conventional dialog acts	
Greet	Conversation opening
Quit	Conversation closing
Apology	Apology
Thank	Thanking (and down-playing)
Feedback/turn management dialog acts	
Clarif-request	Speaker asks addressee for confirmation/repetition of previous utterance for clarification.
Ack	Speaker expresses agreement with previous utterance, or provides feedback to signal understanding of what the addressee said
Filler	Utterance whose main goal is to manage conversational time (i.e. dpeaker taking time while keeping the turn)
Non-interpretable/non-classifiable dialog acts	
Other	Default tag for non-interpretable and non-classifiable utterances

It can be noted that we have decided to retain only the most frequent dialog act types from the schemes that inspired our work. Rather than aspiring to the full discriminative power of possible conversational situations, we have opted for a simple taxonomy that would cover the vast majority of utterances and at the same time would be able to generalize them. Its small number of classes is meant to allow a supervised classification method to achieve reasonable performance with limited data. The taxonomy is currently used by the statistical Dialogue Manager in the ADAMACH EU project (Varges et al., 2008); the limited number of classes allows to reduce the number of hypothesized current dialogue acts, thus reducing the dialogue state space.

Dialog act annotation was performed manually by a linguist on speech transcriptions previously segmented into turns as mentioned above. The annotation unit for dialog acts, is the utterance; however, utterances are complex semantic entities that do not necessarily correspond to turns. Hence, a segmentation of the dialog transcription into utterances was performed by the annotator before dialog act labeling. Both utterance segmentation and dialog act labeling were performed through the MMAX tool (Müller and Strube, 2003).

The annotator proceeded according to the following guidelines:

1. by default, a turn is also an utterance;
2. if more than one tag is applicable to an utterance, choose the tag corresponding to its main function;
3. in case of doubt among several tags, give priority to tags in *core* dialog acts group;
4. when needed, split the turn into several utterances or merge several turns into one utterance.

Utterance segmentation provides the basis not only for dialog act labeling but also for the other semantic annotations. See Fig. 1 for a dialog sample where each line represents an utterance annotated according to the three levels.

PERSON-NAME
Info: Buongiorno, sono Paola.
GREETING *B._NAMED* *Name*
Good morning, this is Paola.

Info-req: Come la posso aiutare?
Benefitted_party *ASSISTANCE*
How may I help you?

CONCEPT *HARDWARE-COMPONENT*
Info: Buongiorno. Ho un problema con la stampante.
GREETING *PR._DESCRIPTION* *Affected_device*
Good morning. I have a problem with the printer.

PART-OF-DAY *NEGAT. ACTION* *ACTION*
Info: Da stamattina non riesco più a stampare
Problem
Since this morning I can't print.

Info-req: Mi può dire nome e cognome per favore?
Addressee *TELLING* *Message*
Can you tell me your name and surname, please?

PERSON-NAME *PERSON-SURNAME*
Answer: Mi chiamo Alessandro Manzoni.
Entity *B._NAMED* *Name*
My name is Alessandro Manzoni.

Figure 1: Annotated dialog extract. Each utterance is preceded by dialog act annotation. Attribute-value annotation appears above the text, PAS annotation below the text.

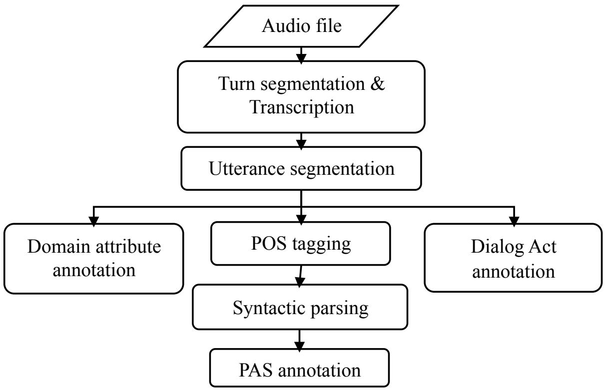
2.2 Predicate Argument annotation

We carried out predicate argument structure annotation applying the FrameNet paradigm as described in (Baker et al., 1998). This model comprises a set of prototypical situations called *frames*, the frame-evoking words or expressions called *lexical units* and the roles or participants involved in these situations, called *frame elements*. The latter are typically the syntactic dependents of the lexical units. All lexical units belonging to the same frame have similar semantics and show the same valence. A particular feature of the FrameNet project both for English and for other languages is its corpus-based nature, i.e. every element described in the resource has to be instantiated in a corpus. To annotate our SDS corpus, we adopted where possible the already existing frame and frame element descriptions defined for the English FrameNet project, and introduced new definitions only in case of missing elements in the original model.

Figure 1 shows a dialog sample with PAS annotation reported below the utterance. All lexical units are underlined and the frame is written in capitals, while the other labels refer to frame elements. In particular, *ASSISTANCE* is evoked by the lexical unit *aiutare* and has one attested frame element (*Benefitted_party*), *GREETING* has no frame element, and *PROBLEM_DESCRIPTION* and *TELLING* have two frame elements each.

Figure 2 gives a comprehensive view of the annotation process, from audio file transcription to the annotation of three semantic layers. Whereas

Figure 2: The annotation process



attribute-value and DA annotation are carried out on the segmented dialogs at utterance level, PAS annotation requires POS-tagging and syntactic parsing (via Bikel's parser trained for Italian (Corazza et al., 2007)). Finally, a shallow manual correction is carried out to make sure that the tree nodes that may carry semantic information have correct constituent boundaries. For the annotation of frame information, we used the *Salto* tool (Burchardt et al., 2006), that stores the dialog file in TIGER-XML format and allows to easily introduce word tags and frame flags. Frame information is recorded on top of parse trees, with target information pointing to terminal words and frame elements pointing to tree nodes.

3 Quantitative comparison of the Annotation

We evaluated the outcome of dialog act and PAS annotation levels on both the human-human (henceforth HH) and human-machine (HM) corpora by not only analyzing frequencies and occurrences in the separate levels, but also their interaction, as discussed in the following sections.

3.1 Dialog Act annotation

Analyzing the annotation of 50 HM and 50 HH dialogs at the dialog act level, we note that an HH dialog is composed in average by 48.9±17.4 (standard deviation) dialog acts, whereas a HM dialog is composed of 18.9±4.4. The difference between average lengths shows how HH spontaneous speech can be redundant, while HM dialogs are more limited to an exchange of essential information. The standard deviation of a conversation in terms of dialog acts is considerably higher in the HH corpus than in the HM one. This can be explained by the fact that the WOZ follows a unique, previously defined task-solving strategy that does not allow for digressions. Utterance segmentation was also performed differently on the two corpora. In HH we performed 167 turn mergings and 225 turn splittings; in HM dialogs, only turn splittings (158) but no turn mergings were performed.

Tables 2 and 3 report the dialog acts occurring in the HM and HH corpora, respectively, ranked by their frequencies.

Table 2: Dialog acts ranked by frequency in the human-machine (HM) corpus

human-machine (HM)		
DA	count	rel. freq.
Info-request	249	26.3%
Answer	171	18.1%
Inform	163	17.2%
Yes-answer	70	7.4%
Quit	60	6.3%
Thank	56	5.9%
Greet	50	5.3%
Offer	49	5.2%
Clarification-request	26	2.7%
Action-request	25	2.6%
Ack	12	1.3%
Filler	6	0.6%
No-answer	5	0.5%
Other, ReportOnAction	2	0.2%
Apology	1	0.1%
TOTAL	947	

From a comparative analysis, we note that:

1. *info-request* is by far the most common dialog act in HM, whereas in HH *ack* and *info* share the top ranking position;

2. the most frequently occurring dialog act in HH, i.e. *ack*, is only ranked 11th in HM;

3. the relative frequency of *clarification-request* (4,7%) is considerably higher in HH than in HM.

We also analyzed the ranking of the most frequent dialog act bigrams in the two corpora. We can summarize our comparative analysis, reported in Table 4, to the following: in both corpora, most bigram types contain *info* and *info-request*,

Table 3: Dialog acts ranked by frequency in the human-human (HH) corpus

	human-human (HH)	
DA	count	rel. freq.
Ack	582	23.8%
Inform	562	23.0%
Info-request	303	12.4%
Answer	192	7.8%
Clarification-request	116	4.7%
Offer	114	4.7%
Yes-answer	112	4.6%
Quit	101	4.1%
ReportOnAction	91	3.7%
Other	70	2.9%
Action-request	69	2.8%
Filler	61	2.5%
Thank	33	1.3%
No-answer	26	1.1%
Greet, Apology	7	0.3%
TOTAL	2446	

as expected in a troubleshooting system. However, the bigram *info-request answer*, which we expected to form the core of a task-solving dialog, is only ranked 5th in the HH corpus, while 5 out of the top 10 bigram types contain *ack*. We believe that this is because HH dialogs primarily contain spontaneous information-providing turns (e.g. several *info info* by the same speaker) and acknowledgements for the purpose of backchannel. Instead, HM dialogs, structured as sequences of *info-request answers* pairs, are more minimal and brittle, showing how users tend to avoid redundancy when addressing a machine.

Table 4: The 10 most frequent dialog act bigrams

human-machine (HM)	human-human (HH)
info-req answer	ack info
answer info-req	info ack
info info-req	info info
info-req y-answer	ack ack
sentence_beginning greet	info-req answer
greet info	info info-req
info quit	info-req y-answer
offer info	ack info-req
thank info	answer ack
y-answer thank	quit *sentence_end*

3.2 Predicate Argument annotation

We annotated 50 HM and 50 HH dialogs with frame information. Differently from the English FrameNet database, we didn't annotate one frame per sentence. On the contrary, we identified all lexical units corresponding to "semantically relevant" verbs, nouns and adjectives with a syntactic subcategorization pattern, eventually skipping the utterances with empty semantics (e.g. disfluencies). In particular, we annotated all lexical units that imply an action, introduce the speaker's opinion or describe the office environment. We introduced 20 new frames out of the 174 identified in the corpus because the original definition of frames related to hardware/software, data-handling and customer assistance was sometimes too coarse-grained. Few new frame elements were introduced as well, mostly expressing syntactic realizations that are typical of spoken Italian.

Table 5 shows some statistics about the corpus dimension and the results of our annotation. The human-human dialogs contain less frame instances in average than the human-machine group, meaning that speech disfluencies, not present in turns uttered by the WOZ, negatively affect the semantic density of a turn. For the same reason, the percentage of turns in HH dialogs that were manually corrected in the pre-processing step (see Section 2.2) is lower than for HM turns, since HH dialogs have more turns that are semantically empty and that were skipped in the correction phase. Besides, HH dialogs show a higher frame variability than HM, which can be explained by the fact that spontaneous conversation may concern minor topics, whereas HM dialogs follow a previously defined structure, designed to solve software/hardware problems.

Tables 6 and 7 report the 10 most frequent frames occurring in the human-machine resp. human-human dialogs. The relative frame frequency in HH dialogs is more sparse than in HM dialogs, meaning that the task-solving strategy followed by the WOZ limits the number of digressions, whereas the semantics of HH dialogs is richer and more variable.

As mentioned above, we had to introduce and define new frames which were not present in the original FrameNet database for English in order to capture all relevant situations described in the dialogs. A number of these frames appear in both tables, suggesting that the latter are indeed rel-

Table 5: Dialog turn and frame statistics for the human-machine (HM) resp. human-human (HH) corpus

	HM	HH
Total number of turns	662	1,997
Mean dialog length (turns)	13.2	39.9
Mean turn length (tokens)	11.4	10.8
Corrected turns (%)	50	39
Total number of annotations	923	1951
Mean number of frame annotations per dialog	18.5	39.0
Mean number of frame elements per frame annotation	1.6	1.7

Table 6: The 10 most frequent frames in the HM corpus (* =newly introduced)

HM corpus		
Frame	count	freq-%
Greeting*	146	15.8
Telling	134	14.5
Recording	83	8.9
Being_named	74	8.0
Contacting	52	5.6
Usefulness	50	5.4
Being_operational	28	3.0
Problem_description*	24	2.6
Inspecting	24	2.6
Perception_experience	21	2.3

Table 7: The 10 most frequent frames in the HH corpus (* =newly introduced)

HH corpus		
Frame	count	freq-%
Telling	143	7.3
Greeting*	124	6.3
Awareness	74	3.8
Contacting	63	3.2
Giving	62	3.2
Navigation*	61	3.1
Change_operational_state	51	2.6
Perception_experience	46	2.3
Insert_data*	46	2.3
Come_to_sight*	38	1.9

evant to model the general semantics of the dialogs we are approaching. The most frequent frame group comprises frames relating to information exchange that is typical of the help-desk activity, including *Telling*, *Greeting*, *Contacting*, *Statement*, *Recording*, *Communication*. Another relevant group encompasses frames related to the operational state of a device, for example *Being_operational*, *Change_operational_state*, *Operational_testing*, *Being_in_operation*.

The two groups also show high variability of lexical units. *Telling*, *Change_operational_state* and *Greeting* have the richest lexical unit set, with 11 verbs/nouns/adjectives each. *Arriving* and *Awareness* are expressed by 10 different lexical units, while *Statement*, *Being_operational*, *Removing* and *Undergo_change_of_operational_state* have 9 different lexical units each. The informal nature of the spoken dialogs influences the composition of the lexical unit sets. In fact, they are rich in verbs and multiwords used only in colloquial contexts, for which there are generally few attestations in the English FrameNet database.

Similarly to the dialog act statistics, we also analyzed the most frequent frame bigrams and trigrams in HM and HH dialogs. Results are reported in Tables 8 and 9. Both HH bigrams and trigrams show a more sparse distribution and lower relative frequency than HM ones, implying that HH dialogs follow a more flexible structure with a richer set of topics, thus the sequence of themes is less predictable. In particular, 79% of HH bigrams and 97% of HH trigrams occur only once (vs. 68% HM bigrams and 82% HM trigrams). On the contrary, HM dialogs deal with a fix sequence of topics driven by the turns uttered by the WOZ. For instance, the most frequent HM bigram and trigram both correspond to the opening utterance of the WOZ:

Help desk buongiorno$_{\text{GREETING}}$, *sono*$_{\text{BEING\_NAMED}}$ *Paola, in cosa posso esserti utile*$_{\text{USEFULNESS}}$*?*

(Good morning, help-desk service, Paola speaking, how can I help you?)

3.3 Mutual information between PAS and dialog acts

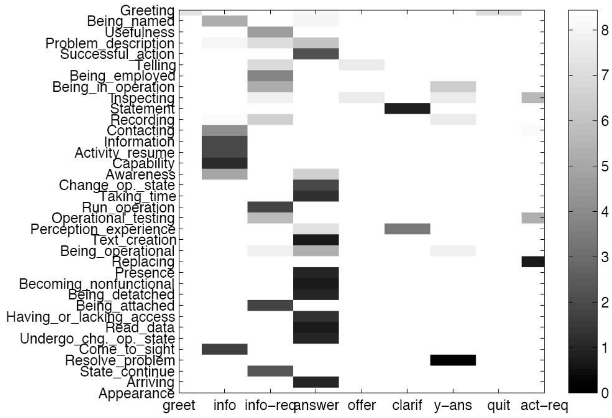
A unique feature of our corpus is the availability of both a semantic and a dialog act annotation level: it is intuitive to seek relationships in the purpose of improving the recognition and understanding of each level by using features from the other. We considered a subset of 20 HH and 50 HM dialogs and computed an initial analysis

Table 8: The 5 most frequent frame bigrams

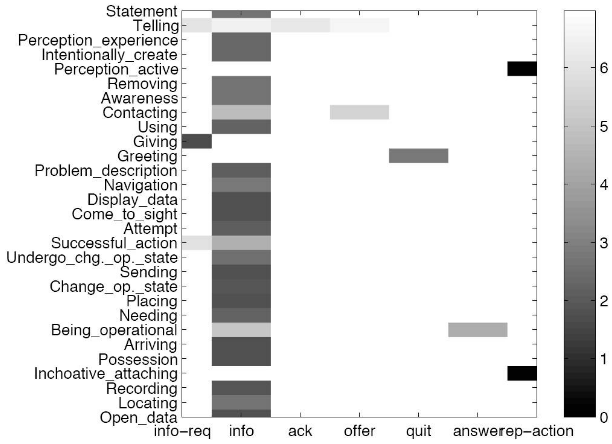
human-machine (HM)	freq-%
Greeting Being_named	17.1
Being_named Usefulness	15.3
Telling Recording	12.9
Recording Contacting	10.9
Contacting Greeting	10.6
human-human (HH)	**freq-%**
Greeting Greeting	4.7
Navigation Navigation	1.2
Telling Telling	1.0
Change_op._state Change_op._state	0.9
Telling Problem_description	0.8

Table 9: The 5 most frequent frame trigrams

human-machine (HM)	freq-%
Greeting Being_named Usefulness	9.5
Recording Contacting Greeting	5.7
Being_named Usefulness Greeting	3.7
Telling Recording Contacting	3.5
Telling Recording Recording	2.2
human-human (HH)	**freq-%**
Greeting Greeting Greeting	1.6
Greeting Being_named Greeting	0.5
Contacting Greeting Greeting	0.3
Navigation Navigation Navigation	0.2
Working_on Greeting Greeting	0.2



(a) human-machine dialogs (filtering co-occurrences below 3)



(b) human-human dialogs (filtering co-occurrences below 5)

Figure 3: Weighted conditional entropy between PAS and dialog acts in the HM (a) and HH corpus (b). To lower entropies correspond higher values of mutual information (darker color in the scale)

of the co-occurrences of dialog acts and PAS. We noted that each PAS tended to co-occur only with a limited subset of the available dialog act tags, and moreover in most cases the co-occurrence happened with only one dialog act. For a more thorough analysis, we computed the weighted conditional entropy between PAS and dialog acts, which yields a direct estimate of the mutual information between the two levels of annotation[2].

Our results are illustrated in Figure 3. In the HM corpus (Fig. 3(a)), we noted some interesting associations between dialog acts and PAS. First, `info-req` has the maximal MI with PAS like *Being_in_operation* and *Being_attached*, as requests are typically used by the operator to get information about the status of device. Several PAS denote a high MI with the `info` dialog act, including *Activity_resume, Information, Being_named, Contacting*, and *Resolve_problem*. *Contacting* refers to the description of the situation and of the speaker's point of view (usually the caller). *Being_named* is primarily employed when the caller introduces himself, while *Activity_resume* usually refers to the operator's description of the sched-

[2]Let $H(y_j|x_i)$ be the weighted conditional entropy of observation y_j of variable Y given observation x_i of variable X:

$$H(y_j|x_i) = -p(x_i; y_j) log \frac{p(x_i; y_j)}{p(x_i)},$$

where $p(x_i; y_j)$ is the probability of co-occurrence of x_i and y_j, and $p(x_i)$ and $p(y_j)$ are the marginal probabilities of occurrence of x_i resp. y_j in the corpus. There is an obvious relation with the weighted mutual information between x_i and y_j, defined following e.g. (Bechet et al., 2004) as:

$$wMI(x_i; y_j) = p(x_i; y_j) log \frac{p(x_i; y_j)}{p(x_i)p(y_j)}.$$

Indeed, the higher is $H(y_j|x_i)$, the lower is $wMI(x_i; y_j)$. We approximate all probabilities using frequency of occurrence.

uled interventions.

As for the remaining acts, `clarif` has the highest MI with *Perception_experience* and *Statement*, used to warn the addressee about understanding problems and asking him to repeat/rephrase an utterance, respectively. The two strategies can be combined in the same utterance, as in the utterance: *Non ho sentito bene: per favore ripeti cercando di parlare più forte.* (I haven't quite heard that, please repeat trying to speak up.).

The `answer` tag is highly informative with *Successful_action, Change_operational_state, Becoming_nonfunctional, Being_detached, Read_data*. These PAS refer to the exchange of information (*Read_data*) or to actions performed by the user after a suggestion of the system (*Change_operational_state*). Action requests (act-req) seem to be correlated to *Replacing* as it usually occurs when the operator requests the caller to carry out an action to solve a problem, typically to replace a component with another. Another frequent request may refer to some device that the operator has to test.

In the HH corpus (Fig. 3(b)), most of the PAS are highly mutually informative with `info`: indeed, as shown in Table 3, this is the most frequently occurring act in HH except for `ack`, which rarely contain verbs that can be annotated by a frame. As for the remaining acts, there is an easily explainable high MI between `quit` and *Greeting*; moreover, `info-req` denote its highest MI with *Giving*, as in requests to give information, while `rep-action` denotes a strong co-occurrence with *Inchoative_attaching*: indeed, interlocutors often report on the action of connecting a device.

These results corroborate our initial observation that for most PAS, the mutual information tends to be very high in correspondence of one dialog act type: this suggests the beneficial effect of including shallow semantic information as features for dialog act classification. The converse is less clear as the same dialog act can relate to a span of words covered by multiple PAS and generally, several PAS co-occur with the same dialog act.

4 Conclusions

In this paper we have proposed an approach to the annotation of spoken dialogs using semantic and discourse features. Such effort is crucial to investigate the complex dependencies between the layers of semantic processing. We have designed the annotation model to incorporate features and models developed both in the speech and language research community and bridging the gap between the two communities. Our multilayer annotation corpus allows the investigation of cross-layer dependencies and across human-machine and human-human dialogs as well as training of semantic models which accounts for predicate interpretation.

References

C. F. Baker, C. J. Fillmore, and J. B. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of ACL/Coling'98*, pages 86–90.

F. Bechet, G. Riccardi, and D. Hakkani-Tur. 2004. Mining spoken dialogue corpora for system evaluation and modeling. In *Proceedings of EMNLP'04*, pages 134–141.

H. Bunt. 2005. A framework for dialogue act specication. In *Proceedings of SIGSEM WG on Representation of Multimodal Semantic Information*.

A. Burchardt, K. Erk, A. Frank, A. Kowalski, S. Padó, and M. Pinkal. 2006. Salto - a versatile multilevel annotation tool. In *Proceedings of LREC 2006*, pages 517–520, Genoa, Italy.

A. Corazza, A. Lavelli, and G. Satta. 2007. Analisi sintattica-statistica basata su costituenti. *Intelligenza Artificiale*, 4(2):38–39.

M. G. Core and J. F. Allen. 1997. Coding dialogs with the DAMSL annotation scheme. In *Proceedings of the AAAI Fall Symposium on Communicative Actions in Humans and Machines*.

R. De Mori, F. Bechet, D. Hakkani-Tur, M. McTear, G. Riccardi, and G. Tur. 2008. Spoken language understanding: A survey. *IEEE Signal Processing magazine*, 25(3):50–58.

G. Leech and A. Wilson. 2006. EAGLES recommendations for the morphosyntactic annotation of corpora. Technical report, ILC-CNR.

C. Müller and M. Strube. 2003. Multi-level annotation in MMAX. In *Proceedings of SIGDIAL'03*.

J. M. Sinclair and R. M. Coulthard. 1975. *Towards an Analysis of Discourse: The English Used by Teachers and Pupils*. Oxford University Press, Oxford.

D. Traum. 1996. Conversational agency: The TRAINS-93 dialogue manager. In *Proceedings of TWLT 11: Dialogue Management in Natural Language Systems*, pages 1–11, June.

S. Varges, G. Riccardi, and S. Quarteroni. 2008. Persistent information state in a data-centric architecture. In *Proceedings of SIGDIAL'08*.