Dialog Intent Structure: A Hierarchical Schema of Linked Dialog Acts

Silvia Pareti, Tatiana Lando
Google Research Europe
{spareti,tlando}@google.com

Abstract

In this paper, we present a new hierarchical and extensible schema for dialog representation. The schema captures the pragmatic intents of the conversation independently from any semantic representation. This schema was developed to support computational applications, be applicable to different types of dialogs and domains and enable large-scale non-expert annotation. The schema models dialog as a structure of linked units of intent, *dialog acts*, that are annotated on minimal spans of text, *functional segments*. Furthermore, we categorise dialog acts based on whether they express a primary or secondary intent and whether the intent is explicit or implicit. We successfully tested the schema on an heterogeneous corpus of human-human dialogs comprising both spoken and chat interactions.

Keywords: Dialog Acts, Speech Acts, Intents, Dialog

1. Introduction

With the increasing popularity of conversational systems and chatbots, we are faced with the challenge of creating suitable dialog representations and annotated data to support the development of dialog modeling systems. Relying on form-filling strategies, with predefined dialog states sequences, falls short of providing a flexible and adaptable system that supports natural conversation, where the interaction is not constrained by a fixed plot with one precise goal. On the other hand, treating utterances or messages independently from each other fails to recognise the interconnected dialog structure and phenomena that are crucial to correctly understand and generate dialog interactions. Part of a larger dynamic and interactive language exchange, dialog messages have both semantic content and communicative functions. The latter provide the conversational context needed to correctly interpret the semantic content and determine what the current dialog state is.

The challenge of representing the communicative intent of dialog segments with dialog acts (DAs) has been addressed multiple times over the past decades. While earlier efforts, e.g. Map Task (Carletta et al., 1997), have focused on representing specific tasks and domains, the attention has shifted to more general purpose schemas such as the Dialogue Act Markup using Several Layers (DAMSL) (Allen, 1997) and the ISO 24617-2(Bunt et al., 2010; Bunt et al., 2012), a standard for dialog act annotation. The latter offers a powerful representation comprising 9 core dimensions and around 60 communicative functions. Communicative functions are linked to functional segments (FSs), defined by (Bunt et al., 2012) as "the unit of dialogue act annotation".

These schemas were not specifically developed to represent human-computer conversation. By developing a schema for the specific application, we can ensure that we identify useful categories that can help us understand the human dialog input as well as generate a suitable machine reaction while also reducing the tagset complexity. The latter is crucial to conduct large-scale non-expert annotation. In particular, such representation should enable non-constrained interactions were the dialog structure is not predefined around a specific scenario with a single task to solve, but transitions between dialog states are learned from data.

For this purpose, we developed a new hierarchical and extensible schema for dialog structure representation. This defines a set of dialog acts that express conversational intents and are tagged as being explicit or implicit and constituting the primary or a secondary intent of a FS. We defined such representation by developing the schema on a corpus of user-assistant dialogs that were collected between human participants. The schema was successfully applied to different dialogs (in terms of written or spoken modality as well as what the dialog is trying to achieve and around which topic), demonstrating that it provides a sufficiently abstract and robust representation of dialog interaction. This schema can be used to annotate resources and support the development of dialog understanding and generation systems that can successfully model the complexity of natural language.

2. Dialog Act Schema

The schema we introduce represents dialog, intended as a conversation between two or more participants, as a graph of interconnected intents. We developed the schema to support large-scale non-expert annotation and piloted it on different types of dialogs to ensure it is robust and meets an acceptable rate of inter-annotator agreement. The goals of the schema are to:

- Represent dialog structure as a sequence of dialog acts expressing a communicative function. This kind of representation supports understanding the intents expressed by each turn in the conversation and generating appropriate follow-up turns.
- Support any kind of dialog relevant to the development of chatbots and dialog agents, in spite of the medium of communication, the number, type and role of the participants or the domain or topic discussed.
- Be independent from semantic or domain specific representations.
- Support extensible granularity through a set of coarse and fine grained tags hierarchically organised. Tags

can be further extended to support finer-grained distinctions, such as capturing the exact type of request to repeat (e.g. louder, rephrase, same).

This representation was developed to support modeling an artificial dialog agent by providing tags that are informative of the conversational expectations at each given point in the dialog and help to determine what the agent response should be. For example, while the ISO 24617-2 (Bunt et al., 2012) would tag 'What?' as AutoNegative thus capturing that the speaker is signaling non-understanding, we instead focus on what reaction the speaker expects to trigger in the listener by tagging this as a request to repeat. This way we can capture that 'What?' and 'Can you repeat?' have the same conversational function and should trigger the same reaction.
We did limit the number of categories by avoiding infrequent and less-useful distinctions. For example, while in the ISO standard, response acts are categorised by the communicative function of the act they are responding to, we considered this as unnecessary since in our representation response FSs are linked to the FS that triggered the response. For example, instead of having 'DeclineRequest, DeclineSuggest, DeclineOffer' we just have one category for 'reject' that can be linked to an instruct, suggest or offer act.

2.1. Turns and Dialog acts

We use the **turn** as a basic unit of dialog, intended as an uninterrupted sequences of messages or speech events from a single participant. Depending on the transcription convention, backchannels can be transcribed as a separate interrupting turn. Turns may comprise multiple sentences or even paragraphs. We avoid using the utterance as a minimal unit of dialog since there is little consensus on how this should be defined and because it only applies to spoken dialogs. Dialog Acts are annotated on **functional segments (FSs)** (Bunt et al., 2010; Bunt et al., 2012): minimal spans of text that express one or more dialog intents. Since in our data FSs were very rarely conveyed more than 2 intents, we decided to constrain the annotation to assign no more than 2 DAs per FS. FS division do not correspond to sentence or clause division, one functional segment can cover several sentences. In Ex.1 the answer consists of several sentences which all correspond to the same intent - provide information in response to a question.

(1) A: Today I want to learn about social psychology.

B: In psychology, social psychology is the scientific study of how people's thoughts, feelings, and behaviors are influenced by the actual, imagined, or implied presence of others. In this definition, scientific refers to the empirical method of investigation...

All text in a dialog should be part of one and only one FS, i.e. we do not allow for text to be left unannotated nor for FSs to overlap or be nested. Nested or overlapping tags can provide additional challenges when processing the data automatically, both for building machine-readable representations and for developing tools to support the annotation. In the ISO standard (Bunt et al., 2012), FSs can theoretically be discontinuous or run across turn boundaries. To simplify the automatic processing and encoding of the annotation, we restricted FSs to be continuous spans of text within a turn. To account for cases where a conversational intent was expressed over a non-continuous span of text, we introduced a specific technical label 'goes_with' to connect the detached parts of the DA. The same label can be used to connect two parts of an utterance interrupted by another participant, which can frequently happen in spoken data (Ex.2).

(2) A: Assistant: For soup, they have Horiatiki, Marouli, Ascolibri...

B: User: Assistant!

C: Assistant:...for the Mesquite Grill they have House Specialty - Arni Paithakia

D: Assistant: Yes?

Spans of text that do not express any identifiable function, such as abandoned or unintelligible spans, are annotated with the 'no-intent' tag.

2.2. Structure of the schema

We devised a hierarchy of categories grouped based on the component that is needed to process and react to a given state in the conversation. Tags are grouped into 3 categories focusing on the assistant reaction: interactional, social, other. These are motivated by the different processing that is needed: interactional acts require that we understand and process the information conveyed; social acts are often formulaic and do not contribute towards the goal of the conversation; other acts include DAs that do not express any intent or any recognised intent or that we need to merge with another FS before processing it.
The tagset inventory (Table 1) is organised in a hierarchy defined on 3 levels of granularity, comprising: 6 higher level tags, 14 middle level tags and 34 lower level tags. The naming of the tags follows a hierarchical approach and inherits the full path through the schema, such as: request.query.open. The act name on the right of the dot is a sub-act of the name on the left. The fine-grained level of DAs can be further splitted into sub-acts to introduce finer distinctions.
The 6 major classes group DAs according to the intent of the speaker and the goal they are trying to achieve:

- **Request**: The set of acts that are intended to elicit some reaction from the listener. Information request usually take the form of a question (directly or indirectly formulated). Action requests are formulated as instructions to accomplish a task or require some action on either or both the speaker and listener's sides.

- **Respond**: Respond acts are complementary to request acts, which they usually follow. Respond acts are answers to information requests and also reactions to an action request.

- **Assert**: Assert acts cover the transmission of information that is not requested in previous discourse, unlike respond acts, nor expect the listener to provide any sort of reaction. These are usually expressions of opinion or statements setting the ground for further interaction.
- **Social**: Social acts have purely social intent and are usually expressed in natural conversation to conform to social expectations. These include greetings, politeness expressions and expressions of sympathy and agreement. These expressions are often formulaic and they can be omitted from the conversation without affecting its structure or comprehension.
- **Other**: A set of acts used to handle cases that do not fit in the 4 previous categories. This includes FSs where the intention of the speaker is unclear or not collaborative, like self talk or abandonment.
- **Add_on**: A set of technical labels used to handle corrections or the continuation of an interrupted segment. These labels should be used to pre-process the data by joining these FSs with those they attach to.

2.3. Dialog Act primary and implicit tagging

When more than one intent is expressed by one single FS, we categorize them as **primary** and **secondary**. Ex.3(B2) conveys two intents: agreeing to the proposal to get some tea and expressing gratitude. These are anchored to the same span of text and cannot be separated as we do in Ex.3(B1) where the pipe denotes the boundary between two distinct FSs. The primary intent must correspond to the most salient intent at that point in the conversation, i.e. the intent that we need to identify to understand the dialog state and provide a conversationally appropriate reaction, in Ex.3 - answering to a proposal. Every functional segment is annotated with one primary dialog act and might also have one secondary act.

(3) A: Would you like some tea?

B(1): No,| thanks!

B(2): Thank you so much.

Both primary and secondary intents can be explicitly or implicitly expressed. In Ex.3(B2) the primary intent (agreeing to a proposal) was implied by explicitly using a polite expression. We introduce a specific label, **implicit**, to mark similar cases where the DA is not expressed by the span of text corresponding to the functional segment, but rather conveyed through entailment or implicature. If a FS has only one intent it is always considered primary and explicit. While other existing schemas allow for a FS to have multiple DAs associated, they do not explicitly mark these as being the most salient intent of the FS nor whether the intent was implicitly expressed. We take the primary intent to connect the DAs into a graph structure and we identify implicit intents since they can require different reaction strategies. Ex.4 and Ex.5 both express a request, however, in the first example this is implicitly expressed through a conversational implicature. While the final goal might be in both cases to find a restaurant, the assistant answers are not interchangeable. In Ex.4, since the information was implicit requested, the assistant can offer to provide it while also making sure that that is indeed the intent. In Ex.5 instead, an offer would not be an appropriate answer and the assistant can directly provide the information.

(4) A: I'm hungry.

B: Would you like help finding nearby restaurants?

(5) A: Find me nearby restaurants

B: Sure. There is an Italian ...

2.4. Dialog Act linking

In addition to assigning DA labels to FSs, we also connect them in a graph. The structure of the dialog is nonlinear, that is, a DA is not necessarily connected to the immediately preceding one, but might relate to any preceding DA. This can be due to the medium of communication causing delays such as a chat message being written while the conversation continues, but also to the way we structure and connect thoughts. One example is when one participant asks a number of questions in a row and another answers them one by one, as in Ex.6:

(6) A(1): who are the actors in Pete's Dragon?

A(2): how are the reviews for this movie?

A(3): Is it directed by anyone famous?

B(1): Of the main characters, Bryce Dallas Howard plays Grace, Robert Redford plays Meacham, Oakes Fegley plays Pete...

B(2): Rotten Tomatos gives it a Tomatometer of 87%, and an Audience Score of 82%

Keeping track on how DAs connect is crucial to computationally model dialog by representing conversational expectations as transition probabilities between states. We therefore represent dialog structure as a graph where each FS (except the very first one in the conversation) attaches to another one already existing in the dialog. We restrict linking to the primary intent expressed by the FS and do not link any secondary intent. In case an intent is not a follow-up of another intent (e.g. a topic starter), it is simply attached to the immediately preceding FS. In this case, we consider the FSs to be connected by a relation of proximity in the dialog. In applying the schema, we follow these principles:

- Each FS has one single primary dialog act associated and can have one additional secondary DA.
- Any DA can be marked as implicit.
- Each primary DA in the dialog is linked with only one preceding DA and with one or more following DAs.

An example of the full annotation is shown in Table 2. The dialog is a shortened version of an actual dialog form the annotated corpus we describe in the next section.

request	instruct query propose check	task, cancel, backtrack open, select, yn suggest, offer align, confirmation, repeat
assert	provide	elaboration, statement, opinion
respond	yes no reply notify	agree, accept disagree, reject open, select acknowledge, buy_time, success, failure
social	greetings politeness interpersonal	opening, closing apology, thanks, acknowledge_thanks feedback
other	other	other_intent, no_intent
add_on	add_on	goes_with, correct

Table 1: Dialog acts tags inventory.

Speaker	Text	Index and link	DA annotation
User	I want Sushi	1.1	Instruct.Task (primary, implicit) Provide.Statement (secondary, explicit)
Assistant	Hanabi sushi near Mountain view serves sushi	2.1	Propose.Suggest (primary, implicit) Provide.Statement (secondary, explicit)
User	When does the place close?	3.1	Query.Open (primary, explicit)
User	I want to have a long dinner.	3.2	Provide.Statement (primary, explicit)
Assistant	It closes at 11 pm.	4.1 linked to 3.1	Reply.Open (primary, explicit)

Table 2: Fully annotated excerpt of a dialog from the corpus

3. Corpus

The scheme was applied to a corpus of scenario-based English dialogs collected between two human participants. The corpus comprises both spoken and chat data. The data was collected in different locations, with different participants and different settings. The scenarios provided the participants with a role and a goal. Participants were asked to act either as an assistant (human or virtual) or a person asking for assistance (user), or as people talking to each other. Goals included locating information (either to make a selection or exploring a topic), performing a task or chitchat and casual conversation. After familiarizing themselves with the instructions participants improvised the dialog.

Chat dialogs were collected among volunteers through Google Hangouts. Spoken dialogs were recorded between hired participants either sitting in the same room separated by a divider to prevent them from seeing each other (to avoid non-verbal interactions) or talking over the phone. Spoken dialogs were then transcribed into text, including some non-verbal phenomena, while annotators had no access to the audio.

The full corpus was annotated with the scheme presented in Sec.2. and comprises 65 English dialogs: 28 spoken and 37 chats. Spoken dialogs are on average longer than chats resulting in the following turn distribution: 617 turns in spoken dialogs and 513 in chats. The corpus provided an heterogeneous testbed (see Table 3) to evaluate the schema applicability on multiple domains and different types of dialog).

Dimensions	
Modality	chat (28) , spoken (37)
Goal	information exploration (18)/information seeking (6) task-based (27), chitchat (14)
Topic	restaurants (13), movies (13), travel (8) sports (6), music (4), other (1-2 per topic)
Participants	person-person (14), person-assistant (51)

Table 3: Corpus composition

3.1. Pilot Corpora

The schema was piloted three times, each time on a new set of data. All the data used in earlier pilots was reannotated according to the final version of the tagset and contributed to the corpus described above. In the next section we will describe in more details how the pilots were organised, show the results and discuss the lessons learnt. The final (third) pilot was conducted on 20 English dialogs (10 spoken and 10 chat) and on an additional dataset of 2 spoken and 8 chat Italian scenario-based dialogs. The different datasets used in the pilots are summarized in Table 4.

4. Inter-annotator Agreement

In order to ensure that the schema is applicable to different types of dialogs with a satisfactory rate of inter-annotator agreement, we conducted three rounds of pilot annotation.

Datasets:	Dialogs	Turns	FSs
Full Corpus	65	1130	2116
- spoken	28	617	1253
- chats	37	513	863
2nd English pilot	24	492	953
3rd English pilot	20	458	558
- spoken	10	234	314
- chats	10	224	244
Italian pilot	10	305	364

Table 4: Corpora summary table

All pilots were conducted with professional linguists, however, minimal training was provided. Each pilot enabled us to identify critical points and further tune the schema and guidelines.
We run the first pilot in an early stage of the tagset development. This allowed us to test the initial design and intuitions on real data and to define a hierarchical and more intuitive structure. We also used the data to group never or rarely-occurring tags and to discover tag distinctions that were not satisfactorily covered by the schema. We then run a second pilot to put the tagset and guidelines to the test. This highlighted some remaining issues. Finally, we run a third pilot with the latest tagset and revised instructions. To provide more substantial training, we asked the annotators to revise the annotation on the data from the previous 2 pilots according to the latest guidelines and to collect and discuss any point of disagreement. In addition to annotating English dialogs, we started to put the applicability of the schema to other languages to test by running the third pilot also on the Italian dataset. While we did not encounter any substantial differences and we achieved comparable agreement results, we foresee that language and cultural characteristics have an impact on the dialog strategies used which could affect the dialog structure, tag distribution and the frequency of intents being conveyed implicitly. We plan to conduct pilot annotations in several languages from different families in the future.
In this section we describe the results from the third pilot and compare them to the second pilot.

4.1. Methodology

The annotation was performed in three steps:

- `functional segments splitting`: each turn in the dialog was split into one or more functional segments. This was performed by 2 annotators and the annotation was then revised and reconciled.

- `dialog act annotation`: 3 annotators assigned one or more dialog act labels to each functional segment. In addition, they added labels for *primary intent* (one per functional segment) and *implicit*.

- `dialog act linking`: 2 annotators identified non-linear links in the dialog and manually added the index of the preceding FS. Annotations were then revised and reconciled.

For the Dialog Act annotation, we introduced fallback tags *other* on the lower level of the hierarchy (e.g. request.instruct.other). The guidelines created for the pilot instructed annotators to use the fallback tags when they could not find a suitable fine-grained tag. This approach ensured that any tag missing from the original tagset would be identified. The initial tagset for the pilot did not contain the tag notify.buy_time. The pilot data from the second pilot showed significant use of the tag notify.other (3.46%) to annotate stalling intents (e.g. *Give me a second*). As a result this was added to the final tagset and the use of a fallback tag decreased in the third pilot to 0.98%.

4.2. Pilot Results

4.2.1. Splitting into Functional Segments

As the first step, annotators were presented with the dialogs consisting of turns and their task was to identify the boundaries of the functional segments. This task was performed two-way. We then reconciled and revised the annotations to create a gold reference annotation. Agreement was calculated as precision and recall with respect to the gold annotation. The set of the boundaries needed was considered the gold annotation and we calculated precision and recall for identifying these boundaries. This metric does not account for all possible boundaries in the corpus: theoretically each turn could be split at any space between words. Thus the metric is quite pessimistic but gives more insight into annotators performance. The average F-score for the annotators was 92.2% for English data and 87.6% for Italian, however it appears that some of the annotators did not fully understand the task and performed significantly worse than others as can be seen in Table 5. More substantial training and examples in the guidelines can increase annotators' performance in the future. The annotation was reconciled before proceeding to the next step: Dialog Act annotation.

English	Annotator 1	Annotator 2	Average
Recall	91.3%	93.2%	92.3%
Precision	98.8%	86.3%	92.5%
Italian	Annotator 1	Annotator 2	Average
Recall	100%	100%	100%
Precision	63%	95.7%	79.3%

Table 5: Annotators' performance on FS splitting task

4.2.2. Dialog Act annotation

At this step annotators were presented with dialogs pre-segmented into functional segments. Their task was to assign one or two Dialog Acts to the FSs and in case they assigned two to identify which Dialog act is the primary intent of the FS and whether any of the intents is implicit. If they assigned only one DA it was considered Primary and Explicit by default. The task was performed three-way. The Kappa inter-annotator agreement (IAA) on Primary Dialog Acts is summarized in Table 6. For English we achieved .71 Kappa and for Italian .64, which are both reasonably good results given the complexity of the task.
The IAA agreement was very low on secondary DAs since annotators could (and often did) choose not to assign any

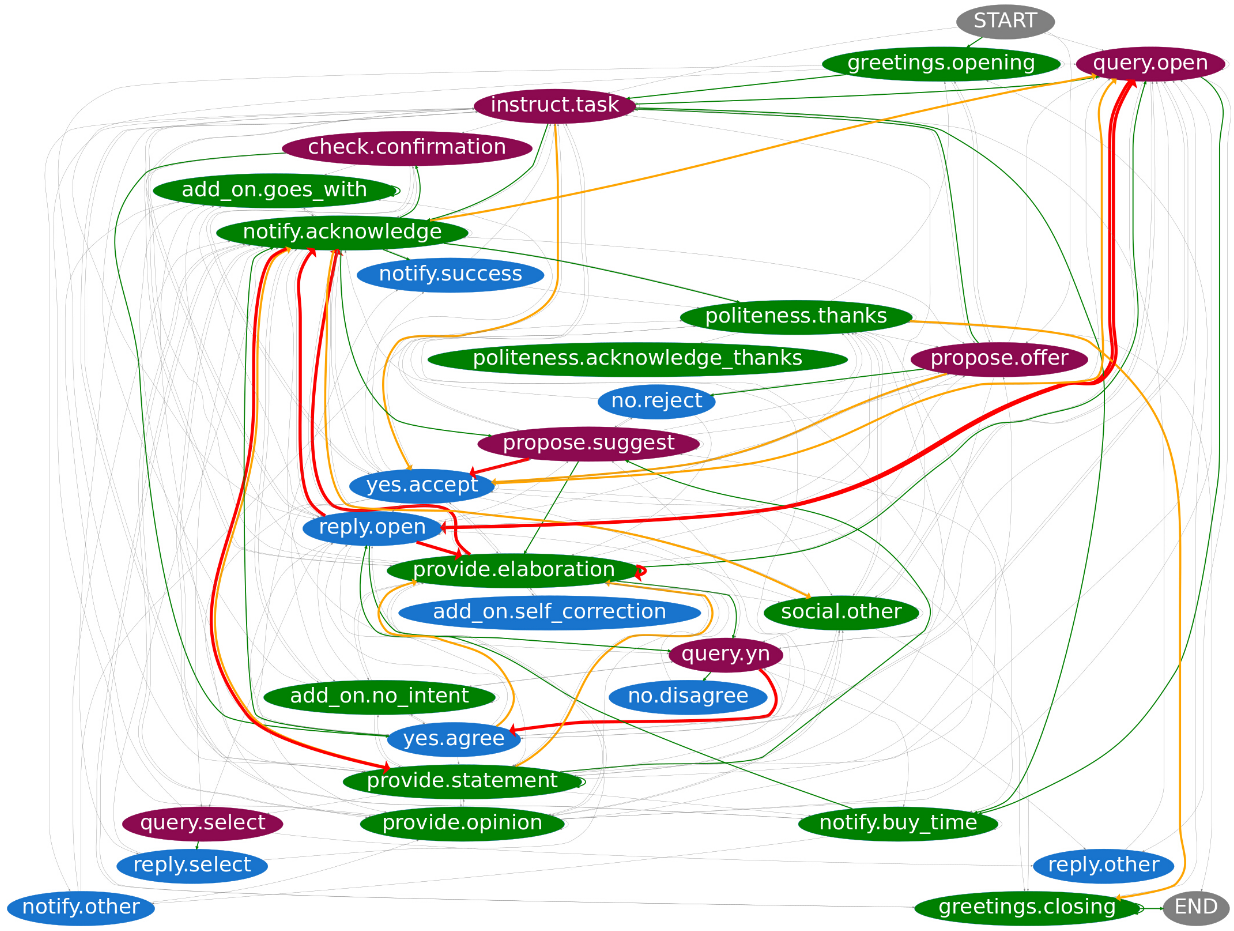

Figure 1: Dialog flow. Social DAs are in blue, request DAs in purple, respond DAs in blue and assert DAs in grey. Links are colour-coded based on how frequently that transition occurred in the corpus: > 5: grey; > 20: green; > 30: orange; > 50: red.

	Fine grained tags	Coarse grained
3rd English pilot	.71	.78
2nd English pilot	.58	.63
Italian pilot	.64	.72

Table 6: Kappa inter-annotator agreement for selecting the primary DA tag of a FS. Fine-grained tags comprised 34 DA labels, while coarse-grained tags 14.

secondary dialog act. Around 23% of the FSs were tagged by at least one of the annotators as having a secondary intent. In the reconciled pilot corpus only 11% of the FSs have two DAs associated.

The IAA was higher on chat dialogs. This could be explained by the lack of context for spoken dialogs and possible transcription errors and simplifications. The significant agreement increase from the second to the third pilot can be explained by several factors:

- The tag hierarchy was reorganised: we introduced a new tag to the inventory and removed rarely used ones. Moreover, there was no consensus on where to draw the boundaries between a request for information (i.e. the request to perform the task of answering a question as in 'tell me the weather forecast for tomorrow') and an instruct act or a question. We therefore collapsed these tags into only two distinctions based on whether we expect the interlocutor to provide some information or perform a task (that does not only result in providing information);
- Several tags were renamed making them more transparent for the annotators;
- We introduced a more substantial training stage and updated the guidelines based on annotators' feedback.

4.2.3. Dialog Act Linking

For this step, annotators were presented with the data pre-segmented into FSs and annotated with DAs and their labels. Their task was to identify cases when the FS expressed an intent that was not a response or reaction to the intent expressed in the immediately preceding FS and identify the correct antecedent. The average F-score for the annotators was 57.3% for English and 83.2% for Italian, detailed metrics are summarized in the Table 7.

The extremely low results from the English Annotator 2 can be explained by a misundertanding of the guidelines. The annotator systematically missed the links in those cases where there was a Notify or Social DA between Request

English	Annotator 1	Annotator 2	Average
Recall	95.2%	19%	57.1%
Precision	78.4%	57.1%	67.8%

Italian	Annotator 1	Annotator 2	Average
Recall	88%	80%	84%
Precision	78.6%	87%	82.8%

Table 7: Annotators' performance on FS linking task

and Response, i.e. in Ex.(7) FS (C) should be linked to FS (A) as the reply is following the answer and not the stalling expression.

(7) A: User: How about a different song from the same album, please?

B: Assistant: Hold on just a second

C: Assistant: here's the song All Over Me by Blake Shelton from the Same album.

4.2.4. Further steps

Although the IAA was very satisfactory for the primary DA annotation, we have identified some measures to further improve the consistency of the annotation:

- Annotators had no access to the audio of spoken dialogs. This can provide additional context (i.e. intonation) and reduce the ambiguity for DA-tagging and FS splitting.
- Annotators had to choose between 34 tags. While expert annotators could successfully perform this task, for large-scale non-expert annotation we plan to split the task into separate sub-tasks. We will ask annotators to make decisions on coarse-grained tags first and therefore simplify the task by reducing the number of tags to select from.
- Limited training was provided for FS splitting and linking which led to some annotators misunderstanding the task. This can be easily corrected with more training.
- Secondary intents are more subtle and harder to identify, for example any indirect request of the form 'Can you ...?' is implicitly deriving from a yes/no question, however this is such a common way to formulate requests that we no longer recognise this as a question. To achieve acceptable agreement also on this subtask, we have revised the guidelines to present more examples and have introduced rules to help annotators spot FS that have more than one intent.

5. Corpus Analysis

All the data from previous pilots was reannoted to match the final version of the tagset. The resulting corpus consists of 65 English dialogs and provides some insights into the intent structure of user-assistant free conversation. We generated the graph in Fig.1 from the corpus data to identify how dialog is structured and how DAs are connected. The distribution of the DAs in the corpus shows that human interaction, even in a user-assistant scenario, is not limited to information exchanges and task accomplishment. The frequency of all Request and Respond DAs in the corpus sums up to 65.4%. Around 29.9% of DAs comprise Assert, Social or Other acts. The rest 4.7% of FSs were annotated as Add_on. These results highlight that by constraining human-machine dialog interaction to be a sequence of requests and responses we are failing to capture all the conversational phenomena that make a dialog natural.

Although turns are a readily available unit in dialog data, they cannot replace FSs as the dialog unit of intention. In the human-human dialogs we collected, participants produce around 1.87 FSs per turn with some difference between spoken and written dialogs (2.03 vs 1.68 respectively).

7.6% of the FSs in the corpus have an implicit intent associated, with 2.4% having a primary implicit intent.

The tag distribution in the corpus is rather balanced, with the 10 most frequently occurring DAs, shown in Table 8, covering around 60% of the data. For the secondary DAs, the most frequent ones are: Provide.statement (28% frequency over secondary tags) which is often associated to an implicit primary act expressing a request; Query.yn (12% frequency) which is usually a question in yes/no form that expects however more information to be supplied; Check.confirmation and Yes.accept (both 9% frequency).

Dialog Act	Occurrences	Frequency
notify.acknowledge	211	9.97%
provide.elaboration	192	9.07%
reply.open	191	9.03%
query.open	177	8.36%
provide.statement	120	5.67%
instruct.task	113	5.34%
yes.accept	106	5.01%
propose.suggest	83	3.92%
politeness.thanks	80	3.78%
query.yn	79	3.73%

Table 8: Top 10 most frequently occurring primary DAs.

6. Conclusion

The dialog schema presented in this paper was created to support computational conversation modelling and represents dialog as a graph of intents providing a structure that can be used by NLG and NLU systems. The schema and annotation tasks were refined through several pilot iterations, leading to the final representation presented in this paper. The latest pilot was conducted on a corpus of 20 English and 10 Italian dialogs of different types. For the identification of the primary intent of a functional segment, we achieved Kappa agreement among three annotators of .71 for English and .64 for Italian. The tasks of splitting turns into functional segments and linking dialog acts achieved reasonable agreement. We plan to adapt the task and pilot the scheme with non-expert annotators and on several different languages.

7. Acknowledgements

We would like to thank Prof. Marilyn Walker from UC Santa Cruz and our colleagues Dilek Hakkani-Tur, Anna Khasin and Soha Sultan for the constructive discussions that laid the foundations of this schema.

8. Bibliographical References

Allen, J. (1997). Draft of DAMSL : Dialog Act Markup in Several Layers.

Bunt, H., Alexandersson, J., Carletta, J., Choe, J.-W., Chengyu Fang, A., Hasida, K., Lee, K., Petukhova, V., Popescu-Belis, A., Romary, L., Soria, C., and Traum, D. (2010). Towards an ISO Standard for Dialogue Act Annotation. In *Seventh conference on International Language Resources and Evaluation (LREC'10)*, La Valette, Malta, May.

Bunt, H., Alex, J., woong Choe, J., Fang, A. C., Hasida, K., Petukhova, V., Popescu-belis, A., and Traum, D. (2012). ISO 24617-2: A Semantically-based Standard for Dialogue Annotation. In *In Proceedings Eighth International Conference on Language Resources and Evaluation (LREC 2012*.

Carletta, J., Isard, S., Doherty-Sneddon, G., Isard, A., Kowtko, J. C., and Anderson, A. H. (1997). The Reliability of a Dialogue Structure Coding Scheme. *Computational linguistics*, 23(1):13–31.