HAL
open science

Dialogue Act Taxonomy Interoperability Using a Meta-Model

Soufian Salim, Nicolas Hernandez, Emmanuel Morin

To cite this version:

Soufian Salim, Nicolas Hernandez, Emmanuel Morin. Dialogue Act Taxonomy Interoperability Using a Meta-Model. 18th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing), May 2017, Budapest, Hungary. 10.1007/978-3-319-77113-7_24 . hal-01539976

HAL Id: hal-01539976

https://hal.science/hal-01539976v1

Submitted on 15 Jun 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Dialogue act taxonomy interoperability using a meta-model

Soufian Salim, Nicolas Hernandez and Emmanuel Morin

Université de Nantes
LS2N UMR 6004
2 rue de la houssinière, 44322 Nantes Cedex 03
{firstname}.{lastname}@univ-nantes.fr

Abstract. Dialogue act taxonomies, such as those of DAMSL, DiAML or the HCRC dialogue structure, can be incorporated into a larger meta-model by breaking down their labels into primitive functional features. Doing so enables the re-exploitation of annotated data for automatic dialogue act recognition tasks across taxonomies, *i.e.* it gives us the means to make a classifier learn from data annotated according to taxonomies different from the target taxonomy. We propose a meta-model covering several well-known taxonomies of dialogue acts, and we demonstrate its usefulness for the task of cross-taxonomy dialogue act recognition.

1 Introduction

Speech act theory [1] attempts to describe utterances in terms of communicative function (*e.g.* question, answer, thanks). Dialogue act theory extends it by incorporating notions of context and common ground, *i.e.* information that needs to be synchronized between participants for the conversation to move forward [2]. Dialogue acts are a fundamental part of the field of dialogue analysis, and the availability of annotations in terms of dialogue acts is essential to the machine learning aspects of many applications, such as automated conversational agents, e-learning tools or customer management systems. However, depending on the applicative or research goals sought, relevant annotations can be hard to come by. This work attempts to alleviate the costs of building systems based on dialogue act statistical learning and recognition. Supervised methods for classification are the norm for dialogue act recognition tasks, and since the annotation of new data is a costly and complicated endeavour, making annotated data reusable as much as possible would be a boon for many researchers.

Several corpora annotated in terms of dialogue acts are available to researchers, such as Switchboard, MapTask, MRDA, *etc.* [3–5]. Most of these corpora are annotated using taxonomies of varying levels of similarity. For example, the Switchboard corpus is annotated using a variation of the DAMSL scheme [6], MapTask and MRDA use their own taxonomies, and BC3 uses the MRDA tagset [5]. Intuitively, it makes sense that different researchers would use different taxonomies since not all information captured by such or such annotation scheme is relevant to each of every possible task, domain, and modality.

In a similar way, general-purpose taxonomies may ignore information that can be crucial to a given task, or specific to a particular domain. This is also why many researchers develop their own taxonomies, or alternatively use a variant or simplification of an existing taxonomy. These taxonomies are then applied to some data used in a few experiments, and often the data isn't even published.

This is all very wasteful, and at the source of an important issue. Annotating data in terms of dialogue acts is expensive and time-consuming, yet most of the resulting annotations aren't used as much as they could be because everyone uses a different taxonomy, or is interested in different domains. There is a need in the community for the availability of diverse corpora sharing the same annotations, as demonstrated by the significant efforts that were recently put in the development of the Tilburg DialogBank [7]. This project aims at publishing annotations for several common corpora using the ISO standard 24617-2 for DiAML [8]. While it is a very useful and commendable venture, it is important to remember that DiAML is not the answer to every task and every problem; there is too much potential information to annotate in dialogues to hope for a comprehensive and complete domain-independent annotation scheme. Even though DiAML is a standard, no standard will ever be sufficient to cover all possible situations of dialogue, and no standard can be useful to all dialogue analysis tasks. Even though ISO 24617-2 does provide guidelines for extending the standard, mainly by extending or reducing sets of annotations, the end result of applying them would always be the creation of a new albeit similar taxonomy.

Thus, rather than attempting to solve the problem of the inter-usability of corpora by proposing a better or more exhaustive standard, which is beyond our capabilities, we propose a different approach: the adoption of a meta-model for the abstraction of dialogue act taxonomies. The meta-model is built by breaking down dialogue act labels into primitive functional features, which are postulated to be aspects of dialogue acts captured by various labels across taxonomies. In this work, we demonstrate that it is possible to use a meta-model of taxonomies for annotation conversion, but also that such a model can be used to train a dialogue act classifier on a corpus annotated with a taxonomy different from the one it is designed to output annotations for.

This article is organized as follows. In Section 2 we discuss standardization efforts and the separation of dialogue act primitive features. We detail our meta-model in Section 3, before presenting our experimental framework in Section 4. In Section 5 we report the results of two sets of experiments. The first one evaluates methods for converting annotations from one taxonomy to another using the meta-model. The second demonstrates that it is possible to train a classifier to output annotations for a taxonomy different than the one used for the data it was trained on. We also experiment with complex taxonomies and show that at least some information can be identified without any annotation by training a DiAML classifier on DAMSL data and evaluating it on the Switchboard corpus. We conclude this article in Section 6.

2 Related work

As we mentioned previously, one approach to the lack of interoperability of dialogue schemes is the development of new standards and their assorted resources. From this perspective, the DialogBank [7] is the most recent effort to bring reliable and generic annotated data to the community. It is essentially a language resource containing dialogues from various sources re-segmented and re-annotated according to the ISO 24617-2 standard. Dialogues come from various corpora, such as HCRC MapTask, DIAMOND and Switchboard.

The authors' efforts are based on their conviction that DiAML is more complete semantically, application-independent and domain-blind. However, we believe that the standardization approach would benefit from efficient tools to improve the interoperability of existing annotations that do not conform to the DiAML recommendations. Firstly, because while it is true that DiAML is more complete semantically and less dependant on application and domain than the other existing annotation schemes, as demonstrated by Chowdhury *et al.* [9], it is not universal. For example, someone working with conversations extracted from internet forums will miss important features of online discourse by using DiAML, such as document-linking or channel-switching, all the while being burdened by a significant number of dimensions and communicative functions that are near absent from his or her data, such as functions of the time or turn dimensions. Secondly, we believe that dialogues are so complex and so rich that we cannot realistically expect a single annotation scheme to capture all of the information that may be relevant to any dialogue analysis system. There will always be missing information that would have been useful for something, and the pursuit of exhaustivity in annotation can sometimes lead to the development of cumbersome and impractical tools. Such ambitions may lead to the phenomenon known as *feature creep*, which is the continuous addition of extra features that are only useful for specific use-cases and go beyond the initial purpose of the tool, which can result in over-complication rather than simple and efficient design.

Perhaps it is preferable to build different taxonomies for different purposes, and focus the efforts put in the standard on making it more interoperable. Petukhova *et al.* [10] provide a good example of such efforts by providing a method to query the HCRC MapTask and the AMI corpora through DiAML. They notably report that the multi-dimensionality of the scheme makes it more accurate: *i.e.*, coding dialogue acts with multiple functions is a good way to make the taxonomy more interoperable. Indeed, the fact that utterances can generally have multiple functions is well known. Traum [11] notes that there are two ways to capture this multiplicity in a taxonomy: either annotate each function separately, which requires that each utterance can bear several labels, or group these functions into coherent sets and code utterances with complex labels.

The first option is the one preferred by DiAML, as it has the advantage of reducing the size of the tagset considered for each tagging decision, and better capture the multi-functionality of utterances. The idea behind this is that it is better to use several mutually exclusive tagsets than one tagset in which labels

may often share functional features. For example, let us consider the following dialogue

```
(Speaker 1) Now take a left
(Speaker 1) And then uuh
(Speaker 2) Turn right?
(Speaker 1) Yeah
```

With DiAML, it would be possible to annotate the second utterance with both the INSTRUCT and the STALLING labels. However, in the HCRC coding scheme, the INSTRUCT tag is separate from the UNCODABLE tag, and therefore the utterance can only be coded with one or the other. The issue here is that it can be difficult to decide how to code an utterance that shares some features with several labels. In effect, what multi-dimensional taxonomies do is separate function features to resolve such problems. But this separation is only meant to ease the annotation of utterances within a single taxonomy: in order to make a coding scheme more compatible with others, we believe that even function features within labels of the same dimension can be identified.

3 The meta-model

The purpose and manner in which dialogue acts (DA) should be defined has been discussed at length in the literature. Traum [11] raises many questions about the different aspects that should be considered when defining DA, such as "should taxonomies used for tagging dialogue corpora given formal semantics?" or "should the same taxonomy be used for different kind of agents?". The purpose of this work is not to promote or depreciate one approach over another, but to suggest a way to join them together.

We postulate that most taxonomies of dialogue acts can be generalized using primitive features as defining attributes of their labels. For example, an ANSWER in DiAML can't have an action-discussion aspect[1], but an ANSWER in DAMSL can. In both cases, the label can only be applied to an utterance elicited by the addressee. We could thus identify a few features of these labels to define the ANSWER label of these two taxonomies. The fact that the answer must be wanted by the addressee would be a common feature, and the fact that the answer cannot have an action-discussion aspect would be a differentiating feature.

We define a meta-model as the set of all features that can be used to define all the labels of a given set of taxonomies. A few benefits of such a tool are illustrated in Figure 1. The figure displays the manifestation of primitive features in utterances according to their label. A few acts are described, for the DiAML and the DAMSL schemes. Going back to our previous example, we see that the ANSWER labels are easy to compare when defined as sets of features, and doing so requires no human discernment: in the columns "`S.believes(p)`" and "`p.isInformation`", the cells are green for DiAML but blue for DAMSL. This

[1] *i.e.* It can't discuss the planning of an action, such as the utterance "ok I'll reboot my computer then".

Fig. 1. Example of a meta-model for six labels from DiAML (top) and DAMSL (bottom). Medium dark (green), "A", is always present in utterances (the definition includes the feature), dark (red), "N", is never present in utterances (the definition includes the negation of the feature), light (blue) is sometimes present in utterances (the definition does not include the feature). Feature designations use several shorthands: `S` stands for "Speaker", `A` for "Addressee", `p` for "(the uttered) proposition" and ¬ for "not". Therefore, `S.believes(p)` could be rewritten as "the speaker believes the uttered proposition to be true", and represents a single feature.

means that an answer must be genuine in DiAML, but answers that are lies are accepted in DAMSL. Moreover, in DiAML an answer must be informational - *i.e.* it cannot be an action-discussion utterance, nor a declarative act - which is not the case in DiAML. For example, answering to a request for action can be an Answer in the sense of DAMSL but not for DiAML. A computer could compare them, which would be impossible if presented with written definitions. We can observe in Figure 1 that when two labels share colour-codes everywhere, they are essentially the same label, when in places a square that is blue in one is red or green in the other, the second label is a specialization of the first one, and when there are opposing green and red squares, they are mutually exclusive.

For the purposes of this work, we built a meta-model including labels from the SWBD-DAMSL annotation scheme, the DiAML standard for the annotation of dialogue acts, and the HCRC dialogue structure coding system.

3.1 Feature formatting

We chose to format the features using a few basic components that can be linked together: *participants* ((`S`)peaker, (`A`)ddressee) use *verbs* (*e.g.* `provides()`, `wants()`, `believes()`) on *objects* (*e.g.* (`p`)roposition, (`f`)eedback, (`a`)ction), and these objects have *properties* (*e.g.* `isPositive`). The following example lists the features of the Auto Negative Feedback label in DiAML, meant for utterances providing negative feedback, such as "I don't get it" for example:

`S.provides(f)` ∧ `f.isAuto` ∧ ¬ `f.isPositive`

Features are separated by conjunction symbols. The first feature means "the speaker provides feedback", the second "the feedback concerns the speaker's understanding of an utterance", and the third "the feedback is negative".

This way of formatting features offers multiple advantages. Notably, it helps to avoid redundancy in features, and it allows for the use of logical operators (*e.g.* not $\neg$, or $\vee$, and $\wedge$). Moreover, using such a format makes it possible to break down features into learnable bits that can be used to train a classifier (for example, the presence of the object (`a`)ction in the feature). We also chose to make it similar to logical predicates so that it can be parsed and evaluated: although representing dialogue within a logical framework is an idea that has been explored in the literature before [12], we did not come across any work attempting to utilize the individual representation of dialogue act classes for data analysis and recognition. This aspect of our research however - parsing utterances to match logical definitions - is out of the scope of this paper. At the moment, each feature is treated as a boolean by the algorithms and the naming convention does not impact the experiments, *i.e.* "`S.provides(f)` $\wedge$ `f.isAuto` $\wedge$ $\neg$ `f.isPositive`" is equivalent to "$feature_a = true, feature_b = true, feature_c = true$".

However, the main advantage of this formulation is that it allows us to use concepts such as "belief" or "feedback" across multiple features, and implement theoretically grounded notions in the meta-model's building blocks. These elements reflect the conceptual foundation of the taxonomies comprised within the meta-model. In the meta-model used in this work, the primitive features used hint at the fact that the researchers behind DAMSL, DiAML and HCRC subscribed to a certain vision of dialogue structure. Indeed, the features are predominantly built around the notions of belief, desire and intention [13, 14], as well as the linguistic notion of *grounding* [2]. However it is important to note that the meta-model itself is not linguistically motivated, and could incorporate elements from any theory. For example, should a meta-model integrate Verbal Response Modes [15], its features would necessarily capture notions such as the frame of reference or the source of experience. In effect, primitive features can describe characteristics of knowledge, intention and belief of the speaker and the addressee, as well as characteristics of action and acknowledgement.

3.2 Feature extraction

We based our features on the exact written definitions of their labels, as published in the literature by their authors. For example, the Auto-Negative Feedback label used in our earlier example, the written definition as found in the ISO 24617-2 guidelines is the following:

"Communicative function of a dialogue act performed by the sender, S, in order to inform the addressee, A that S's processing of the previous utterance(s) encountered a problem."

Theoretically, any number of features can be extracted from such a definition. Perhaps a feature signifying that the utterance bears an information, another one to signal that it is not information related to the task, another to mark

the utterance as potentially non-verbal *etc.* Our formalization of the label is "`S.provides(f)` ∧ `f.isAuto` ∧ ¬ `f.isPositive`". To reach that result from the definition, we used a simple principle: new features should only be introduced as a mean to distinguish the label from its parent or siblings[2].

All three of these features are therefore used to distinguish Auto-Negative Feedback from other labels. "`S.provides(f)`" means that the utterance informs the processing of a previous utterance's execution, and in doing so distinguishes feedback functions from general-purpose functions [3]. "`f.isAuto`" means that the feedback pertains to the speaker's own processing, and is used to distinguish the label from Allo-Negative Feedback, which pertains to the addressee's processing of an utterance. "¬ `f.isPositive`" means that the feedback signals a problem; this feature is used to distinguish it from Auto-Positive Feedback. No more than these three features are required to efficiently distinguish each of the feedback labels. This method aims at building a meta-model suited to label comparison, not at capturing all the information contained in an annotation.

4 Experimental framework

The experiments detailed in this paper deal with the conversion and recognition of dialogue acts across taxonomies. First we present the corpora we perform the experiments on, and then our implementation of the meta-model.

4.1 Corpora and taxonomies

Two corpora seem most relevant for our task: the Switchboard corpus [3][4] and the MapTask corpus [4][5].

Switchboard [3][6] is a very large corpus (over 200 000 annotated utterances) annotated with the SWBD-DAMSL coding scheme [16]. DAMSL is the first annotation scheme to implement a multidimensional approach (*i.e.* which allows multiple labels to be applied to a single utterance) and is a *de facto* standard in dialogue analysis. SWBD-DAMSL is a DAMSL variant meant to reduce the multidimensionality of the latter [6]. A portion of the Switchboard corpus, about 750 utterances, has also been annotated with the ISO standard 24617-2 for DiAML [7]. The standard is inspired by DAMSL, but expands on it and attempts to annotate dialogue with a more theoretically-grounded approach.

The MapTask corpus [4][7] is also a relatively large corpus (over 2 700 annotations) annotated using the HCRC dialogue structure coding system [17], which

[2] If the taxonomy is "flat", *i.e.* not hierarchical, all labels are treated as siblings.

[3] While not specified in the guidelines, Inform and in some cases Answer could arguably be considered a parent of all feedback labels

[4] `https://catalog.ldc.upenn.edu/ldc97s62`

[5] `http://groups.inf.ed.ac.uk/maptask/`

[6] `https://catalog.ldc.upenn.edu/ldc97s62`

[7] `http://groups.inf.ed.ac.uk/maptask/`

comprises twelve labels. A portion of this corpus, a little over 200 utterances, has also been annotated using the DiAML scheme, which makes it an ideal candidate for our first task, converting annotations from one taxonomy to another.

4.2 Experimental meta-model

We built a meta-model for the labels in the taxonomies of SWBD-DAMSL, DiAML and the HCRC coding system in the manner described in Section 3.2. It contains 108 different features built from 2 participant types, 19 verbs, 6 object types and 32 object properties.

5 Experiments

First, we experiment with annotation conversion within the same corpus to demonstrate the ability of the meta-model to act as an effective bridge between taxonomies. Then, we present our results with cross-taxonomy classifiers, that are trained on data annotated with a different taxonomy than the one they output annotations for.

5.1 Annotation conversion

In the context of the construction of the Tilburg DialogBank, significant efforts were put towards the re-annotation of corpora with DiAML annotations, such as the Switchboard corpus [18]. Such endeavours were met with some difficulties [19]. Some automation was employed, in the form of manually defined mappings between labels that had a many-to-one or one-to-one relation. Our experiment explores a new automated method for label conversion.

For this experiment we do not apply any supervised algorithm for dialogue act classification. We merely attempt to use the meta-model to convert annotations from one taxonomy to another on the same data. Since some data from the Switchboard corpus is annotated with both SWBD-DAMSL and DiAML tags, we use it in this experiment. We also use the utterances from the MapTask corpus that are annotated with both the HCRC dialogue structure coding system and the ISO 24617-2 annotation scheme.

Annotations of the source taxonomy are first converted to primitive features (the set of all features of all labels for the utterance), then reassembled into new annotations for the target taxonomy (including the NONE label). We first attempted to perform the second step by computing the cosine similarity between the features of the original label and the features of labels in the target taxonomy. The system would choose the label with the feature set most similar to that of the original label. We then repeated the experiment using a NaiveBayes algorithm. The system would classify sets of features into target labels. This system was evaluated through cross-validation, over ten folds. Results for both methods are reported in Table 1.

We compare our results to a simple baseline, called the direct conversion approach. It consists of using a NaiveBayes classifier trained on the combinations of tags from the source and target taxonomy. The baseline classifier does not make use of the meta-model at all.

Results were evaluated on a sample of 746 DA for the Switchboard (SWBD) corpus and 675 DA for the MapTask corpus. They are reported in Table 1.

Corpus	Source → Target	Accuracy
Baseline: direct conversion approach		
MapTask	DiAML → HCRC	0.60
MapTask	HCRC → DiAML	0.70
SWBD	DiAML → SWBD-DAMSL	0.60
SWBD	SWBD-DAMSL → DiAML	0.78
Labels recovered with similarity algorithm		
MapTask	DiAML → HCRC	0.60
MapTask	HCRC → DiAML	0.76
SWBD	DiAML → SWBD-DAMSL	**0.65**
SWBD	SWBD-DAMSL → DiAML	0.87
Labels recovered with NaiveBayes algorithm		
MapTask	DiAML → HCRC	**0.71**
MapTask	HCRC → DiAML	**0.82**
SWBD	DiAML → SWBD-DAMSL	0.64
SWBD	SWBD-DAMSL → DiAML	**0.93**

Table 1. Label conversion scores.

We see that both methods outperform the direct conversion baseline. We also observe that a simple classifier trained on very little data can have stronger performances for the task of converting annotations than using a similarity metric. The exception being the DiAML to SWBD-DAMSL conversion, for which results are almost identical. This confirms that the meta-model has value for the task of automated annotation conversion.

5.2 Cross-taxonomy classification

Three sets of results are reported for this experiment. The first one is our baseline: it comprises results for a straightforward DA recognition task: over ten folds of a corpus, a model is trained on nine tenth of the data and evaluated on the rest. This method requires data annotated with the target taxonomy to function. The next two sets of results are those of systems that attempt to reach similar levels of accuracy, but this time using data from annotations in a different taxonomy from the output annotations.

The first of those systems, system A, works as follows: 1) a model is trained on correct labels from the source corpus annotated according to the source taxonomy, 2) labels from the source taxonomy are projected on data from the target

corpus, 3) projected labels are converted into labels from the target taxonomy according to the method described in subsection 5.1.

The second system, system B, attempts to learn primitive features instead of labels: 1) a model is trained on correct *primitive features* from the source corpus annotated according to the source taxonomy, 2) the target corpus is automatically annotated in terms of primitive features, 3) labels from the target taxonomy are recognized from primitive features according to the method described in subsection 5.1.

5.3 Method

For classification, we use an SVM for our algorithm and tokens, lemmas and parts-of-speech tags as features. Each feature type is used to build a bag-of-n-grams model. The SVM classifier was implemented using the *liblinear* library for text classification and analysis (Fan *et al*, 2008). We use a bigram model without stopword removal. We use a heuristic based on WordNet [20] for lemmatization and the Stanford toolkit [21] for part-of-speech tagging.

Since one of our taxonomies is multidimensional, allowing each instance to be tagged separately (and optionally) in several different dimensions (*i.e.* categories), a system that would attempt to pick one tag out of a tagset comprising all labels for the taxonomy would not be appropriate. Rather than using a multiclass SVM on the entire set of labels, which would not be entirely appropriate either since in DiAML only one label per dimension can be applied to an utterance, we chose to split them into dimensional tagsets. We then added the None label to each tagset to capture utterances that should not receive any label. Therefore, for DiAML the provided results are averaged over the results obtained over each dimension. If some results seem high for DiAML, it's because a few dimensions - such as Allo Feedback for example - will mostly be annotated with the None label. This is not an issue for our evaluation since the systems used as baselines also benefit from it.

5.4 Results

Results are provided in Table 2. We observe that system B has much weaker performances than system A. Its accuracy is 22 and 13 points behind the direct dialogue act classifier, for DIAML and HCRC respectively. System A, by contrast, is only outperformed by 9 and 8 points. This suggests that many features are hard to learn, comparatively to DA classes.

We can see that while the system B performs poorly, the system A is fairly efficient, less than ten points behind the results of a direct dialogue act recognition classifier. Accuracy loss can be attributed to two factors: (1) error rates of label conversion, and (2) increased error rates from the classifier due to structural and linguistic differences between the corpora used in this experiment.

Source	Target	Accuracy
Baseline: direct dialogue act recognition		
SWBD (DiAML)	SWBD (DiAML)	0.83
MapTask (HCRC)	MapTask (HCRC)	0.59
A: DA recognition, decomposition then recomposition		
SWBD (DAMSL)	SWBD (DiAML)	0.74 (-0.09)
SWBD (DAMSL)	MapTask (HCRC)	0.51 (-0.08)
B: DA decomposition, recognition then recomposition		
SWBD (DAMSL)	SWBD (DiAML)	0.61 (-0.22)
SWBD (DAMSL)	MapTask (HCRC)	0.46 (-0.13)

Table 2. Macro and micro accuracies of a baseline classifier (label-to-label) and an indirect cross-taxonomy dialogue act classifier (label-to-features-to-label).

6 Conclusion

In this paper, we presented a meta-model for the abstraction of dialogue act taxonomies. We believe the meta-model to have many useful applications for dialogue analysis and taxonomical research. The main contribution of this work is to provide a method to build supervised dialogue act recognition systems that do not require data annotated with the target taxonomy, but merely data annotated with a taxonomy which captures relevant information. We showed that a classifier trained on SWBD-DAMSL annotations could output DiAML or HCRC annotations at an accuracy not much lower than a regular classifier.

In future work, we will start a more data-driven approach to primitive feature identification by experimenting with clustering methods on annotated data. We believe an automated method will remove author bias in feature selection and allow for greater reproducibility. In order to further establish the relevance of the system, we also plan to replicate methods used in state-of-the-art dialogue act recognition systems to better understand how well a classifier can perform without a large corpus of data annotated in the appropriate taxonomy.

References

1. Austin, J.L.: How to Do Things With Words. Oxford University Press (1975)
2. Traum, D.R., Hinkelman, E.A.: Conversation Acts in Task-Oriented Spoken Dialogue. Computational Intelligence **8** (1992) 575–599
3. Godfrey, J.J., Holliman, E.C., McDaniel, J.: SWITCHBOARD: Telephone speech corpus for research and development. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-92). Volume 1., San Francisco, CA, USA, IEEE (1992) 517–520
4. Anderson, A.H., Bader, M., Bard, E.G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., et al.: The HCRC map task corpus. Language and speech **34** (1991) 351–366
5. Shriberg, E., Dhillon, R., Bhagat, S., Ang, J., Carvey, H.: The ICSI meeting recorder dialog act (MRDA) corpus. Technical report, DTIC Document (2004)

6. Core, M., Allen, J.: Coding Dialogs with the DAMSL Annotation Scheme. In: Working Notes of AAAI Fall Symposium on Communicative Action in Humans and Machines, Boston, MA, USA (1997) 28–35
7. Bunt, H., Petukhova, V., Malchanau, A., Fang, A., Wijnhoven, K.: The DialogBank. In: Proceedings of the 2016 International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia (2016) 3151–3158
8. Bunt, H., Alexandersson, J., Choe, J.W., Fang, A.C., Hasida, K., Petukhova, V., Popescu-Belis, A., Traum, D.R.: ISO 24617-2: A semantically-based standard for dialogue annotation. In: Proceedings of the 2012 International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, Turkey (2012) 430–437
9. Chowdhury, S.A., Stepanov, E.A., Riccardi, G.: Transfer of Corpus-Specific Dialogue Act Annotation to ISO Standard: Is it worth it? In: Proceedings of the 2016 International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia (2016) 132–135
10. Petukhova, V., Malchanau, A., Bunt, H.: Interoperability of Dialogue Corpora through ISO 24617-2-based Querying. In: Proceedings of the 2014 International Conference on Language Resources and Evaluation (LREC 2014), Reykjavik, Iceland (2014) 4407–4414
11. Traum, D.R.: 20 Questions on Dialogue Act Taxonomies. Journal of Semantics **17** (2000) 7–30
12. Sadek, M.D.: Dialogue acts are rational plans. In: Proceedings of the ESCA/ETRW Workshop on The structure of multimodal dialogue" (VENACO II), Maratea, Italy (1991) 19–48
13. Grosz, B.J., Sidner, C.L.: Attention, intentions, and the structure of discourse. Computational linguistics **12** (1986) 175–204
14. Georgeff, M., Pell, B., Pollack, M., Tambe, M., Wooldridge, M.: The belief-desire-intention model of agency. In: Proceedings of the International Workshop on Agent Theories, Architectures, and Languages (LNAI 1555), Berlin, Germany, Springer (1998) 1–10
15. Stiles, W.B.: Verbal response modes and dimensions of interpersonal roles: A method of discourse analysis. Journal of Personality and Social Psychology **36** (1978) 693
16. Jurafsky, D., Shriberg, E., Biasca, D.: Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual. Technical report (1997)
17. Carletta, J., Isard, A., Kowtko, J., Doherty-Sneddon, G.: HCRC dialogue structure coding manual. Technical report (1996)
18. Fang, A.C., Cao, J., Bunt, H., Liu, X.: The annotation of the Switchboard corpus with the new ISO standard for dialogue act analysis. In: Proceedings of the 8th joint ISO-ACL Sigsem workshop on interoperable semantic annotation, Pisa, Italy (2012) 13
19. Bunt, H., Fang, A.C., Liu, X., Cao, J., Petukhova, V.: Issues in the addition of ISO standard annotations to the Switchboard corpus. In: Proceedings of the 9th Joint ISO ACL SIGSEM Workshop on Interoperable Semantic Annotation (ISA-9), Potsdam, Germany (2013) 59–70
20. Miller, G.A.: WordNet: a lexical database for English. Communications of the ACM **38** (1995) 39–41
21. Manning, C.D., Surdeanu, M., Bauer, J., Finkel, J.R., Bethard, S., McClosky, D.: The Stanford CoreNLP Natural Language Processing Toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Baltimore, MD, USA (2014) 55–60