

# Issues in the addition of ISO standard annotations to the Switchboard corpus

<b>Harry Bunt</b>	<b>Alex C. Fang and Xiaoyue Liu</b>	<b>Jing Cao</b>	<b>Volha Petukhova</b>
TiCC, Tilburg Center for Cognition and Communication Tilburg University Tilburg, The Netherlands harry.bunt@uvt.nl	City University of Hong Kong Department of Chinese, Translation, and Linguistics Hong Kong SAR a.c.fang@cityu.edu.hk xyliu@cityu.edu.hk	College of Foreign Languages Zhongnan University of Economics and Law Wuhan, China c-jinhk@yahoo.cn	Department of Spoken Language Systems Saarland University Saarbrücken, Germany v.petukhova@lsv. uni-saarland.de

## Abstract

This paper analyzes the issues that arise when trying to add annotations to the dialogues in the Switchboard corpus according to ISO standard 24617-2, exploiting the existing SWBD-DAMSL annotations. These issues relate to differences between the two tag sets; to the highly multidimensional view that underlies the ISO standard; to differences in segmenting the dialogues into functional units; to the use of in-line markups for certain phenomena in Switchboard, and to the use of intra-dialogue dependence relations as defined in the ISO standard.

The analysis is supplemented by a discussion of how the existing annotations may be helpful to semi-automatically create a fully-fledged ISO standard annotation alongside the existing SWBD-DAMSL annotation.

## 1 Introduction

In September 2013 the International Organisation for Standardisation ISO published the international standard 24617-2<sup>1</sup>, a comprehensive application-independent scheme for dialogue act annotation that is both empirically and theoretically well-founded, that can deal with typed, spoken, and multimodal dialogue, and that can be used effectively by human annotators and by automatic annotation methods.

With the aim of building a large corpus of dialogues, annotated according to this standard, an effort was initiated to create ISO 24617-2 annotations for the dialogues in the Switchboard corpus, which forms a valuable resource for the study of spoken dialogue.<sup>2</sup> In particular, this effort ex-

ploits the similarities between the ISO 24617-2 and the SWBD-DAMSL scheme (Jurafsky et al., 1997) by semi-automatically converting SWBD-DAMSL annotations into ISO 24617-2 annotations where possible. An additional benefit of this approach is that it allows an in-depth comparison between the two annotation schemes.

Fang et al. (2011) have described initial explorations in this project, and Fang et al. (2012) have described the possibilities and limitations of automatically converting SWBD-DAMSL tags to ISO 24617-2 tags. This paper deals with other issues, relating in particular to (1) the highly multidimensional approach to annotation that underlies the ISO standard more clearly than the annotations in the Switchboard corpus; (2) the segmentation of the Switchboard dialogues into ‘slash-units’ rather than into ‘functional segments’, as the ISO standard requires; (3) the use of certain in-line markups and tagging of non-functional phenomena in the Switchboard dialogues; and (4) the annotation of dependence relations between units in a dialogue according to the ISO standard.

Example (1), showing a small dialogue fragment (from Switchboard dialogue sw01-0105), as marked up in the Switchboard corpus and as annotated according to ISO 24617-2, illustrates some of the differences between the two approaches.

- (1) a. (dialogue sw01-0105 lines 0007-0008)  
A003: qw^d {D So } when you say the  
morning news, or evening news  
or national news is when? /  
B004: sd {F Uh, } evening news at six  
thirty I believe /
- b. ISO-24617-2 segmentation:  
fs1 = So  
fs2 = when you say the morning  
news, or evening news, or  
national news, is when?

<sup>1</sup>See the official description of the standard in ISO 24617-2:2013, and summary descriptions in Bunt et al. (2010; 2012).

<sup>2</sup>The Switchboard Dialogue Act Corpus is distributed by LDC.

fs3 = Uh  
 fs4 = evening news is at six thirty  
 I believe /

c. ISO 24617-2 annotation:

```
<diaml xmlns:
  "http://www.iso.org/diaml/" />
<dialogueAct xml:id="a1"
  target="#fs1"
  sender="#a" addressee="#b"
  communicativeFunction=
    "turnTake"
  dimension="turnManagement" />
<dialogueAct xml:id="a2"
  target="#fs1"
  sender="#a" addressee="#b"
  communicativeFunction=
    "stalling"
  dimension="timeManagement" />
<dialogueAct xml:id="a3"
  target="#fs2"
  sender="#a" addressee="#b"
  communicativeFunction=
    "propositionalQuestion"
  dimension="task" />
<dialogueAct xml:id="a4"
  target="#fs3"
  sender="#b" addressee="#a"
  communicativeFunction=
    "stalling"
  dimension="timeManagement" />
<dialogueAct xml:id="a5"
  target="#fs3"
  sender="#b" addressee="#a"
  comm.Function="turnTake"
  dimension="turnManagement" />
<dialogueAct xml:id="a6"
  target="#fs4"
  sender="#b" addressee="#a"
  communicativeFunction=
    "answer"
  certainty="uncertain"
  dimension="task"
  functionalDependence="#a3" />
</diaml>
```

SWBD-DAMSL annotations and ISO 24617-2 annotations clearly use very different representation formats. SWBD-DAMSL makes use of functional tags like  $qw^d$  (which stands for “Declarative Wh-Question”) and  $sd$  (for “Statement non-opinion”), in the form of strings attached to stretches of text delineated by “/”, so-called “slash-units” (see Section 2.1). Other information is encoded as in-line markups, such as in (1a) a discourse marker by ‘{D So }’ and a filled pause by ‘{F Uh }’; and the identity of the speaker is encoded in line numbers like ‘A003’ and ‘B004’.

ISO standard annotations represent all the information in the form of XML-expressions, making use of the XML-based annotation language DiAML (Dialogue Act Markup Language) which is defined as part of the standard. These annotations are in stand-off form, with an attribute @target whose value identifies the stretch of dialogue that the annotation applies to (a ‘functional segment’, see Section 2.1). The annotations in DiAML include not only an identification of the speaker, as in Switchboard, but also of one or more addressees (the attribute @addressee may have multiple values); a specification not only of the communicative function of a dialogue act expressed by the functional segment but also of the communicative dimension that the act belongs to (such as the task that motivates the dialogue, the dimension of turn-taking, or the dimension of time management)<sup>3</sup>; and an indication of relations among dialogue acts, in this example an indication of the question that is answered by an Answer act.

An analysis of the similarities and differences between the SWBD-DAMSL and ISO 24617-2 tag sets in Fang et al. (2011; 2012) shows that 14 of the SWBD-DAMSL tags exactly match an ISO 24617-2 communicative function tag, and 27 SWBD-DAMSL tags correspond to 9 ISO standard tags. The latter is due to the fact that SWBD-DAMSL sometimes makes distinctions which are not motivated semantically but syntactically or lexically; for example, the tags Yes-answer, Affirmative non-yes answer, No-answer, and Negative non-no answer all correspond to the single ISO tag Answer. In the case of exact matches and many-to-one matches, the conversion from SWBD-DAMSL tags to ISO communicative function tags can be done automatically; Fang et al. (2012) report that this can be done for 187,768 of the 223,606 units annotated in the Switchboard corpus, which amounts to 84,0% of the corpus.

Replacing SWBD-DAMSL tags by ISO communicative function tags does not create full ISO standard annotations, however, as example (1) showed; not only do we have to replace the tags  $qw^d$  and  $sd$  by the appropriate ISO tags (Set-Question and Inform, respectively) but we also have to consider (1) for each communicative

<sup>3</sup>The ISO standard distinguishes nine dimensions: Task, Turn Management, Time Management, Auto-Feedback, Allo-Feedback, Own Communication Management, Partner Communication Management, Discourse Structuring, and Social Obligations Management. For definitions see Bunt (2009).

function the dimension in which it is used; (2) the addition of communicative functions in those dimensions where SWBD-DAMSL doesn't have any, such as turn management; (3) what to do with the in-line markup of discourse connectives like and filled pauses; (4) how to produce the ISO qualifiers, like `certainty="uncertain"` and relations between dialogue acts, like `functionalDependence="#a3"`.

This paper is structured as follows. Section 2 discusses issues relating to the segmentation of dialogues into meaningful units. Section 3 discusses the annotation of in-line markups. Section 4 discusses the treatment of some phenomena that are not annotated in Switchboard. The concluding Section 6 summarizes the analysis of the main issues involved in adding ISO standard annotations to the Switchboard corpus, and indicates for each of these issues how the additions could be made, exploiting the existing SWBD-DAMSL annotations and the in-line markups of various phenomena.

## 2 Segmentation

### 2.1 Slash units versus functional segments

The annotations in the Switchboard corpus make use of a segmentation of dialogues into so-called 'slash units', defined by Meteer & Taylor (1995), as "*Maximally a sentence, but possibly a smaller unit. Intuitively, slash-units below the sentence level correspond to those parts of the narrative which are not sentential but which the annotator interprets as complete*". Slash units are allowed to span (parts of) multiple turns by the same speaker, separated by a contribution from another speaker, and in that sense to be discontinuous, as in the following example (from Core & Allen, 1997):

- (2) u: take the product to  
 s: yes?  
 u: to Corning

The Switchboard segmentation follows the strategy for dialogue annotation with DAMSL tags described by Core and Allen (1997), who call these units 'utterances'. Utterances are allowed to be discontinuous only in case of an interruption by another speaker, as in (2), and are not allowed to overlap with other units. Disfluencies such as hesitations (like *uh* or *um*), and restarts like *I mean*, are thus not treated as units with a communicative function. With reference to the repair in ex-

ample (3), Core and Allen (1997) note that they do not view *Tuesday I mean Friday* as a functional unit, since that "*would mean cutting off "Friday" from "we'll go Tuesday"*". DAMSL is not designed for annotating speech repairs, reference, or other intra-clause relations so we decided to use a simple definition of an utterance that leaves out such phenomena".

- (3) we'll go Tuesday I mean Friday

This strategy is clearly inadequate for annotating phenomena of own communication management and time management. ISO 24617-2 supports the annotation of communicative functions in these dimensions, in view of the frequent occurrence of stallings and self-corrections in spontaneous speech, and takes over the approach to segmentation developed for dialogue analysis using the DIT<sup>++</sup> annotation scheme (Bunt, 2009). This approach defines a *functional unit* as a minimal stretch of communicative behaviour that has a communicative function (and possibly more than one function) (Geertzen et al., 2007). Utterance (3) would be segmented as shown in (4), where the parts in boldface form the discontinuous segment *we'll go Friday*, expressing an inform act, and the underlined part *Tuesday I mean Friday* forms an overlapping functional segment that expresses a self-correction.

- (4) **we'll go** Tuesday I mean Friday

A disadvantage of treating an entire utterance like (3) as a single unit, is that any self-correction which it contains is associated with the entire utterance, which is not accurate. This causes a serious problem when a slash unit contains more than one stalling or self-correction, since the annotation cannot distinguish between these. For example, in (5) (from Switchboard dialogue sw00-0004, line 30) a stalling is expressed by the filled pauses {F uh, } {F uh, } and another one by the repetition [ to the, + to the].

- (5) you wouldn't have this {F uh, } {F uh, } theatrics where the lawyer jumps up and presents it [ to the, + to the] jury /

Some 25-30% of the slash units in the Switchboard corpus contain a stalling or a self-correction, and an estimated 6% more than one of these, so the inability to correctly annotate these is a serious limitation. The in-line markup indicates each

filled pause, but does not assign an interpretation to it. The annotation of disfluencies is discussed further in Section 3.

## 2.2 Mono- versus multifunctionality

The annotations in the Switchboard corpus are monofunctional, in the sense that only one SWBD-DAMSL tag is assigned to each slash unit. There are only a few cases in the corpus where more than one tag has been assigned; see the examples in (6):

- (6) a. (dialogue sw07-0701 line 0161-B108-03)  
 B: school's very important I'm an educator myself and my wife teaches/
- b. (dialogue sw07-0701 line 0083-B060-05)  
 B: {C but } I think that if you did something, for example, to an individual and caused them to lose the ability to earn a living, I remember a man drove by randomly shot a woman in the head while she was driving –
- c. (dialogue sw07-0701 line 0716-A107-01)  
 # I think this giving excuses # is pretty prevalent, {F uh, } [yo-, + ] I work in the school district /

These cases all seem to involve segmentation problems: in (6b), when the speaker says *I remember* it seems that a new thought is starting, which would plausibly correspond to the start of a new slash unit; in the other two cases it would seem preferable to segment into a sequence of two slash units, in case (6c) rather fairly signaled by the hesitation {F uh, } and the restart [ yo-, + ].

The SWBD-DAMSL tags are composite, and have been characterized as ‘tag clusters’ (Jurafsky et al., 1997), but different from the composite tags introduced by Popescu-Belis (2008) they do not represent dialogue act combinations. For example, the tag  $q_w \hat{d}$  can be decomposed into  $q$  for question,  $w$  for WH-, and  $\hat{d}$  for declarative, but only the sub-tag  $q_w$  denotes a communicative function.

The ISO standard is intended to be used for annotating all the communicative functions of dialogue units. The slash units in the Switchboard corpus on average have 1.8 communicative functions, and 62% of the slash units has two or more communicative functions. This means, for the creation of fully-fledged ISO 24617-2 annotations, that in addition to the function tags which can

be obtained through the conversion of SWBD-DAMSL tags, further functional tags have to be generated through a more comprehensive interpretation of the dialogues. This is partly possible by interpreting the in-line mark up of certain dialogue phenomena, as discussed in the next section.

## 3 Interpreting in-line markups

The Switchboard dialogues include the in-line markup (often occurring within slash units) of the following types of disfluencies:

1. restarts, marked up [ X + Y ] (more detail below);
2. filled pauses, marked up { F ... };
3. explicit editing terms, marked up { E ... };
4. discourse connectives and discourse markers, marked up { C } and { D }.

Asides, such as self-talk and third-person talk, are marked up by means of SWBD-DAMSL tags and are also considered in this Section.

### 3.1 Restarts and repairs

Following Shriberg (1994), restarts are expressions of the form shown in (7), in which the part RM is called the ‘reparandum’, a stretch of text to be replaced; RR is the replacing material, and IM is intermediate material (such as a filled pause or an editing term), that separates the two and typically signals that a replacement is going to follow. This is marked up in the Switchboard corpus as shown in (7a).

- a. Show me flights [ from Boston on +  
 RM  
 {F uh } from Denver on ] Monday  
 IM RR
- b. **Show me flights** from Boston on uh  
from Denver on Monday

The ISO 24617-2 annotation makes use of the segmentation shown in (7b), consisting of the segment **Show me flights from Denver on Monday** in the Task dimension, expressing a request; and the segment from Boston on uh from Denver on in the Own Communication Management dimension, expressing a self-correction.

The description of a restart in terms of a ‘reparandum’ (RM) and replacing material (RR) strongly suggests that restarts are self-corrections.

This is not always correct, however, since the RM part of a restart may be identical to the reparandum, in which case we have a repetition rather than a replacement, and it may be empty; in both cases the ISO 24617-2 definition of a self-correction would not apply. Repetitions often do not signal that the speaker wants to correct what he just said, but rather that he hasn't quite made up his mind yet as to how he wants to express himself, which makes this behaviour a case of stalling, rather than a case of self-correction. In cases where the RR part is empty, the speaker decides not to go on saying what he started to say; this corresponds to what in ISO 24617-2 is called a *retraction*.

Even if the RR part of a restart is not empty and not identical to the reparandum, we do not necessarily have a self-correction, as the examples in (8) show (from sw00-0004, lines 68 and 25, respectively):

- (8) a. .. to begin with, [ you would -, + you would have, ] -  
 b. .. if they did it [ with the + {F uh } just with the ] judges, the police have to do..

In such cases, where the reparandum re-appears in the RR part, Meteer & Taylor (1995) speak of an 'insertion'. An insertion has one of the following two forms, where XM denotes the inserted material (and IM may be empty):

- (9) a. ... RM IM RM XM ...  
 b. ... RM IM XM RM ...

In an insertion of the form (9a) the speaker does not so much correct himself; the repetition of the reparandum rather seems to indicate that the speaker needs some time to decide to say what he already started to say; that makes this behaviour a case of stalling rather than self-correction.

The following guideline can be formulated for interpreting the markings of restarts in the Switchboard corpus in terms of the ISO standard :

- if the RR part is empty, then the marked up segment is a Retraction;
- if the RR part is of the form RM XM, then the marked up segment is a Stalling;
- if the RR part is not empty and not of the form RM XM, then the marked up segment is a Self-Correction.

Note that this is no more than a guideline; each individual case has to be inspected in order to be certain about the correct interpretation in the given context.

### 3.2 Filled pauses

Filled pauses typically signal that the speaker needs a little time to decide how to continue his contribution, and are annotated according to ISO 24617-2 as stalling acts. (The ISO tag 'pausing' is used for those cases where the speaker temporarily suspends the dialogue, as in *just a moment*).

Stalling acts occurring at the beginning of a turn (like *um*, or *well*,) additionally signal that the participant takes the turn; those occurring at the beginning of a slash unit but not at the beginning of a turn (*and* occurs frequently in that position) often indicate that the speaker wants to keep the turn. See further Section 5.2.

Filled pauses may also be indicators of Own Communication Management acts, viz. retractions and self-corrections, or of struggling to find the right words for something and eliciting a collaborative completion (an act in the Partner Communication Management dimension).

### 3.3 Explicit editing terms

Explicit editing terms are marked up in the Switchboard corpus as { E .... } and often occur after the reparandum part of a restart. In ISO 24617-2 explicit editing terms are regarded as indicators of Own Communication Management acts (a Retraction or as a Self-Correction), or of a Partner Communication Management act (eliciting help); see also Section 3.1.

### 3.4 Discourse markers

A distinction is made in the Switchboard corpus between 'discourse markers', such as 'Well' and 'So', indicated by { D ... }, and 'coordinating conjunctions', such as *and*, *but*, and *because*, marked up by { C ... }. In the literature the term 'discourse marker' is commonly understood to include coordinating conjunctions (at utterance level, rather than propositional level), and we follow this convention in this paper.

Discourse markers are important for segmenting a dialogue into meaningful units, since they very often 'bracket' functional segments, and they may also be functional segments on their own. With reference to the AMI corpus, Petukhova & Bunt (2009) have shown that discourse markers

are nearly always multifunctional. The most frequently occurring discourse marker, *and*, has an average of 2.6 communicative functions; other frequent ones are *so* (average multifunctionality 2.0); *well* (2.1); *but* (1.9); and *because* (1.2). *And* is used 57% of the time with a small pause as a speaker continuation signal, i.e. as a turn-keeping act; *well* is mostly a turn-taking signal (also mostly with a small pause); *so*, used with or without a small pause, can be both. An example of the characteristic use of *and* is shown in (10), with durations of micro-pauses:

- (10) like you said a problem was how many components are in there  
 (0.28) {C and } (0.12) the power is basically a factor of that  
 (0.55) {F um } (0.47) {C and } (0.32) this affects you in terms of the size of your device  
 (0.52) {F um } (0.26) {C and } (0.16) that would have some impact

The importance of discourse markers for segmentation is evidenced in the Switchboard corpus by the fact that an estimated 35% of all slash units begin with a discourse marker. As a discourse marker (rather than a propositional connective), *and* occurs almost exclusively at the start of a slash unit inside a turn; *well* typically occurs in turn-initial position and has a turn-taking or turn-accepting function, as illustrated in (11) (from sw03-0304 line 0087-A049-01).

- B: {C So } [ what do you, + what kind of hobbies are you ] in?/  
 (11) A: {C Well, } I'm a mother of four, /

Discourse markers may also have a feedback function, a time management function, or a discourse structuring function. Clark and Shaefer (1989) and Clark (1996) claim that *and* has an important feedback function; this claim is not supported by the Switchboard data, where *and* occurs predominantly inside a speaker turn, whereas feedback tends to be expressed at the beginning of a turn.

The markup of discourse markers in the Switchboard corpus is useful for the recognition of slash units; to correctly annotate discourse markers that by themselves have one or more communicative functions according to the ISO 24617-2 standard, a resegmentation is required that treats such occurrences of discourse markers as separate slash units.

### 3.5 Asides

Asides do have a communicative function, but in a sense do not belong to the dialogue, as (12) illustrates. In the Switchboard corpus, asides like the one in (12) (from sw03-0304, lines 180-A099-01 through 184- A101-02) are annotated with the non-communicative tag 't3' (third-party talk).

- (12) A: I keep hearing these marvelous things –  
 B: Yeah, /  
 B: haven't either. /  
 A: – about Dear Valley and,  
 A: {F um, } <to child> {A don't, Adam, } ...

Since an aside typically expresses a dialogue act, it could be annotated with the appropriate communicative function tag(s); moreover, ISO standard annotation includes indicating for each dialogue act the identities of the speaker and the addressee(s); in an aside like the one in the bottom line in (12) (sw03-0304 line 184), this is possible if the addressee ('Adam') has been introduced in the metadata as one of the participants in the communicative situation.

As for the conversion of Switchboard annotations to the ISO standard, all cases labelled t3 have to be re-annotated, taking their context of occurrence into account.

## 4 Phenomena not annotated in Switchboard

### 4.1 Nonverbal behaviour

Nonverbal behaviour is marked up in-line in Switchboard transcriptions with pointed brackets, and when it occurs as a separate turn it is annotated (even though it is not considered as a slash unit) with the SWBD-DAMSL tag 'x'. An example is seen in the second line of (13) (from dialogue sw03-0304):

- (13) sd A: {C so } basically I'm just, <laughter>/  
 x B: <laughter>

While marked as being a stretch of nonverbal behaviour, no functional annotation is associated with nonverbal behaviour in the Switchboard corpus, as illustrated by (13) and (14) (line 0014 from dialogue sw00-0004).

- sv I think what they need to do is, they  
 (14) need to somehow <lipsmack> take the money out of it. /

The ISO standard makes use of nonverbal and multimodal functional segments besides purely verbal segments (see Petukhova and Bunt, 2012), and supports the functional annotation of such segments

Laughter often expresses a positive emphatic sentiment concerning something that another participant just said, and thus indicates that the laughing participant understood what was said. The appropriate functional ISO tag is thus Auto-Positive (in the Auto-Feedback dimension).

Example (14) would be treated in ISO 24617-2 by distinguishing the discontinuous verbal segment *I think what they need to do is, they need to somehow take the money out of it*, which would be annotated as having the communicative function Inform, and the vocal functional segment defined by its begin and end point being just after the end of *somehow* and before the start of *take*; this segment would be annotated as having a Stalling function. (See ISO 2461702:2012, Annex D, and Petukhova & Bunt, 2012 for more details.)

While <laughter> and <lipsmack> can mostly be mapped to the ISO function tags Auto-Feedback and Stalling (although each occurrence has to be checked for its function in the context in which it occurs), the addition of these annotations to the Switchboard corpus would require a resegmentation of the dialogues, using functional segments rather than slash units.

## 4.2 Turn Management

Turn management functions are not annotated in (SWBD-)DAMSL. In the ISO standard they are, the guidelines instructing the annotation of communicative behaviour with turn management functions if and only if a dialogue participant explicitly signals the wish to have or keep the speaker role, or to release it or to give it to another participant. The background of this guideline is that speakers often take the turn simply by starting to speak, like participant B in dialogue fragment (15):

- A: Anyone wants to add something?  
 (15) B: I would like to add that the controls should be really easy to use.

Any time a dialogue participant (B) starts to speak after another participant (A) has ceased to speak, he (B) can be said to perform a turn-taking (or turn-accepting) act *by implication* of performing a dialogue act which is expressed by what he

(B) says.<sup>4</sup> For dialogue act annotation, more interesting are those cases where a speaker *explicitly* indicates that he wants to take on the speaker role, for example by starting to speak without producing any content, such as a filler (*Um...*) or a discourse marker (e.g. *Well...* or *You know...*). The Switchboard examples in (16) illustrate this.

- (16) a. (dialogue sw01-0105 lines 01-02)  
 A: Jimmy, {D so } how do you get most of your news?/  
 B: {D Well, } [ I kind of, + {F uh, } I ] watch the national news every day
- b. (dialogue sw01-0105 lines 07-08)  
 A: {D so } when you say the morning news, or evening news or national news is when? /  
 B: {F Uh, } evening news at six thirty I believe
- c. (dialogue sw03-0304 lines 01-02)  
 A: Tell me what you like to do. /  
 B: {D Well, } <laughter> [ I, +I ] collect antique tools ... /

Whereas stalling when starting to speak is typically a sign of wishing or agreeing to have the speaker role, ceasing to speak while fixating the gaze on another participant (and naming that other participant, especially in multi-party dialogue) is a sign of giving the speaker role to that participant. Slowing down and stalling at the end of an utterance, and a rising intonation, often signals that the speaker wants to keep the speaker role.

Turn management signals are often quite subtle, with an important role being played by nonverbal behaviour accompanying the speech. Since the Switchboard corpus consists of transcriptions of telephone dialogues, the annotation of turn management functions has to be based exclusively on verbal and vocal turn management signals. Turn-initial stallings, slash unit-initial and slash unit-final stallings, and interruptions are the main sources for adding ISO 24617-2 turn management functions to units in the Switchboard corpus. These could be added semi-automatically by identifying the turn-initial, slash-unit initial, and slash-unit final stallings and certain discourse markers, but each individual case would have to be checked for its communicative function in the context in

<sup>4</sup>See Bunt (2011) for a discussion of implications and other semantic relations between dialogue acts.

which it occurs; moreover, a partial resegmentation of the dialogues would be required in order to isolate the units to be annotated with turn-management functions.

### 4.3 Allo-Feedback

Feedback is communicative behaviour that provides or elicits information about the processing of utterances earlier in a conversation. The ISO standard follows the DIT<sup>++</sup> annotation scheme in dividing feedback behaviour into those where the speaker provides information about his own processing of previous utterances (Auto-Feedback) and those which provide or elicit information about the addressee's (or addressees') processing (Allo-Feedback). SWBD-DAMSL has tags for annotating Auto-Feedback acts, but not for Allo-Feedback acts.

Examples of Allo-Feedback acts in the Switchboard corpus are shown in (17) line 19 (Switchboard dialogue sw01-0105, lines 0012-A0005-03 to 0019-A0009-01), and in (18) line 66 (from sw00-0004, lines 61-66; in-line markups suppressed):

- (17)
- 12. A: I don't, uh, subscribe to cable
  - 13. B: Uh-huh.
  - 14. A: be- because of the poor service and also, uh, because,
  - 15. A: well, I, uh, I give to the United Way
  - 16. A: and so I figured that amount of money I just donate that.
  - 17. B: Uh-huh.
  - 18. as opposed to paying for cable.
  - 19. A: Yeah.

In line 18 in (17) B checks the correctness of his understanding of what A said, performing a Check Question (which is commonly expressed by a declarative sentence) in the Auto-Feedback dimension, to which A responds by a confirmation of B's understanding; this constitutes a Confirm act in the Allo-Feedback dimension. The SWBD-DAMSL annotation tags line 18 as *bf* ("Summarize/reformulate") and line 19 as *aa* (Accept/agree), which is not very satisfactory; in line 18 speaker B does neither summarize nor reformulate something that A has said, but rather adds a consideration to clarify what A said and offers this for confirmation, which A does in line 19, where he does not really express agreement with

what B said, but confirms the correctness of his interpretation.

- 61. B: I've nailed the problem
- 62. but I
- 63. A: <laughter>
- (18) 64. B: <laughter>
- 65. A: Leave the details up to someone else, huh?
- 66. B: Yeah,

In line 65 in (18) A provides a tentative completion of what B was trying to say in line 62, with a check of correctness, to which B replies with an allo-feedback Confirm act. The SWBD-DAMSL annotation tags line 65 as  $\hat{2}$  ("Collaborative Completion") and line 19 as *aa* (Accept/agree). Assigning only  $\hat{2}$  to the slash unit in line 65 fails to account for the , *huh?* part of that unit, which indicates that the speaker is not only performing a completion but also checks the correctness of his understanding on which the completion is based. In line 66 B confirms that correctness, which makes it a Confirm act in the Allo-Feedback dimension.

Identifying the units in the Switchboard dialogues which have an Allo-Feedback function seems quite hard on the basis of the existing SWBD-DAMSL annotations. An important clue is that allo-feedback acts mostly occur in response to allo-feedback acts, but the tagging of auto-feedback acts in the corpus is not very reliable, as example (17) illustrates, and does not seem to provide a solid basis for automatically identifying these acts.

### 4.4 Communicative function qualifiers

In natural dialogue, speakers often use expressions to qualify their communicative activity for (un)certainty, (un)conditionality, or sentiment. The ISO standard makes use of so-called 'qualifiers' (Petukhova & Bunt, 2010) for representing this in dialogue act annotation. SWBD-DAMSL does not have a device with the same generality, but does use the tag component  $\hat{e}$  to express uncertainty (but also other possible qualifications; 'e' stands for elaboration'), as in (19) line 27, and the tag 'am' ('accept maybe/partial accept' - see (19)) line 28, which can be used for some of the cases where ISO 24617-2 uses the qualifier 'uncertain' applied to the communicative function that interprets SWBD-DAMSL's 'accept' tag (which corre-



sponds to a number of more specific tags in the ISO standard).

(19) (dialogue sw03-0304, lines 25-28)

25. qy A: Are you going to move your whole family over there then?  
26. mn B: No, /  
27. sd ^e actually, {F uh, } I'm not even sure, /  
28. am B: I may, /

In a fully-fledged ISO 24617-2 annotation, it would be necessary to add function qualifiers wherever they apply, including interpretations of the cases where the SWBD-DAMSL tags and tag components 'h' (for 'hold'), 'am', and 'e' are used.

#### 4.5 Relations between dialogue acts

Responsive dialogue acts, such as answers, (dis-)confirmations, (dis-)agreements, acceptance and rejection of offers and requests, acceptance of apologies, return greetings, and so on, all presuppose a particular kind of preceding dialogue act, to which they have a 'functional dependence relation'. The ISO standard annotates these relations in a dialogue; SWBD-DAMSL does not.

Similarly, the ISO standard annotates the relations between a feedback act and the preceding dialogue contribution that the feedback is about, whereas SWBD-DAMSL does not support the annotation of such relations.

Again, in a fully-fledged ISO 24617-2 annotation of the Switchboard dialogues, it would be necessary to add functional and feedback relations wherever they would apply. Examples occur all over the place, for example in (20a) the slash unit in line 155 would be tagged as an answer that is linked to the question in line 153 by a 'functional dependence' relation, and in (20b) the feedback utterance in line 80 is tagged as an 'autoPositive' act that is linked to the preceding Inform by a 'feedback dependence' relation.

(20) a. (dialogue sw03-0304, lines 153-155)

153. B: [ You guys, + are you guys ]  
getting snow?  
154. A: We, - /  
155. it is snowing right now. /

b. (dialogue sw03-0304, lines 79-80)

79. A: {C so, } it's been a real interesting thing for them ../  
80. B: That's great. /

The addition of functional and feedback dependence relations to Switchboard annotations can probably be done semi-automatically, because of the following regularities that govern the dependency relations:

- for functional dependence:
  - these occur (always) for a particular set of dialogue act types, the 'responsive' ones, which are specified in the ISO standard;
  - for each type of responsive dialogue act the 'functional antecedent' is a dialogue act with a specific communicative function (like the functional antecedent of a Confirm being a Check Question) and a specific speaker;
  - the functional antecedent of a responsive dialogue act is nearly always the most recent dialogue act of the appropriate type (Petukhova et al., 2011).
- for feedback dependence:
  - these occur (always) for dialogue acts in one of the two feedback dimensions;
  - the 'antecedent' of a feedback act is in the vast majority of cases either the most recent dialogue act contributed by the previous speaker, or a subdialogue that ends there, intervening dialogue acts being mainly turn management acts, time management acts, and own communication management acts. In the latter case it may be difficult, however, to (automatically) determine the start of such a subdialogue.

## 5 Discussion and Conclusions

A comparison of annotation schemes is often thought of as comparing the respective typologies of dialogue acts and their encodings, but we have seen in this paper that the construction of ISO 24617-2 annotations for the dialogues in the Switchboard corpus, starting from the existing SWBD-DAMSL tagging, is much more complicated than that. Fang et al. (2012) have shown that the replacement of SWBD-DAMSL tags by ISO 24617-2 communicative functions can be done

automatically for 84% of the Switchboard corpus, which is a promising start. In this paper we addressed the following additional aspects of adding fully-fledged ISO 24617-2 annotations to the Switchboard corpus:

1. The ISO standard is intended for annotating all the communicative functions of dialogue units in the nine dimensions defined in the standard. The slash units in the Switchboard corpus have only one functional tag from the SWBD-DAMSL scheme, while on average they would have 1.8 communicative functions according to ISO 24617-2. This means that the number of functional tags in the corpus should be almost doubled. In some cases it is possible to derive the appropriate ISO tags from in-line markups (see 2, 3, and 7); in other cases this does not seem feasible (see 4, 5, 6, and 8). In nearly all cases, the addition of communicative functions requires the dialogues to be partly re-segmented, using the more fine-grained DIT<sup>++</sup> segmentation of dialogues into functional segments.
2. The in-line markup of restarts, repairs, and edit terms in the Switchboard corpus can be replaced semi-automatically by functional annotations in the ISO dimension of Own Communication Management, making use of the markup to automatically resegment the slash units in which these markups occur. The results must be manually checked, however, since edit terms and repetitions sometimes have other functions, e.g. as indicators of dialogue acts in the Partner Communication Management dimension.
3. The in-line markups of filled pauses can be used to resegment the utterances in which they occur, and to annotate these segments with Time Management functions. This can be done automatically with manual checks, since filled pauses can have functions in other dimensions than Time Management.
4. Discourse markers, as marked up in-line in the Switchboard corpus, have to be identified as separate functional segments if they express one or more dialogue acts by themselves. Their communicative functions cannot be derived from the Switchboard tagging, and require a re-annotation taking their context of occurrence into consideration.
5. Asides, such as third-party talk, have communicative functions just like other functional segments (and slash units), which can only be constructed through re-annotation with the ISO 24617-2 scheme.
6. Stretches of nonverbal communicative behaviour, such as laughter, chuckles, sighs, and lip smacks, should be treated as functional segments not only when they occur as a separate turn, but also when they occur inside a slash unit; their ISO 24617-2 annotation cannot be derived from the Switchboard markups.
7. Turn Management functions can be added semi-automatically to Switchboard once discourse markers have been treated as indicated in 4 and filled pauses as in 3, if detailed information is available about small pauses associated with turn-initial, segment-initial and segment-final discourse markers and filled pauses.
8. The addition of Allo-Feedback functions to Switchboard can partly be done automatically by identifying responsive dialogue acts that respond to a dialogue act in the Auto-Feedback dimension. The SWBD-DAMSL tagging is very crude in indicating dimensions, however; the tag component  $\hat{c}$  is used to represent “about communication”, so an Auto-Feedback Check Question could be tagged as  $qd\hat{c}$ , but this has not been done systematically in the Switchboard corpus (moreover, there is no SWBD-DAMSL tag corresponding exactly to ISO’s Check Question). In the absence of detailed encodings of functions in the Auto-Feedback dimension, it hardly seems feasible to derive Allo-Feedback functions automatically.
9. The ISO communicative function qualifiers for (un-)certainty and (un-)conditionality have no counterparts in SWBD-DAMSL; the tags and tag components ‘h’, ‘am’, ‘ $\hat{e}$ ’ can be used to automatically identify cases which are relevant to examine.
10. The functional and feedback relations that form an important part of the ISO 24617-

2 annotation of a dialogue can be added largely automatically for functional dependences, since these relations are known to occur always (and only) for certain types of dialogue acts (the ‘responsive’ ones) and nearly always relate to the most recent dialogue act of a specific type performed by the previous speaker. For feedback relations, similarly a good guess that could be used in a semi-automatic process is to take the last dialogue act performed by the previous speaker. Manual checks are needed to verify the correctness of the relations generated in this way, especially for feedback dependence relations, which may have a wider scope (for details see Petukhova et al., 2011).

With respect to the resegmentation and re-annotation that several of these aspects necessitate, it may be noted that Petukhova and Bunt (2011) have developed a highly successful machine-learning based approach for the automatic segmentation and annotation of raw spoken dialogue. A variant of this method could conceivably be defined for the ISO-compliant resegmentation and reannotation of Switchboard dialogues that makes use of the information encoded in the Switchboard transcriptions, in particular in the in-line markups.

### Acknowledgement

The research described in this article was supported in part by grants received from the General Research Fund of the Research Grants Council of the Hong Kong Special Administrative Region, China (RGC Project No. 142711) and City University of Hong Kong (Project Nos 7002793, 9610226, 9041694, 9610188, and 7008062). The authors would like to acknowledge the academic input and the technical support received from the members of the Dialogue Systems Group (<http://dsg.ctl.cityu.edu.hk>) based at the Department of Chinese, Translation and Linguistics, City University of Hong Kong

### References

Allen, J. and M. Core (1997). *DAMSL: Dialogue Act Markup in Several Layers (Draft 2.1). Technical Report*. Rochester, NY: University of Rochester.

Bunt, H. (2009). The DIT++ taxonomy for functional dialogue act markup. In D. Heylen,

C. Pelachaud, R. Catizone, and D. Traum (Eds.), *Proceedings of EDAML–AAMAS Workshop “Towards a Standard Markup Language for Embodied Dialogue Acts”*, Budapest, pp. 13–24.

Bunt, H. (2011). Multifunctionality in dialogue. *Computer, Speech and Language* 25(2), 225 – 245.

Bunt, H., J. Alexandersson, J. Carletta, J.-W. Choe, A. Fang, K. Hasida, K. Lee, V. Petuhova, A. Popescu-Belis, L. Romary, and D. Soria, C. Traum (2010). Towards an ISO standard for dialogue act annotation. In *Proceedings Seventh International Conference on Language Resources and Evaluation (LREC 2010)*, Malta.

Bunt, H., J. Alexandersson, J.-W. Choe, A. Fang, K. Hasida, V. Petuhova, A. Popescu-Belis, and D. Traum (2012). Iso 24617-2: A semantically-based standard for dialogue annotation. In *Proceedings Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul.

Clark, H. (1996). *Using Language*. Cambridge, UK: Cambridge University Press.

Clark, H. and E. Shaefer (1989). Contributing to discourse. *Cognitive Science* 13, 259–294.

Core, M. and J. Allen (1997). Coding dialogs with the DAMSL annotation schema. In *AAAI Fall Symposium on Communicative Action in Humans and Machines*, Boston, MA.

Fang, A., H. Bunt, J. Cao, and X. Liu (2011). Relating the semantics of dialogue acts to linguistic properties: A machine learning perspective through lexical cues. In *Proceedings 5th IEEE International Conference on Semantic Computing*, Stanford University, Palo Alto.

Fang, A., H. Bunt, J. Cao, and X. Liu (2012). The annotation of the Switchboard corpus with the new iso standard for dialogue act analysis. In *Proceedings 8th Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation (ISA-8)*, Pisa.

Geertzen, J., V. Petukhova, and H. Bunt (2007). A multidimensional approach to utterance segmentation and dialogue act classification. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, Antwerp, pp. 140–149.

- Jurafsky, D., E. Shriberg, and D. Biasca (1997). *Switchboard SWBD-DAMSL Shallow-Discourse-Annotation Coders Manual*. Available at <http://stripe.colorado.edu/>.
- Meteer and R. A. Taylor (1995). *Dysfluency annotation stylebook for the Switchboard corpus*. <ftp://ftp.cis.upenn.edu/pub/treebank/swbd/docs/DFL-book.ps>.
- Petukhova, V. and H. Bunt (2009). Towards a multidimensional semantics of discourse markers in spoken dialogue. In *Proceedings of the Eighth International Workshop on Computational Semantics (IWCS-8)*, Tilburg, the Netherlands, pp. 157–168.
- Petukhova, V. and H. Bunt (2010). Introducing communicative function qualifiers. In *Proceedings Second International Conference on Global Interoperability for Language resources (ICGL-2)*, Hong Kong, pp. 123 – 131.
- Petukhova, V. and H. Bunt (2011). Incremental dialogue act understanding. In *Proceedings Ninth International Conference on Computational Semantics (IWCS 2011)*, Oxford, pp. 235 – 244.
- Petukhova, V. and H. Bunt (2012). The coding and annotation of multimodal dialogue acts. In *Proceedings Eighth International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul.
- Petukhova, V., L. Prévot, and H. Bunt (2011). Multi-level discourse relations between dialogue units. In *Proceedings 6th Joint ACL-ISO Workshop on Interoperable Semantic Annotation (ISA-6)*, Oxford, pp. 18–27.
- Popescu-Belis, A. (2008). Dimensionality of dialogue act tagsets. *Language Resources and Evaluation 14*, 99–107.
- Shriberg, E. (1994). *Preliminaries to a theory of speech disfluencies*. Ph.D. Thesis, University of Berkeley.