Relating the Semantics of Dialogue Acts to Linguistic Properties: A machine learning perspective through lexical cues

Alex C. Fang[1]
Department of Chinese, Translation and Linguistics
City University of Hong Kong
Hong Kong SAR
acfang@cityu.edu.hk

Harry Bunt[2]
Tilburg Centre for Cognition and Communication
Tilburg University
The Netherland
harry.bunt@uvt.nl

Jing Cao[3]
College of Foreign Languages
Zhongnan University of Economics and Law
Wuhan, China
cjinhk2010@gmail.com

Xiaoyue Liu[4]
Department of Chinese, Translation and Linguistics
City University of Hong Kong
Hong Kong SAR
xyliu0@cityu.edu.hk

Abstract— **This paper describes a corpus-based investigation of dialogue acts. In particular, it attempts to answer questions about the empirical distribution of dialogue acts and to what extent dialogue acts can be automatically predicted from their lexical features. The Switchboard Dialogue Act Corpus is adopted and the SWBD-DAMSL tags used for automatic prediction. We show that 60-70% of the dialogue acts can be predicted from lexical features alone depending on different levels of granularity. We also present a mapping from SWBD-DAMSL tags to the tags of the new ISO standard for dialogue act annotation, as part of an ongoing investigation into the relationship between the structure and granularity of the tag set and classification accuracy. The paper concludes with discussions and suggestions for future work.**

Keywords-dialogue act; Switchboard corpus; SWBD-DAMSL; automatic classification; ISO dialogue annotation standard

I. INTRODUCTION

This paper describes a preliminary corpus-based investigation into the lexical characteristics of dialogue acts (DA) as part of a wider research programme bidding to relate communicative functions of spoken utterances to their underlying linguistic properties definable in terms of lexis, grammar and syntax. Different from previous studies, this study examines the full set of dialogue acts observed in a corpus of transcribed conversations through the automatic selection of sets of distinguishing features, defined in lexical, grammatical and syntactic terms, that maximally predict their corresponding DAs. In particular, the current study reported here attempts to chart the frequency distribution and hence the probability of occurrence of DAs in the Switchboard Dialogue Act (SWBD) Corpus that has been annotated with a variant of the DAMSL scheme (SWBD-DAMSL). At the same time, it also attempts to answer the question to what extent dialogue acts can be automatically predicted according to their lexical features such as word unigrams. In doing so, valuable empirical evidence can be collected to substantiate, in statistical terms, the important notion of multifunctionality and, indeed, multidimensionality of spoken utterances. It is felt that such statistical information has been generally lacking in dialogue act analysis while it is much needed in, for example, man-machine interactive systems for the challenging task of dialogue understanding. Finally, this paper will describe some preliminary work to map the SWBD-DAMSL scheme to the newly developed ISO standard for DA annotation, an attempt that we hope will help deepen our understanding of the different aspects in concern in the study of communicative functions of dialogue.

First of all, here are some preliminaries. DAs play a key role in the interpretation of the communicative behaviour of dialogue participants and offer valuable insight into the design of human-machine dialogue systems [1]. More recently, the newly developed ISO/DIS 24617-2 standard for dialogue act annotation defines dialogue acts as the "communicative activity of a participant in dialogue interpreted as having a certain communicative function and semantic content, and possibly also having certain functional dependence relations, rhetorical relations and feedback dependence relations" (p. 3). The semantic content specifies the objects, relations, events, etc. that the dialogue act is about; the communicative function can be viewed as a specification of the way an addressee uses the semantic content to update his or her information state when he or she understands the corresponding stretch of dialogue.

Continuing efforts have been made to identify and classify the dialogue acts expressed in dialogue utterances taking into account the empirically proven multifunctionality of utterances, i.e. the fact that utterances often express more than one dialogue act (see [2] and [3]). In other words, an utterance in dialogue typically serves several functions. See Example (1) taken from the SWBD Corpus (`sw_0097_3798.utt`).

(1) A: *Well, Michael, what do you think about, uh, funding for AIDS research? Do you…*
B: *Well, uh, uh, that's something I've thought a lot about.*

978-0-7695-4492-2/11 $26.00 © 2011 IEEE
DOI 10.1109/ICSC.2011.32

With the first utterance, *A* performs two dialogue acts: he (a) assigns the next turn to the participant Michael, and (b) formulates an open question. *B*, in his response, (a) accepts the turn, (b) stalls for time, and (c) answers the question by making a statement.

Human annotators often find themselves in a situation where they have a hard time deciding which of several DA tags to apply. Consider Example (2), taken from the Schiphol Corpus [4]:

(2) C: *Can you tell me the departure times for flights to Munich on Saturday?*
I: *For Munich I have 08:15, 9:30, 12:20, 18:30, and 20:45*
C: *And that's on Saturday too.*
I: *And that's on Saturday too.*

In this example, the last utterance by participant I can be interpreted as an answer or as a signal of what this participant has understood. In this case, the answer to the question which tag is appropriate would be: both! An annotator who can assign only one DA tag to an utterance, as is the case in the SWBD Corpus, finds himself in a predicament here.[1] It may be noted that this is not a case of ambiguity, where only one of two or more meanings is the intended one, but a case of multifunctionality, where two meanings are both intended. The same utterance also illustrates the phenomenon of ambiguity; taken in isolation, its communicative function can also be that of a check question, as the identical preceding utterance by participant C shows. (The two utterances were not only lexically identical but also had identical intonation.) As is typical for cases of ambiguity, this ambiguity is resolved by the dialogue context.

Three major features are often examined for automatic dialogue act recognition: prosodic information, words, and syntactic information [5]. Previous studies have shown that lexical cues (or cue phrases) and certain syntactic constructions in DAs demonstrate a high degree of correlation to DA recognition [5]. Four DA types were examined in [5], viz. Agreements, Continuer, Incipient Speaker and Yes-Answer. More recently, based on the Switchboard Corpus, [6] employed 16 manually identified features for automatic DA classification including long utterances with more than 10 words, question marks and exclamation marks. An F-score of 57% was reported. Again using the Switchboard Corpus, [7] studied words according to the Linguistic Inquiry and Word Count taxonomy (see [8]) where words are selected and grouped from a psycholinguistic point of view.

In the rest of the paper, we shall describe the SWBD Corpus and the procedures of automatic classification of DAs before a discussion of the findings. In addition, the mapping of SWBD-DAMSL to the ISO/DIS 24617-2 standard DA tag-set is also discussed as part of an ongoing investigation into the relationship between the structure and granularity of tag sets and classification accuracy. The paper will conclude with discussions and suggestions for future work.

II. Corpus Resource

This study uses the Switchboard Dialog Act (SWBD) Corpus as the corpus resource[2], where each segmented utterance is annotated for its communicative function according to a set of dialogue acts specified in the "SWBD-DAMSL" scheme [9]. The corpus contains 1,155 5-minute conversations, orthographically transcribed in about 1.5 million word tokens. It should be noted that the minimal unit of utterances for DA annotation in the SWBD Corpus is the so-called "slash unit" [10], defined as "maximally a sentence but can be smaller unit" (p. 16), and "slash-units below the sentence level correspond to those parts of the narrative which are not sentential but which the annotator interprets as complete" (p. 16). See Table I for the basic statistics of the SWBD Corpus.

TABLE I. Basic Statistics of the SWBD Corpus

Folder	Conversations	Slash-units	Tokens	Types
sw00	99	14,277	103,045	5,574
sw01	100	17,430	119,864	6,250
sw02	100	20,032	132,889	6,651
sw03	100	18,514	127,050	6,447
sw04	100	19,592	132,553	6,436
sw05	100	20,056	131,783	6,573
sw06	100	19,696	135,588	6,735
sw07	100	20,345	136,630	6,598
sw08	100	19,970	134,802	6,450
sw09	100	20,159	133,676	6,384
sw10	100	22,230	143,205	6,407
sw11	16	3,213	20,493	1,987
sw12	11	2,773	18,164	2,140
sw13	29	5,319	37,337	3,271
Total	1,155	223,606	1,507,079	20,895

The corpus comprises 223,606 slash-units, each marked up with only one type of DAs. See Example (3) taken from `sw_0002_4330.utt`, where `qy` is the code for "yes/no question".

(3) qy A.1 utt1: *{D Well, } {F uh, } does the company you work for test for drugs? /*

Altogether 303 different DA tags are identified throughout the corpus, which were clustered at three levels of granularity, reflecting the structure of the SWBD-DAMSL annotation scheme. Level 1 corresponds to the four top-level dimensions (or 'layers') of DAMSL; Level 2 corresponds to the nine classes of forward-looking functions (FLF) and backward-looking functions (BLF) defined in DAMSL, and Level 3 is that of the individual communicative functions. Table II summarizes these levels of dialogue act classification and shows some examples.

[1] It should be noted however that the corpus indeed contains a negligible number of utterances, 19 in total, with multiple DA tags. In such cases, the first tag was used in the experiment.

[2] The corpus is available online from the Linguistic Data Consortium (www.ldc.upenn.edu).

TABLE II. RE-CLUSTERING OF DAS

Granularity	Group size	Examples
Level 1	4	Communicative-status
Level 2	9	Statement; Agreement
Level 3	60	Statement-non-opinion; Accept

III. AUTOMATIC CLASSIFICATION OF DIALOGUE ACTS

The question to what extent dialogue acts can be learned or predicted automatically through machine learning methods according to linguistic cues can be answered through experiments that measure the classification accuracy for a set of annotated utterances. Given the fact that dialogue utterances are functionally ambiguous, especially if only lexical cues are taken into account, a low accuracy is generally to be expected, but the detailed results can provide information about the ambiguity of certain DA tags and about the usefulness of the structuring of an annotation scheme, making use of clusters and groupings of various degrees of granularity. This section explores these issues.

A. Data Pre-processing

For the benefit of the current study and potential follow-up work, the banners between folders were removed and each slash-unit was extracted to create a set of five files, which are named using the following format: `Folder-Text-Utterance-Speaker+Turn-Unit.txt` and `Folder-Text-Utterance-Speaker+Turn-Unit-Level.da`, where `.txt` contains the original utterance and `.da` contains DA tag labels for the various levels. For the first slash-unit in `sw_0001_4325.utt` the following set of files is created:

`sw00-0001-0001-A001-01.txt`	the original utterance
`sw00-0001-0001-A001-01-1.da`	the first level DA tag (out of 4)
`sw00-0001-0001-A001-01-2.da`	the second level DA tag (out of 9)
`sw00-0001-0001-A001-01-3.da`	the third level DA tag (out of 60)
`sw00-0001-0001-A001-01-4.da`	the original DA tag (out of 303)

The standoff markup allows utterances to be marked up according to different levels of granularity without augmenting data size for the classification tasks. More importantly, it is also possible to plug in a file with another tag set, such as that of the ISO standard (ISO/DIS 24617-2), which we intend to apply to the same corpus for a comparative study. Next, the component lexical items for each utterance are extracted from `.txt` files, and for the current investigation, word unigrams were extracted and orthographical forms retained without lemmatization.

B. Classification Tasks

Three classification tasks were defined according to the three levels of dialogue act clustering. Tables III and IV present the basic statistics for classification tasks 1 and 2 respectively.

TABLE III. BASIC STATISTICS OF DAS AT LEVEL 1

Level 1 DA clusters	# of Slash-units	%
backwards-communicative-functions	68,541	30.653
communicative-status	15,902	7.112
forward-communicative-functions	113,401	50.715
other	25,762	11.521
Total	223,606	100.00

The DA clusters are arranged alphabetically in Table III. As can be noted, `forward-communicative-functions` is the largest DA group, accounting for over 50% of the utterances, while `backwards-communicative-functions` is the second largest type which consists of about 31% of utterances. See Appendix for a complete listing of the 60 DA tags for Level 3, sorted according to frequency in descending order. The frequency distribution is uneven, with the top 12 tags accounting for over 90% of the total occurrence of DA tags. This observation also suggests that some directed effort on the resolution of the top ranking DAs will lead to cost-effective returns on the overall DA prediction performance.

Table IV illustrates how Level 1 DA clusters are further divided into Level 2 clusters. The statistics shows that `statement` is the largest DA group, taking up about 45% of the total utterances, followed by `understanding` (22%). An unbalanced distribution of DA types can be observed here, with one extreme case, where only 103 utterances are annotated as `committing-speaker-future-action`, accounting for only 0.046% of the total number of utterances. It should be pointed out that `information relation` (see [11]) is not coded in the SWBD Corpus [9].

TABLE IV. BASIC STATISTICS OF DAS AT LEVEL 2

Level 1 DA clusters	Level 2 DA clusters	Slash-units	%
backwards-communicative-functions	agreement	12,187	5.450
	answer	8,129	3.635
	understanding	48,225	21.567
communicative-status	communicative-status	15,902	7.112
forward-communicative-functions	committing-speaker-future-action	103	0.046
	influencing-addressee-future-action	10,344	4.626
	other-forward-functions	3,120	1.395
	statement	99,834	44.647
other	other	25,762	11.521
	Total	223,606	100.00

C. Experiments and Results

In these three classification tasks, the Naïve Byes Multinomial classifier was employed, which is available from Waikato Environment for Knowledge Analysis (Weka; [12]). 10-fold cross validation was used and the results were evaluated in terms of precision, recall and F-score (F_1). Results from Level 3 are not reported here since only 19 out of 60 DA types achieve an F-score above zero, but some of the DAs at this level will be used as cases in the granularity analysis.

TABLE V. RESUTLS FROM TASK 1

Level 1 DA clusters	Precision	Recall	F_1
backwards-communicative-functions	0.896	0.861	0.878
forward-communicative-functions	0.705	0.925	0.800
communicative-status	0.416	0.088	0.146
other	0.251	0.055	0.090
Weighted Average	0.691	0.746	0.696

Table V presents the results for classification task 1. The DA clusters are arranged according to F-score in descending order. As can be noted, `backwards-communicative-functions` achieves the best F-score of 0.878, followed by `forward-communicative-functions` with an F-score of

0.800. The cluster `backwards-communicative-functions` has the highest precision, of about 90%, whereas `forward-communicative-functions` has the highest recall of over 92%. It can also be noted that the F-score of both `communicative-status` and `other` is only around 1%. The confusion matrix in Fig. 1 reveals more details of the performance.

a	b	c	d	<-- classified as
1416	22334	1647	364	a = other
3981	**104885**	3099	1436	b = forward-communicative-function
200	9150	**59016**	175	c = backwards-communicative-function
48	12329	2120	**1405**	d = communicative-status

Figure 1. Confusion matrix for Task 1

The matrix shows for instance that `backwards-communicative-functions` has some lexical similarity with `forward-communicative-functions`, evidenced by the 9,150 instances classified into the latter group, an indication of the multifunctionality of utterances for sure but a source of ambiguity for automatic learning. The matrix also helps to explain the poor performance of `communicative-status` and `other`. It can be noted that a majority of instances of `other` has been identified as `forward-communicative-functions`, which seems to suggest that a substantial common lexicon is used in both DA groups, which may cause ambiguity for automatic DA identification. The same is true for `communicative-status`.

A breakdown of the DA clusters at Level 2 will further reveal the performance of lexical features of DAs. Table VI presents the results for classification task 2. Again the DA clusters are arranged according to F-score in descending order.

TABLE VI. RESUTLS FROM TASK 2

Level 2 DA clusters	Precision	Recall	F_1
understanding	0.738	0.896	0.809
statement	0.644	0.928	0.760
other-forward-functions	0.952	0.357	0.520
influencing-addressee-future-action	0.700	0.359	0.475
agreement	0.561	0.172	0.263
communicative-status	0.426	0.100	0.162
answer	0.432	0.059	0.104
other	0.257	0.061	0.099
committing-speaker-future-action	0.000	0.000	0.000
Weighted Average	0.599	0.655	0.584

The data reveal interesting patterns of the DA groups. First, it is within our expectation that with the increase of the number of DA groups, the performance drops from 69.6% in task 1 to 58.4% in task 2. Secondly, `understanding` achieves the highest F-score (80.9%), much higher than the other two subdivisions of `backwards-communicative-functions` (i.e. `agreement`, 26.3%; `answer`, 10.4%). This seems to indicate that the lexicon used in `backwards-communicative-functions` is more likely to signal understanding than to articulate an agreement or to provide an answer. Thirdly, `statement` performs the second best with an F-score of 76%. Finally, it is worth noting that lexical cues fail to identify any instances of `committing-speaker-future-action`.

Again, the confusion matrix reveals the ambiguity among the DA groups; see Fig. 2. The matrix shows that within `agreement`, about 61% of the instances have been identified as `understanding`, and 18.5% as `statement`. The granularity analysis on the DAs at Level 3 also shows that `accept`, as a sub-division of `agreement` and with an F-score of 0.325, has 58.7 % of instances that are classified as a sub-division of `understanding`, and 16.3% that fall into the sub-divisions of `statement`. For example, the most frequently used words in `acknowledge` under `understanding` are *uh-huh*, *right* and *yeah*, which are also frequently used in `accept`. Such a shared lexicon causes ambiguity and thus difficulty in automatic DA classification.

a	b	c	d	e	f	g	h	i	<-- classified as
3717	305	5884	34	293	15	89	7	0	a = influ-addressee-fut-actn
388	**1579**	21889	126	1322	80	357	20	0	b = other
798	3923	**92611**	222	679	189	1399	13	0	c = statement
34	14	2260	**2094**	7464	265	56	0	0	d = agreement
191	173	4061	360	**43233**	64	138	5	0	e = understanding
22	67	3252	831	3402	**482**	70	3	0	f = answer
122	58	12584	49	1474	17	**1590**	8	0	g = communicative-status
32	28	1154	13	746	5	27	**1115**	0	h = other-forward-function
9	2	88	3	1	0	0	0	**0**	i = com-speaker-fut-actn

Figure 2. Confusion matrix for Task 2

In the case of `answer`, 41.9% of the instances fall into the category of `understanding`, 40% into `statement`, and 10% into `agreement`. It seems to suggest that a large number of lexical items used in `answer` also occur in `understanding` and `statement`. In addition, when most of instances of `other` fall into `forward-communicative-functions` at the first level classification, the granularity analysis of Level 2 DA clusters reveals that about 90.2% of the instances of `other` are mistakenly identified as `statement`.

D. Discussion

Lexical items that are cues for more than one DA type can be a reflection of the multifunctionality of dialogue utterances as well as of their functional ambiguity for automatic DA classification or recognition, thus illustrating the potential speech understanding challenges in interactive man-machine dialogue systems. To establish new ways to further improve the performance, case studies are made with the help of ISO/DIS 24617-2 to analyze the classification difficulty in the SWBD Corpus and to exploit the advantages of the ISO standard in DA annotation. Consider the DA type of `accept` in SWBD-DAMSL. It is a broad function applicable to a range of different situations. For instance, `accept` annotated as `aa` in Example (4) taken from `sw_0005_4646.utt` corresponds to `agreement` in ISO.

(4) sd A.25 utt1: *{C Or } people send you there as a last resort. /*
aa B.26 utt1: *Right, /*

However, `accept` (`aa`) in Example (5) taken from `sw_0423_3325.utt` actually corresponds to `accpetSuggestion (addressSuggestion)` in ISO which takes into consideration the context features, i.e., the previous utterance in this case.

(5) ad B.128 utt2: *{C so } we'll just wait. /*
aa A.129 utt1: *Okay, /*

As a matter of fact, `accept` in SWBD-DAMSL may correspond to four different ISO DAs:

- `agreement`,
- `acceptRequest(addressRequest)`,
- `acceptSuggestion(addressSuggestion)` and
- `acceptOffer(addressOffer)`.

The broad definition of `accept` (as well as of `reject`) in SWBD-DAMSL makes it applicable in a variety of different contexts and makes it harder to recognise than the corresponding narrower DA types of ISO. This carries over to the `agreement` cluster at Level 2, which suffers from the same broadness and is therefore hard to recognise/predict, with an F-score of 26.3% only. In other words, the empirical evidence shows that the definitions of `accept (part)`, `reject (part)` and `maybe`, and therefore of `agreement`, are not well chosen and thus difficult to apply in human language and speech application systems. The more precisely defined corresponding functions in the ISO DA annotation scheme, in comparison, may be expected to perform better.

The second case concerns `hedge`, a sub-division of `other` in SWBD-DAMSL. A hedge expresses uncertainty or lack of confidence on the part of the speaker, and it typically corresponds to an uncertain statement used to qualify an earlier statement or answer. See Example (6) from `sw_0093_3227.utt`.

(6) qy B.18 utt1: *Do you like rap? /*
no A.19 utt1: *{F Um,} some of it, {F um,}/*
H A.19 utt2: *it depends. /*

Results show that 90.2% of the instances in `hedge` have been identified as `statement`, which suggests a high degree of similarity in terms of lexical cues between `hedge` and `statement`. However, from the viewpoint of ISO, hedges are categorized together with statements in the same DA function of `inform`, qualified as being conditional or uncertain. Therefore, using the ISO standard we may expect a better performance of a more clearly defined DA function (i.e. `inform`).

The results obtained from the experiments reported in this section have provided empirical evidence in support of the multifunctionality of utterances in terms of their communicative functions. In addition, it is also seen that DAs at different granularity levels exhibit different degrees of accuracy when it comes to automatic classification or detection. While a coarse granularity is expected to offer an F-score of 69.6% at Level 1, the weighted average dropped to 58.4% with a finer granularity at Level 2, suggesting an ambiguity problem related to the use of lexical cues without contextual information. Manual inspection of problematic cases suggests that a different grouping of DAs may produce better results (such as the treatment of hedge as belonging to the statement group by the ISO scheme). This observation has prompted the need to map the DAMSL DA tags in the SWBD Corpus to ISO tags so that the two schemes can be comparatively evaluated to produce better insight into the relationship between the linguistic properties and granularity of the DA tags through classification accuracy.

IV. Matching SWBD-DAMSL to ISO

The case of the broadly defined `agreement` cluster in SWBD-DAMSL illustrates that the accuracy of dialogue act prediction may depend on the granularity of analysis and the structuring of the dialogue act tag set. This section describes some ongoing work to map SWBD-DAMSL tags to ISO tags, in order to create a new version of the Switchboard Corpus that is tagged with ISO tags.

A. Description of the ISO standard

A basic premise of the ISO standard for dialogue act annotation (ISO/DIS 24617-2) is that utterances in dialogue are often multifunctional; hence the standard supports so-called ‘multidimensional tagging’, i.e. the tagging of utterances with multiple DA tags. It does so in two ways: First of all, it defines nine dimensions to which a dialogue act can belong:

- Task
- Auto-Feedback
- Allo-Feedback
- Turn Management
- Time Management
- Discourse Structuring
- Social Obligations Management
- Own Communication Management
- Partner Communication Management

Secondly, as the unit in dialogue to be tagged with DA information, it takes a so-called ‘functional segment’, defined as a “minimal stretch of communicative behavior that has one or more communicative functions” [13]. A functional segment is allowed to be discontinuous, and to overlap with or be included in another functional segment. A functional segment may be tagged with at most one DA tag for each dimension. Another important feature is that an ISO DA tag consists not only of a communicative function encoding, but also of a dimension indication, with optional attributes for representing certainty, conditionality, sentiment, and links to other dialogue units expressing semantic, rhetorical and feedback relations.

Thus, two broad differences can be observed between SWBD-DAMSL and ISO. The first concerns the treatment of the basic unit of analysis. While in SWBD-DAMSL this is the slash-unit, ISO employs the functional segment, which serves well to emphasise the multifunctionality of dialogue utterances. An important difference here is that ISO identifies multiple DAs per segment and assigns multiple tags via the stand-off annotation mechanism. Secondly, each slash-unit (or utterance) in the SWBD Corpus is annotated with one SWBD-DAMSL label, while each DA tag in ISO is additionally associated with a dimension tag and, when appropriate, with function qualifiers and relations to other dialogue units. See the following example taken from the Schiphol Corpus.

(7) A: *I’m most grateful for your help*

While the utterance in Example (7) would be annotated with only a functional tag in SWBD-DAMSL, it is annotated to contain the communicative function "inform" and in addition the dimension of social obligation management:

```
communicativeFunction = "inform"
dimension = "socialObligationManagement"
```

B. *Mapping SWBD-DAMSL to ISO*

When mapping SWBD-DAMSL tags to functional ISO tags, five cases can be distinguished, namely, exact matches, many-to-one matches, one-to-many matches, tags unique to ISO, and tags unique to SWBD-DAMSL. It is worth mentioning that the mapping is achieved in terms of semantic contents rather than the surface labels. Therefore, even for the exact matches, the naming in SWBD-DAMSL is not always the same as that in the ISO scheme, but they have the same or very similar meaning. Table VII lists the exact matches.

TABLE VII. EXACT MATCHES

SWBD-DAMSL	ISO
Open-question	Question
Dispreferred answers	Disconfirm
Offer	Offer
Commit	Promise
Open-option	Suggest
Hold before answer/ agreement	Stalling
Completion	Completion
Correct-misspeaking	CorrectMisspeaking
Apology	Apology
Downplayer	AcceptApology
Thanking	Thanking
You're-welcome	AcceptThanking
Signal-non-understanding	AutoNegative
Conventional-closing	InitialGoodbye

Table VIII shows the many-to-one matches. Such matches occur because semantically identical functions are sometimes given different names in SWBD-DAMSL in order to distinguish differences in lexical or syntactic form. For example, an "affirmative non-yes answer" is defined as an affirmative answer that does not contain the word *yes* or one of its variants (like *yeah* and *yep*). It is worth pointing out that `understanding` in the SWBD-DAMSL scheme groups together a variety of DAs under this header that are not always justifiable. As a result, when mapped to `AutoPositive` in ISO, some DA tags concerning signaling understanding such as `completion`, `downplayer` and `correct-misspeaking` are left out.

TABLE VIII. MANY-TO-ONE MATCHES

SWBD-DAMSL	ISO
Wh-question; Declarative wh-question	SetQuestion
Or-question; Or-clause	ChoiceQuestion
Yes-no-question; Backchannel in question form	PropositionalQuestion
Tag-question; Declarative Yes-no-question	CheckQuestion
Statement-non-opinion; Statement-opinion; Rhetorical-question; Statement expanding y/n answer; Hedge	Inform
Maybe; Yes-answer; Affirmative non-yes answers; Yes plus expansion; No-answer; Negative non-no answers; No plus expansion	Answer
Accept-part; Reject-part	Correction
Acknowledge; Acknowledge answer, Appreciation; Sympathy; Summarize/reformulate; Repeat-phrase	AutoPositive

The most complex issue is with the one-to-many matches, where a DA function in SWBD-DAMSL is too general and corresponds to a set of different DAs in ISO. The case of the `accept` function discussed earlier illustrates that some SWBD-DAMSL tags do not correspond to semantically well-defined functions. Other cases include `reject`, `action-directive` and `other answers`. The best solution is to re-cluster them according to the previous utterance by the other speaker. Again, consider `accept` as an example. Table IX illustrates how `accept` can be further divided to match ISO tags.

TABLE IX. SUB-DIVISIONS OF ACCEPT

Previous DA in SWBD-DAMSL	ISO
Statement-non-opinion; Statement-opinion; Rhetorical-question; Statement expanding y/n answer, Hedge	Agreement
Offer	AcceptOffer
(the rest of cases of accept)	AcceptRequest
Open-option	AcceptSuggestion

Since there is no individual DA tag for `request`, the cases that correspond to `acceptRequest` in ISO will be determined after the other three matches have been made. Nevertheless, some of the matches in this category have to be performed manually. For example, `other answers` is a SWBD-DAMSL label for responses to `yes/no questions` which does not fall in any of the other SWBD-DAMSL answer categories. So in each case where a segment of dialogue is annotated as `other answers`, we shall determine what the communicative function exactly is of that segment. If the utterance contains "*I don't know*", it corresponds to negative auto-feedback in ISO whereas the other cases will have to be matched manually.

TABLE X. TAGS UNIQUE TO ISO

Dimension	DA Clusters
Information-Providing	Confirm
Allo-Feedback	AlloPositive; AlloNegative; FeedbackElicitation
Time Management	Pausing
Turn Management	TurnAccept; TurnAssign; TurnGrab; TurnKeep; TurnRelease; TurnTake
Discourse Structuring	InteractionStructuting
Own Communication Management	SingalSpeakingError; Retraction; SelfCorrection
Social Obligation Management	ReturnGreeting; InitialSelfIntroduction; ReturnSelfIntroduction; ReturnGoodbye

Tags unique to the ISO standard occur because the SWBD-DAMSL tag set does not cover certain dialogue phenomena, such as turn-taking, speech editing ("own communication management") and feedback about the addressee's processing of previous utterances ("allo-feedback"), or because it does not cover certain phenomena which do not occur in SWBD dialogues, such as requests to perform physical actions. As can be noted from Table X, SWBD-DAMSL lacks the dimensions *Allo-Feedback*, *Turn Management*, and *Own Communication Management*, and some DA types in *Time Management*, *Discourse Structuring*, and *Social Obligation Management*. Therefore, the ISO annotation scheme is more multidimensional than the SWBD-DAMSL scheme.

Finally, the tags unique to SWBD-DAMSL mainly concern the marking up of other phenomena than dialogue acts, such as self-talk and abandoned speech, which include

- `explicit-performative`,
- `exclamation`,
- `other-forward-function`,
- `quoted material`,
- `uninterpretable`,
- `abandoned`,
- `self-talk`,
- `3rd-party-talk`,
- `segment (multi-utterance)`, and
- `double labels`.

V. Conclusion and Future Work

This paper presents a preliminary corpus-based investigation into the lexical characteristics of dialogue acts in the Switchboard Dialog Act Corpus. In particular, it attempts to answer questions about the relative distribution of dialogue acts and to what extent dialogue acts can be automatically predicted according to their lexical features. This question is potentially important since it may provide empirical information about the multifunctionality and ambiguity of dialogue utterances. While results lend themselves to the understanding that utterances are multifunctional and hence ambiguous for automatic processing in man-machine dialogue systems, they nonetheless suggest that a granular approach to the DAMSL scheme and re-grouping of the DA tags may produce better results, a suggestion that emerged from a manual inspection of some problematic cases. A DAMSL-to-ISO mapping is discussed as an essential first step of an attempt to investigate the relationship between, on the one hand, the granularity of analysis and the structure of the classification scheme, and, on the other, classification accuracy. In the future, we plan to examine the effect of grammatical and syntactic cues on the performance of DA classification, with a specific view on whether dialogue acts exhibit differentiating preferences for grammatical and syntactic constructions that have been overlooked before. Indeed, we have already started to perform part-of-speech tagging and syntactic parsing on the SWBD Corpus, with a linguistically fine-grained analysis that allows for in-depth investigation. Once the afore-mentioned annotations are completed, it is possible to conduct similar experiments using grammatical and syntactic cues in combination with lexical features.

Acknowledgment

Research described in this article was supported in part by grants received from City University of Hong Kong (Project Nos 6354005, 9610188, 7008062 and 7008002) and the General Research Fund of the University Grant Council of Hong Kong (Project No 142711). The authors would like to thank the members of the Dialogue Systems Group at the Department of Chinese, Translation and Linguistics, City University of Hong Kong for comments and suggestions on an earlier draft.

References

[1] H. Bunt, J. Alexandersson, J. Carletta, J.-W. Choe, A.C. Fang, K. Hasida, K. Lee, V. Petukhova, A. Popescu-Belis, L. Romary, C. Soria, and D. Traum. "Towards an ISO standard for dialogue act annotation," In Proceedings of the Seventh International Conference on Language Resources and Evaluation. Valletta, MALTA, 17-23 May 2010.

[2] H. Bunt. "Multifunctionality and multidimensional dialogue semantics," In Proceedings of DiaHolmia Workshop on the Semantics and Pragmatics of Dialogue, Stockholm, 2009.

[3] H. Bunt. "Multifunctionality in dialogue and its interpretation," Computer, Speech and Language, Special issue on dialogue modeling. 2010.

[4] H. Prüst, G. Minnen and R.-J. Beun, "Transcriptie dialoogexperiment". Report No. 481, Institute for Perception Research, Eindhoven 1984.

[5] D. Jurafsky, E. Shriberg, B. Fox and T. Curl. "Lexical, prosodic, and syntactic cues for dialog acts," ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers, 1998.

[6] J. Araki, "Dialogue act recognition using cue phrases," available at http://nlp.stanford.edu/courses/cs224n/2010/reports/junaraki.pdf.

[7] N. Novielli and C. Strapparava, "Studying the lexicon of dialogue acts," In Proceedings of Irec, 2010.

[8] J. Pennebaker and M. Francis, Linguistic Inquiry and Word Count: LIWC. Erlbaum Publisher, 2001.

[9] D. Jurafsky, E. Shriberg, and D. Biasca, "Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual, Draft 13," University of Colorado, Boulder Institute of Cognitive Science Technical Report 97-02. 1997.

[10] M. Meeter and A. Taylor, "Dysfluency annotation stylebook for the Switchboard Corpus," available online at ftp://ftp.cis.upenn.edu/pub/treebank/swbd/doc/DFL-book.ps, 1995.

[11] M. Core and J. Allen, "Coding dialogs with the DAMSL annotation schema," AAAI Fall Symposium on Communicative Action in Humans and Machines, Boston, MA, 1997.

[12] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," SIGKDD Explorations, vol. 11, Issue 1, 2009, pp. 10-18.

[13] H. Bunt and A. Schiffrin, "Methodologial aspects of semantic annotation," In Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC 2006). Genova, Italy, May 24-26, 2006.

APPENDIX DAs AT LEVEL 3 RANKED IN DESCENDING ORDER ACCORDING TO FREQUENCY

Rank	Level 3 DAs	#	%	Cum %
1.	statement-non-opinion	73,435	32.84	32.84
2.	acknowledge-(backchannel)	38,372	17.16	50.00
3.	statement-opinion	26,399	11.81	61.81
4.	segment-(multi-utterance)	18,691	8.36	70.17
5.	abandoned	12,986	5.81	75.97
6.	accept	11,157	4.99	80.96
7.	appreciation	4,663	2.09	83.05
8.	yes-no-question	4,488	2.01	85.06
9.	double-labels	3,678	1.64	86.70
10.	yes-answers	3,040	1.36	88.06
11.	uninterpretable	2,696	1.21	89.27
12.	conventional-closing	2,585	1.16	90.42
13.	statement-expanding-y/n-answer	2,087	0.93	91.36
14.	wh-question	1,985	0.89	92.24
15.	no-answers	1,378	0.62	92.86
16.	acknowledge-answer	1,309	0.59	93.45
17.	declarative-yes-no-question	1,252	0.56	94.01
18.	hedge	1,227	0.55	94.55
19.	backchannel-in-question-form	1,057	0.47	95.03
20.	quoted-material	986	0.44	95.47
21.	summarize/reformulate	961	0.43	95.90
22.	affirmative-non-yes-answers	849	0.38	96.28
23.	o	815	0.36	96.64
24.	action-directive	752	0.34	96.98
25.	completion	730	0.33	97.30
26.	repeat-phrase	698	0.31	97.62
27.	open-question	657	0.29	97.91
28.	rhetorical-questions	578	0.26	98.17
29.	hold-before-answer/agreement	557	0.25	98.42
30.	reject	345	0.15	98.57
31.	transcription-errors:-slash-units	339	0.15	98.72
32.	signal-non-understanding	299	0.13	98.86
33.	negative-non-no-answers	298	0.13	98.99
34.	other-answers	286	0.13	99.12
35.	or-question	237	0.11	99.22
36.	conventional-opening	225	0.10	99.33
37.	or-clause	210	0.09	99.42
38.	dispreferred-answers	184	0.08	99.50
39.	exclamation	134	0.06	99.56
40.	3rd-party-talk	117	0.05	99.61
41.	downplayer	104	0.05	99.66
42.	self-talk	103	0.05	99.71
43.	tag-question	93	0.04	99.75
44.	declarative-wh-question	85	0.04	99.79
45.	apology	79	0.04	99.82
46.	thanking	78	0.03	99.86
47.	offer	65	0.03	99.89
48.	accept-part	59	0.03	99.91
49.	maybe	46	0.02	99.93
50.	commit	38	0.02	99.95
51.	quotation marks	26	0.01	99.96
52.	reject-part	23	0.01	99.97
53.	sympathy	19	0.01	99.98
54.	correct-misspeaking	13	0.01	99.99
55.	explicit-performative	9	0.00	99.99
56.	open-option	7	0.00	99.99
57.	other-forward-function	6	0.00	99.99
58.	no-plus-expansion	5	0.00	99.99
59.	you're-welcome	4	0.00	99.99
60.	yes-plus-expansion	2	0.00	100.00