Collaborative Annotation of Dialogue Acts: Application of a New ISO Standard to the Switchboard Corpus

Alex C. Fang[1], Harry Bunt[2], Jing Cao[3], and Xiaoyue Liu[4]

[1,3,4]The Dialogue Systems Group, Department of Chinese, Translation and Linguistics
City University of Hong Kong, Hong Kong, SAR
[2]Tilburg Centre for Cognition and Communication
Tilburg University, The Netherlands
[3]School of Foreign Languages, Zhongnan University of Economics and Law, China
E-mail: {[1]acfang, [3]cjing3, [4]xyliu0}@cityu.edu.hk, [2]harry.bunt@uvt.nl

Abstract

This article reports some initial results from the collaborative work on converting SWBD-DAMSL annotation scheme used in the Switchboard Dialogue Act Corpus to ISO DA annotation framework, as part of our on-going research on the interoperability of standardized linguistic annotations. A qualitative assessment of the conversion between the two annotation schemes was performed to verify the applicability of the new ISO standard using authentic transcribed speech. The results show that in addition to a major part of the SWBD-DAMSL tag set that can be converted to the ISO DA scheme automatically, some problematic SWBD-DAMSL tags still need to be handled manually. We shall report the evaluation of such an application based on the preliminary results from automatic mapping via machine learning techniques. The paper will also describe a user-friendly graphical interface that was designed for manual manipulation. The paper concludes with discussions and suggestions for future work.

1. Introduction

This article describes the collaborative work on applying the newly proposed ISO standard for dialogue act annotation to the Switchboard Dialogue Act (SWBD-DA) Corpus, as part of our on-going effort to promote interoperability of standardized linguistic annotations with the ultimate goal of developing shared and open language resources.

Dialogue acts (DA) 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 (Bunt et al., 2010). More recently, the emerging ISO DIS 24617-2 (2010) 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 Bunt, 2009 and 2011). In other words, an utterance in dialogue typically serves several functions. See Example (1) taken from the SWBD-DA 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.*

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

Our concern in this paper is to explore the applicability of the new ISO Standard to the existing Switchboard corpus with joint efforts of automatic and manual mapping. In the rest of the paper, we shall first describe the Switchboard Dialogue Act (SWBD-DA) Corpus and its annotation scheme (i.e. SWBD-DAMSL). We shall then describe the new ISO Standard and explain our mapping of SWBD-DAMSL to the ISO DIS 24617-2 DA tag set. In addition, machine learning techniques are employed for automatic DA classification on the basis of lexical features to evaluate the application of the new ISO DA scheme using authentic transcribed speech. We shall then introduce the user interface designed for manual mapping and explain the annotation guidelines. Finally, the paper will conclude with discussions and suggestions for future work.

2. Corpus Resource

This study uses the Switchboard Dialog Act (SWBD-DA) Corpus as the corpus resource, which is available online from the Linguistic Data Consortium [1]. The corpus

[1] http://www.ldc.upenn.edu/

Proceedings of the Workshop on Innovative Hybrid Approaches to the Processing of Textual Data (Hybrid2012), EACL 2012, pages 61–68,
Avignon, France, April 23 2012.

contains 1,155 5-minute conversations[2], 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" (Meteer and Taylor, 1995), 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 1 for the basic statistics of the SWBD-DA Corpus.

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

Table 1: Basic Statistics of the SWBD-DA Corpus

Altogether, the corpus comprises 223,606 slash-units and each is annotated for its communicative function according to a set of dialogue acts specified in the SWBD-DAMSL scheme (Jurafsky et al., 1997) and assigned a DA tag. See Example (2) taken from `sw_0002_4330.utt`, where `qy` is the DA tag for yes/no questions.

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

A total of 303 different DA tags are identified throughout the corpus, which is different from the total number of 220 tags mentioned in Jurafsky et al. (1997: 3). To ensure enough instances for the different DA tags, we also conflated the DA tags together with their secondary carat-dimensions, and yet we did not use the seven special groupings by Jurafsky et al. (1997) as we kept them as separate DA types (see Section 4 for further explanations). In the end, the 303 tags were clustered into 60 different individual communicative functions. See Table 2 for the basic statistics of the 60 DA clusters.

According to Table 2, we observe that the 60 DA clusters range from 780,570 word tokens for the top-ranking `statement-non-opinion` to only 4 word tokens for `you're-welcome`. In Table 2, the *Token %* column lists the relative importance of DA types measured as the proportion of the word tokens in the SWBD-DA corpus as whole. It can be observed that, as yet another example to illustrate the uneven use of DA types, `statement-opinion` accounts for 21.04% of the total number of word tokens in the corpus.

60 DAs	*Tokens*	*Token %*	*Cum %*
Statement-non-opinion	780,570	51.79	51.79
Statement-opinion	317,021	21.04	72.83
Segment-(multi-utterance)	135,632	9.00	81.83
Acknowledge-(backchannel)	40,696	2.70	84.53
Abandoned	35,214	2.34	86.87
Yes-no-question	34,817	2.31	89.18
Accept	20,670	1.37	90.55
Statement-expanding-y/n-answer	14,479	0.96	91.51
Wh-question	14,207	0.94	92.45
Appreciation	13,957	0.93	93.38
Declarative-yes-no-question	10,062	0.67	94.05
Conventional-closing	9,017	0.60	94.65
Quoted-material	7,591	0.50	95.15
Summarize/reformulate	6,750	0.45	95.60
Action-directive	5,860	0.39	95.99
Rhetorical-questions	5,759	0.38	96.37
Hedge	5,636	0.37	96.74
Open-question	4,884	0.32	97.06
Affirmative-non-yes-answers	4,199	0.28	97.34
Uninterpretable	4,138	0.27	97.61
Yes-answers	3,512	0.23	97.84
Completion	2,906	0.19	98.03
Hold-before-answer/agreement	2,860	0.19	98.22
Or-question	2,589	0.17	98.39
Backchannel-in-question-form	2,384	0.16	98.55
Acknowledge-answer	2,038	0.14	98.69
Negative-non-no-answers	1,828	0.12	98.81
Other-answers	1,727	0.11	98.92
No-answers	1,632	0.11	99.03
Or-clause	1,623	0.11	99.14
Other	1,578	0.10	99.24
Dispreferred-answers	1,531	0.10	99.34
Repeat-phrase	1,410	0.09	99.43
Reject	891	0.06	99.49
Transcription-errors:-slash-units	873	0.06	99.55
Declarative-wh-question	855	0.06	99.61
Signal-non-understanding	770	0.05	99.66
Self-talk	605	0.04	99.70
Offer	522	0.03	99.73
Conventional-opening	521	0.03	99.76
3rd-party-talk	458	0.03	99.79
Accept-part	399	0.03	99.82
Downplayer	341	0.02	99.84
Apology	316	0.02	99.86
Exclamation	274	0.02	99.88
Commit	267	0.02	99.90
Thanking	213	0.01	99.91
Double-quote	183	0.01	99.92
Reject-part	164	0.01	99.93
Tag-question	143	0.01	99.94
Maybe	140	0.01	99.95
Sympathy	80	0.01	99.96
Explicit-performative	78	0.01	99.97
Open-option	76	0.01	99.98
Other-forward-function	42	0.00	99.98
Correct-misspeaking	37	0.00	99.98
No-plus-expansion	26	0.00	99.98
Yes-plus-expansion	22	0.00	99.98
You're-welcome	4	0.00	99.98
Double-labels	2	0.00	100.00
Total	1,507,079	100.00	100.00

Table 2: Basic Statistics of the 60 DAs

If the cumulative proportion (*Cum%)* is considered, we

[2] Past studies (e.g. Stolcke et al., 2000; Jurafsky et al., 1997; Jurafsky et al., 1998a; Jurafsky et al., 1998b) have been focused on only 1115 conversations in the SWBD-DA Corpus as the training set. As there is no clear description which 40 conversations have been used as the testing set or for future use, we use all the 1155 conversations.

see that the top 10 DA types alone account for 93.38% of the whole corpus, suggesting again the uneven occurrence of DA types in the corpus and hence the disproportional use of communication functions in conversational discourse.

It is particularly worth mentioning that `segment-(multi-utterance)` is not really a DA type indicating communicative function and yet it is the third most frequent DA tag in SWBD-DAMSL. As a matter of fact, the SWBD-DAMSL annotation scheme contains quite a number of such non-communicative DA tags, such as `abandoned`, and `quoted-material`.

3. ISO DIS 24617-2 (2010)

A basic premise of the emerging ISO standard for dialogue act annotation, i.e., ISO DIS 24617-2 (2010), 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, it takes a so-called 'functional segment' as the unit in dialogue to be tagged with DA information, defined as a 'minimal stretch of communicative behavior that has one or more communicative functions' (Bunt et al., 2010). 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 DIS 24617-2 (2010) employs the functional segment, which serves well to emphasise the multifunctionality of dialogue utterances. An important difference here is that the ISO scheme identifies multiple DAs per segment and assigns multiple tags via the stand-off annotation mechanism.

The second difference is that each slash-unit (or utterance) in the SWBD-DA Corpus is annotated with one SWBD-DAMSL label, while each DA tag in the ISO scheme 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.

(3) A: *I'm most grateful for your help*

While the utterance in Example (3) 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"
```

4. Mapping SWBD-DAMSL to ISO

4.1 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 files. See Example (4), the tenth slash-unit taken from the file `sw_0052_4378.utt` in the folder `sw00`.

(4) sd B.7 utt1: *{C And,} {F uh,} <inhaling> we've done <sigh> lots to it. /*

The following set of files is created:

`sw00-0052-0010-B007-01.txt`	the original utterance
`sw00-0052-0010-B007-01-S.da`	SWBD-DAMSL tag

In the .txt file, there is the original utterance:

{C And,} {F uh,} <inhaling> we've done <sigh> lots to it. /

While the *-S.da file only contains the DA label: sd^t. Still another one or more files (depending on the number of dimensions) will be added to this set after converting the SWBD-DAMSL to the ISO tag sets. Take Example (4) for instance. Two more files will be created, namely,

`sw00-0052-0010-B007-01-ISO-0.da`	ISO DA tag
`sw00-0052-0010-B007-01-ISO-1.da`	ISO DA tag

The *-ISO-0.da file will contain in this case:

```
communicativeFunction = "inform"
dimension = "task"[3]
```

and the *-ISO-1.da file will contain[4]:

```
communicativeFunction = "stalling"
dimension = "timeManagement"
```

[3] The same function `Inform` have been observed to occur in different dimensions. See ISO DIS 24617-2 (2010) for detailed description.

[4] See Section 4.2 for more explanation of the multi-layer annotations in ISO standard.

4.2 Assessment of the Conversion

When mapping SWBD-DAMSL tags to functional ISO tags, it is achieved in terms of semantic contents rather than the surface labels. To be more exact, four situations were identified in the matching process.

The first is what is named as "exact matches". It is worth mentioning that since we are not matching the labels in the two annotation schemes, 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 3 lists the 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 3: Exact Matches

It can also be noted that in the previous study on the 42 DA types in SWBD-DAMSL, `open-option` (`oo`), `offer` (`co`), `commit` (`cc`) are treated as one DA type. In the current study, they are treated as individual DA types, which makes more sense especially when mapping to the ISO DA tag sets since each of them corresponds to a different ISO tag, `suggest`, `offer`, and `promise` respectively. The same is also true for the `you're-welcome` (`fw`) and `correct-misspeaking` (`bc`), which are combined together in SWBD-DAMSL and correspond to different ISO DA label.

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
Acknowledge (backchannel); Acknowledge answer; Appreciation; Sympathy; Summarize/reformulate; Repeat-phrase	AutoPositive
Accept-part; Reject-part	Correction

Table 4: Many-to-one Matches

The second situation is where more than one SWBD-DAMSL tags can be matched to the one ISO DA type, as defined as many-to-one matches. Table 4 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`).

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 the ISO scheme. 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 (5) taken from `sw_0005_4646.utt` corresponds to `Agreement` in ISO DIS 24617-2 (2010).

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

However, `accept` (`aa`) in Example (6) taken from `sw_0098_3830.utt` actually corresponds to `acceptOffer` in ISO/DIS 24617-2 (2010).

(6) co B.26 utt1: *I can tell you my last job or --/*
aa A.27 utt1: *Okay, /*

As a matter of fact, `accept` in SWBD-DAMSL may correspond to several different DAs in the ISO tag set such as:

- Agreement
- AcceptRequest (addressRequest)
- AccpetSuggestion (addressSuggestion)
- AcceptOffer (addressOffer)
- etc.

Other cases include `reject`, `action-directive` and `other answers`.

Finally, the remaining tags are unique to SWBD-DAMSL, including

- quoted material
- uninterpretable
- abandoned
- self-talk
- 3rd-party-talk
- double labels
- explicit-performative
- exclamation
- other-forward-function

It is not difficult to notice that 6 out of the 9 DA types mainly concern the marking up of other phenomena than dialogue acts. The last three unique DA types only account for a marginal portion of the whole set, about 0.03% all together (See Table 2).

In addition, multi-layer annotations of ISO can be added to the original markup of SWBD (Meteer and Taylor 1995), especially in cases such as Stalling and Self-Correction. See Example (7) taken from `sw_0052_4378.utt`.

(7) sd A.12 utt2 : *[I, + {F uh, } two months ago I] went to Massachusetts -- /*

According to Meteer and Taylor (1995), the *{F ...}* is used to mark up "filler" in utterances, which corresponds to Stalling in ISO DIS 24617-2 (2010). In addition, the markup of *[... + ...]* indicates the repairs (Meteer and Taylor, 1995), which suits well the definition of Self-correction in the ISO standard. As a result, the utterance in Example (7) is thus annotated in three dimensions:

```
communicativeFunction = "inform"
dimension = "task"

communicativeFunction = "stalling"
dimension = "timeManagement"

communicativeFunction = "self-correction"
dimension = "ownCommManagement"
```

4.3 Mapping Principles

Given the four setting of the matching, there major principles were made:

1) Cases in both "exact matches" and "many-to-one matches" can be automatically mapped to ISO tags by programming.

2) Tags that are unique to SWBD-DAMSL would not be considered at the current stage due to the absence of ISO counterparts and their marginal proportion.

3) Cases in "one-to-many matches" are more complex and call for manual mapping, which will be further discussed in Section 6.

4) Different DA dimensions will be also automatically added accordingly to each utterance in the format of stand-off annotation.

5. Application Verification

To evaluate the applicability of mapping SWBD-DAMSL tag set to the new ISO standard (ISO DIS 24617-2, 2010), machine learning techniques are employed, based on the preliminary results from the automatic mapping, to see how well the SWBD-ISO DA tags can be automatically identified and classified based on lexical features. The result is also compared with that obtained from the Top-15 SWBD-DAMSL tags. It will be particularly interesting to find out whether the emerging ISO DA annotation standard will produce better automatic prediction accuracy. In this paper, we evaluate the performance of automatic DA classification in the two DA annotation schemes by employing the unigrams as the feature set.

Two classification tasks were then identified according to the two DA annotation schemes. Task 1 is to automatically classify the DA types in the SWBD-DAMSL. Based on the observations mentioned above, it was decided to use the top 15 DA types to investigate the distribution of word types in order to ascertain the lexical characteristics of DAs. Furthermore, since `segment-(multi-utterance)`, `abandoned`, and `quoted-material` do not relate to dialogue acts per se, these three were replaced with `rhetorical-questions`, `open-question` and `affirmative-non-yes-answers`. We thus derive Table 6 below, showing that the revised list of top 15 DA types account for 85.13% of the SWBD corpus. The DA types are arranged according to *Token%* in descending order.

Top-15 SWBD-DAMSL DAs	*Tokens*	*Token %*	*Cum %*
Statement-non-opinion	780,570	51.79	51.79
Statement-opinion	317,021	21.04	72.83
Acknowledge-(backchannel)	40,696	2.70	75.53
Yes-no-question	34,817	2.31	77.84
Accept	20,670	1.37	79.21
Statement-expanding-y/n-answer	14,479	0.96	80.17
Wh-question	14,207	0.94	81.11
Appreciation	13,957	0.93	82.04
Declarative-yes-no-question	10,062	0.67	82.71
Conventional-closing	9,017	0.60	83.31
Summarize/reformulate	6,750	0.45	83.76
Action-directive	5,860	0.39	84.15
Rhetorical-questions	5,759	0.38	84.53
Open-question	4,884	0.32	84.85
Affirmative-non-yes-answers	4,199	0.28	85.13
Total	1,282,948	85.13	

Table 6: Top-15 SWBD-DAMSL DA types

Next, accordingly, task 2 is to classify the top 15 ISO DAs based on the results from the automatic mapping. It should be pointed out that only one layer of annotation in the ISO DA tags is considered in order to make the result comparable to that from SWBD-DAMSL, and the dimension of `task` is the priority when it comes to multi-layer annotations.

Top-15 SWBD-ISO DAs	*Tokens*	*Token %*	*Cum %*
Inform	1,117,829	74.17	74.17
AutoPositive	64,851	4.30	78.47
PropositionalQuestion	37,201	2.47	80.94
SetQuestion	15,062	1.00	81.94
Answer	11,171	0.74	82.68
CheckQuestion	10,062	0.67	83.35
InitialGoodbye	9,017	0.60	83.95
Question	4,884	0.32	84.27
ChoiceQuestion	4,212	0.28	84.55
Completion	2,906	0.19	84.75
Stalling	2,860	0.19	84.94
Disconfirm	1,531	0.10	85.04
AutoNegative	770	0.05	85.09
Offer	522	0.03	85.12
AcceptApology	341	0.02	85.15
Total	1,283,219	85.15	

Table 7: Top-15 SWBD-ISO DA types

The Naïve Bayes Multinomial classifier was employed, which is available from Waikato Environment for Knowledge Analysis, known as Weka (Hall et al., 2009). 10-fold cross validation was performed and the

results evaluated in terms of precision, recall and F-score (*F1*).

Table 8 presents the results for classification task 1. The SWBD-DAMSL DAs are arranged according to F-score in descending order.

Top 15 SWBD-DAMSL DAs	*Precision*	*Recall*	*F1*
Acknowledge-(backchannel)	0.821	0.968	0.888
Statement-non-opinion	0.732	0.862	0.792
Appreciation	0.859	0.541	0.664
Statement-opinion	0.538	0.584	0.560
Conventional-closing	0.980	0.384	0.552
Accept	0.717	0.246	0.367
Yes-no-question	0.644	0.204	0.309
Wh-question	0.760	0.189	0.303
Open-question	0.932	0.084	0.154
Action-directive	1.000	0.007	0.013
Statement-expanding-y/n-answer	0.017	0	0.001
Declarative-yes-no-question	0	0	0
Summarize/reformulate	0	0	0
Rhetorical-questions	0	0	0
Affirmative-non-yes-answers	0	0	0
Weighted Average	0.704	0.725	0.692

Table 8: Results from Task 1

As can be noted, the weighted average F-score is 69.2%. To be more specific, `acknowledge-(backchannel)` achieves the best F-score of 0.888, followed by `statement-non-opinion` with an F-score of 0.792. Surprisingly, the `action-directive` has the highest precision of 100%, but has the second lowest recall of over 0.7%. It can also be noted that the last four types of DAs cannot be classified with the F-score of 0%.

Top 15 SWBD-ISO DAs	*Precision*	*Recall*	*F1*
Inform	0.879	0.987	0.930
Answer	0.782	0.767	0.775
AutoPositive	0.711	0.507	0.592
InitialGoodbye	0.972	0.351	0.516
PropositionalQuestion	0.521	0.143	0.224
SetQuestion	0.668	0.120	0.203
Question	0.854	0.051	0.097
AutoNegative	0.889	0.026	0.051
ChoiceQuestion	0.286	0.008	0.015
Stalling	0.400	0.003	0.007
CheckQuestion	0.042	0.001	0.001
AcceptApology	0	0	0
Completion	0	0	0
Disconfirm	0	0	0
Offer	0	0	0
Weighted Average	0.832	0.865	0.831

Table 9: Results from Task 2

Table 9 presents the results for classification task 2. The DAs are arranged according to F-score in descending order. As can be noted, the weighted average F-score is 83.1%, over 10% higher than task 1. To be more specific, `Inform` achieves the best F-score of 0.93, followed by `Answer` with an F-score of 0.775. The DA `InitialGoodbye` has the highest precision, of about 97%, whereas `Inform` has the highest recall of over 98%. Similar to the results obtained in Task 1, the last four types of DAs in Task 2 also cannot be classified with the F-score of 0%.

Meanwhile, as mentioned earlier, when the data size for each DA type is taken into consideration, Task 2 may be more challenging than Task 1 in that 6 out of the 15 SWBD-ISO DA types has a total number of word tokens fewer than 4,000 whereas all the 15 SWBD-DAMSL DA types has a total number of over 4,000. Therefore, the much higher average F-score suggests that the application of ISO standard DA scheme could lead to better classification performance, suggesting that the ISO DA standard represents a better option for automatic DA classification.

To sum up, with a comparable version of the SWBD-DA Corpus, results from the automatic DA classification tasks show that the ISO DA annotation scheme produces better automatic prediction accuracy, which encourages the completion of the manual mapping.

6. Manual Mapping

6.1 Analysis of Problematic DA Types

As mentioned earlier, there are mainly four problematic SWBD-DAMSL tags, namely, `accept` (`aa`), `reject` (`ar`), `action-directive` (`ad`) and `other answers` (`no`). They are problematic in that they carry a broad function applicable to a range of different situations according to the new ISO standard, as evidenced in the case of `accept` discussed in Section 4.2. Consequently, to map the problematic SWBD-DAMSL tags to the ISO tags calls for manual manipulation.

A close look into those four types shows that the mapping could be further divided into two setting. Again, take `accept` (`aa`) for example. In the first setting, a sub-division of `accept` (`aa`) can also be automatically matched according to the previous utterance by the other speaker in the adjacent pair. See Example (8) taken from `sw_0001_4325.utt`.

(8) sv A.49 utt3: *take a long time to find the right place /*
x A.49 utt4: *<laughter>.*
aa B.50 utt1: *Yeah, /*

Here `accept` (`aa`) corresponds to `Agreement` because of the DA type in `A.49 utt3` but not the immediate previous DA as in `A.49 utt4`. With this principle, the particular sub-groups for automatic mapping were identified for `accept` (`aa`). See Table 10.

SWBD-DAMSL		*ISO*
Previous DA	*Current DA*	
Statement-non-opinion; Statement-opinion; Hedge Rhetorical-question; Statement expanding y/n answer,	accept	Agreement
Offer		AcceptOffer
Open-option		AcceptRequest
Thanking		AcceptThanking
Apology		AcceptApology

Table 10: Sub-groups of `accept` for Auto Mapping

The remaining cases, in the second setting, call for manual annotation. For instance, when the previous DA type is also a problematic one, annotators need to decide

the corresponding ISO DA tag for the previous SWBD-DAMSL one before converting the `accept` (`aa`). See Example (9) taken from `sw_0423_3325.utt`.

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

Here, `action-directive` (`ad`) is first decided as a suggestion, and therefore `accept` (`aa`) turns out to actually correspond to `acceptSuggestion` (`addressSuggestion`) in ISO/DIS 24617-2 (2010).

6.2 Design of a User Interface

Given the analysis of those four DA tags, a user-friendly interface was then designed to assist annotators to maximize the inter-annotator agreement. See Figure 1.

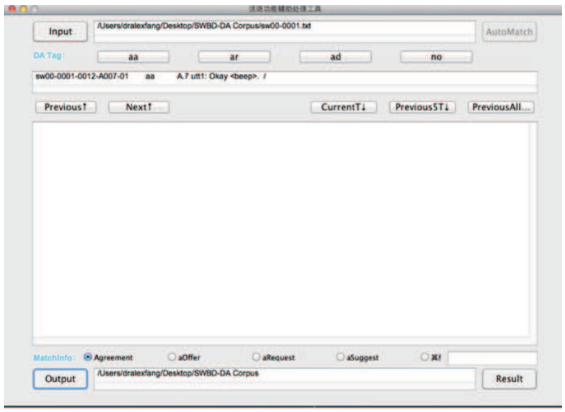

Figure 1: User Interface

Figure 1 shows the screenshot when the targeted SWBD-DAMSL type is `accept` (`aa`). As can be noted above, the basic functional bars have been designed, including:

- *Input*: the path of the input
- *Automatch*: to filter out the sub-groups that can be automatically matched
- *DA Tag*: the targeted problematic DAs, namely,
  - `aa` (`accept`)
  - `ar` (`reject`)
  - `ad` (`action-directive`) and
  - `no` (`other answers`)
- *Previous*: to go back to the previous instance of the targeted DA type
- *Next*: to move on to the next instance of the targeted DA type
- *Current*: the extraction of the adjacent turns
- Previous5T: the extraction of the previous five turns when necessary
- *PreviousAll*: the extraction of all the previous turns when necessary
- *MatchInfo*: Bars for mapping information with five options:
  - Four pre-defined ISO DA types
  - Other: a user-defined mapping with a two-fold function: for user defined ISO DA type and for extra pre-defined ISO DA types (since the pre-defined DA types differ for the four targeted SWBD-DAMSL types).
- *Output*: the path of the output
- *Result*: export the results to the chosen path

With this computer-aided interface, three annotators are invited to carry out the manual mapping. They are all postgraduates with linguistic background. After a month of training on the understanding of the two annotation schemes (in process), they will work on the SWBD-DAMSL DA instances from 115 randomly chosen files, and map them into ISO DA tags independently. The kappa value will be calculated to measure the inter-annotator agreement.

7. Conclusion

In this paper, we reported our efforts in applying the ISO-standardized dialogue act annotations to the Switchboard Dialogue Act (SWBD-DA) Corpus. In particular, the SWBD-DAMSL tags employed in the SWBD-DA Corpus were analyzed and mapped onto the ISO DA tag set (ISO DIS 24617-2 2010) according to their communicative functions and semantic contents. Such a conversion is a collaborative process involving both automatic mapping and manual manipulation. With the results from the automatic mapping, machine learning techniques were employed to evaluate the applicability of the new ISO standard for dialogue act annotation in practice. With the encouraging results from the evaluation, the manual mapping was carried out. A user-friendly interface was designed to assist annotators. The immediate future work would be finish the manual mapping and thus to produce a comparable version of the SWBD-DA Corpus was produced so that the two annotation schemes (i.e. SWBD-DAMSL vs. SWBD-ISO) can be effectively compared on the basis of empirical data. Furthermore, with the newly built resource, i.e., SWBD-ISO, 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.

8. Acknowledgements

Research described in this article was supported in part by grants received from City University of Hong Kong (Project Nos 7008002, 9610188, 7008062 and 6454005). It was also partially supported by the General Research Fund of the Research Grants Council of Hong Kong (Project No 142711).

9. References

Bunt, H. (2009). Multifunctionality and multidimensional dialogue semantics. In *Proceedings of DiaHolmia Workshop on the Semantics and Pragmatics of*

Dialogue, Stockholm, 2009.

Bunt, H. (2011). Multifunctionality in dialogue and its interpretation. *Computer, Speech and Language,* 25 (2), pp. 225--245.

Bunt, H., Alexandersson, J., Carletta, J., Choe, J.-W., Fang, A.C., 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 *Proceedings of the Seventh International Conference on Language Resources and Evaluation*. Valletta, MALTA, 17-23 May 2010.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explorations*, 11 (1), pp. 10--18.

ISO DIS 24617-2. (2010). *Language resource management – Semantic annotation framework (SemAF), Part 2: Dialogue acts*. ISO, Geneva, January 2010.

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

Jurafsky, D., Bates, R., Coccaro, N., Martin, R., Meteer, M., Ries, K., Shriberg, E., Stolcke, A., Taylor, P. and Ess-Dykema, C. V. (1998a). Switchbaod Discourse Language Modeling Project and Report. *Research Note 30, Center for Language and Speech Processing*, Johns Hopkins University, Baltimore, MD, January.

Jurafsky, D., Shriberg, E., Fox B. and Curl, T. (1998b). Lexical, prosodic, and syntactic cues for dialog acts. *ACL/COLING-98 Workshop on Discourse Relations and Discourse Markers*.

Meeter, M., Taylor, A. (1995). Dysfluency annotation stylebook for the Switchboard Corpus. Available at ftp://ftp.cis.upenn.edu/pub/treebank/swbd/doc/DFL-book.ps.

Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurfsky, D., Taylor, P., Martin, R., Ess-Dykema, C.V. and Meteer, M. (2000). Dialogue Act Modeling for Automatic Tagging and Recognition of Conversational Speech. *Computational Linguistics*, 26 (3), pp. 339--373.