Revised SPAAC Annotation Scheme

1. Introduction

The Speech Act Annotated Corpus (SPAAC) annotation scheme was applied to the British Telecom OASIS Corpus of 1200 telephone dialogues at Lancaster University. This document describes the main aspects of the scheme, aiming in particular at the revised version used in the NXT release of the OASIS material. Much of the content is taken unchanged from the original Lancaster description of the annotation approach.

The OASIS dialogues were supplied directly to the University of Lancaster by British Telecom, in the form of orthographic transcriptions. The annotated version of the corpus was later passed on to the University of Edinburgh along with a number of recommendations for changes. Edinburgh effected these changes and made some other alterations to make the corpus compatible with NXT, allowing simpler representation and searching.

The Lancaster annotators did not refer to the original sound recordings. Where anomalies occurred in the original transcription, the annotators did not attempt to correct or otherwise change the transcription.

> Note [a]: In referring to the two speakers in a dialogue, we will use the terms 'speaker x' and 'speaker y', as where there is a frequent change of speaker, the terms 'speaker' and 'addressee' can cause confusion.

> Note [b]: It is a general rule that each turn contains at least one utterance. Exceptions occur in the BT transcripts, where an "empty turn" sometimes occurs, containing some non-verbal action or comment.

An utterance can contain more than one **C-unit** (a term to be discussed below, and previously used and defined for spoken discourse in Biber et al. (1999: 1069-1082)), but usually contains one C-unit only (see the next paragraph). C-units, which can be defined as independent syntactic units of spoken English grammar, are assigned speech-act attributes and other attributes.). In the NXT version of the corpus, the SEGMENT element is used to identify C-units.

Utterances are not given separate attributes, and so the present annotation scheme pays attention only to C-units. The segmentation of a dialogue into turns is a 'given' provided by the transcription. The segmentation of a turn into C-units, on the other hand, is a part of annotation. It was accomplished automatically by a **splitting routine**, which takes account of pauses and other phenomena.

An utterance cannot contain less than a C-unit. If an utterance contains more than one C-unit, then one of the C-units will be just a **discourse marker**, such as *well* or *now* or *okay*. The

structure of an utterance, therefore, is a C-unit plus (optionally) one or more discourse markers associated with it.

2. C-units and Form Categories

A C-unit is the basic unit for speech act annotation. Syntactically, it is an independent clausal or a non-clausal unit (a non-clausal unit is here labelled as a 'fragment' – **frag**). Functionally, it represents a unit which can be assigned to a given communicative function, represented by its **speech act attribute**.

C-units are the basic units (or 'envelopes') for conveying **speech acts** (also called 'moves' or 'dialogue acts'), which can be regarded as the minimal communicative actions performed in a dialogue. C-units are initially classified in terms of their grammatical form using one of the following **form labels:**

- **decl** (declarative clause),
- **q_yn** (yes-no question),
- **q_wh** (wh-question),
- **imp** (imperative),
- **frag** (fragment – i.e. a non-clausal unit, which has no finite verb – e.g. *London Euston* – or (rarely) a clause which is incomplete because it lacks a subject – e.g. *arrives at 10.10*),
- **dm** (discourse marker – normally a single word which has no denotative or referential content, but simply signals a particular pragmatic role in the dialogue),
- **yes** (affirmative reply),
- **no** (negative reply).

Of these, **yes** and **no** could be treated as special cases of discourse markers, but because of their significance in the dialogues, they are placed in distinct categories, and given special form labels.

As noted above, in the distributed version of the corpus, C-units are represented by the SEGMENT element, and each of the form labels above is represented as a type attribute; for example:

```
<segment type='decl' sp-act='answ' id='le000130.decl.1'>
   ......
</segment>
```

Brief characterizations of the form label categories follow.

2.1. decl (declarative clause)

A segment whose type is **decl** is an independent clause in which a subject precedes a finite verb:

e.g. *It takes three hours.* A complex clause structure containing more than one subject + finite

verb also counts as a **decl**, but with exceptions noted here:

(a) two coordinated clauses may count as a single **decl** where they are run together without a

pause, and convey a single speech act function. But where they are separated by a pause, and/or

convey distinct speech act functions, they are annotated as more than one **decl.**

(b) an adverbial clause, especially when introduced by *(be)cause, if* or *when*, counts as a separate

decl if it is judged to convey a distinct speech act function.

2.2. q_yn (yes-no question or alternative question)

A **q_yn** is an interrogative C-unit recognized by the reversal of the position of subject and

(auxiliary) verb, as in *Does it matter? Are you ready yet?*

Alternative questions, which offer a choice between alternatives separated by *or*, are included in

this category: e.g. *Do you want a forward-facing or backward-facing seat?* (However, they are

distinguished by the mode label "alternative"):

> Note: Declaratives followed by tag questions (e.g. *That's your home number, is it?*) are classed as a single
> **decl**, although in speech-act terms, they are classed as elicitations of a verbal response: ***reqInfo*** is their usual
> speech-act label. Questions which have a declarative form (e.g. *That's you home number* – spoken with rising
> intonation) are also classed as **decl**.

2.3. q_wh (wh-question)

A **q_wh** is an interrogative C-unit recognized by the occurrence of one of the wh-words *what,*

which, who, whom, whose, when, where, how, why normally at or near the beginning of the unit,

and occupying one of the major syntactic roles such as subject, object, adverbial.

> Note [a]: In wh-questions, the reversal of the position of subject and (auxiliary) verbs (as in 2.2) often occurs,
> too: *When **do you** want to return? How far **is it**?* Such units are annotated: **q_wh**.
> Note [b]: Sometimes, however, the *wh*-word comes at the end: *It arrives **when**?* This type is also classed as a
> **q_wh**.
> Note [c]: A question which has a wh-word but no finite verb is considered to be a wh-question (**q_wh**), not a
> fragment (**frag** – see below): e.g. *What about this handset? Who else?*

2.4. imp (imperative)

An **imp** is an imperative clause, with a verb in its base (uninflected) form which normally occurs

in initial position: *Look where you're going. Come on*. A clause beginning *Let's* or *Let me* is

classed as an **imp**, although in its function it is likely to imply action on the part of the speaker rather than of the addressee.

> Note: An utterance beginning with a base form of the verb is not always imperative. For example: *A: What did you do? B:* **Phone** *the police*. Here the construction is elliptical (=*What I did was phone the police*), and the attribute value **frag** is used instead (as the unit lacks a finite verb.)

2.5. frag (fragment)

The form category **frag** applies to C-units which do not contain an independent finite clause. (However, see also 2.6-2.8 below.) A **frag** may be

- a single stand-alone word (e.g. *Really? Today.*)
- or a stand-alone phrase (e.g. *Sorry about that. Birmingham New Street.*)
- or a non-finite construction, containing an infinitive or a participle (e.g. *Arriving at Euston at 9.45.*)
- or a combination of structural types: e.g. *Just a moment, please*.

It should also be noted that, if a statement lacks one of the elements of the subject – finite verb combination criterial for declaratives (perhaps because of ellipsis), it is treated as a **frag** rather than as a **decl** or **q_yn**: e.g. *Can't help it* and *You coming* have **frag** as their form category.

2.6. dm (discourse marker)

A discourse marker is normally a single word which stands alone as an utterance or else is loosely attached to a larger C-unit and is part of the same utterance. (A **dm**, therefore, is an element peripheral to syntax, which does not enter into larger integrated constructions.) Discourse markers have a range of functions and can be multi-functional in dialogue. But in general they act as

(a) indicators of how the speaker is reacting to what is being said and/or

(b) indicators of the direction in which the speaker intends to take the discourse. Most discourse markers are small words such as *well, so, right, alright, ok, now*. Occasionally a **dm** contains more than one word: e.g. *okay now*.

2.7. yes (positive response)

The response word *yes*, and variants such as *yeah, yep*, are assigned the form category **yes**, which also includes combinations such as *yes please*.

2.8. no (negative response)

As for **yes**, **no** is a form category for negative responses: viz. *no* and its variants, including combinations such as *no no, no way.*

3. XML mark-up

Each dialogue was automatically converted to XML (eXtensible Markup Language) by Lancaster University and annotated in XML. The SPAACy annotation tool (created by Martin Weisser) enables a dialogue to be annotated through a combination of automatic, interactive and manual procedures.

After the conversion into a simple XML format, Lancaster ran various analysis stages in order to determine the type of segment (C-unit) in hand. Below is an illustration of the annotation of a turn in the original Lancaster format:

```
<turn id="5" speaker="A">
   <utt id="7">
     <q_yn sp-act="reqInfo" polarity="positive"
           topic="creditcard" mode="closed-alternative">
              do you hold a current debit or credit card
     </q_yn>
   </utt>
</turn>
```
Figure 1.1 – illustration of the basic turn structure

The attributes on the **q_yn** tag convey further information about the content of this C-unit; **sp-act** (speech act), **polarity**, **topic** and **mode**. However, the focus of this annotation scheme is on the form category and the speech act attribute, and in the NXT version the structure above appears as:

```
<lturn id="le000130.a.lturn.5">
    <utt id="le000130.utt.7">
      <segment type="q_yn" sp-act="reqInfo" alternative="yes" id="le000130.q_yn.2">
         <nite:child href="unit-range"/>
      </segment>
    </utt>
</lturn>
```

Figure 2.2 – NXT version of the basic turn structure

The **topic** and **polarity** attributes were viewed as of secondary importance, having a diagnostic function in the process of annotation, and were deleted.   Two aspects of the **mode** value, however, are retained in the NXT version.   One of these cases is where the **mode** attribute value signalled that the move had been abandoned (left incomplete by the speaker). The NXT version of the corpus uses the attribute SECONDARY-ACT for this information, for example:

```
<segment type='decl' sp-act='reqInfo' id='le752150.decl.13' secondary-act='abandon'>
```

Any kind of speech act can be abandoned. The other mode value which is retained is 'alternative', which is used to signal alternative-OR questions, as in *Would you like coffee or tea?*; this only occurs with **q_yn** segments, and is indicated using the **alternative** attribute. The example **lturn** structure in Figure 1.2 above contains an instance of this.

4. The Function of this Document: an Annotation Scheme

The purpose of this document is to specify the task of assigning speech-act labels to the corpus in enough detail to ensure a consistent practice of annotation. This is an ideal goal which cannot be achieved 100 per cent. An annotation scheme has a dual role.

(a) For annotators of future corpora, it provides guidelines to be consulted in the course of assigning and post-editing (checking and correcting) annotations.

(b) For users of the present corpora, it provides an explanation of the symbolic practices used in annotation.

Apart from **form categories**, which have already been characterized, the most important part of the annotation is the **speech-act categories** (a convenient term for what are strictly speech act attribute values assigned to form categories).

Speech-act categories are both form-determined and context-determined. In practice, a set of speech-act categories can be relatively more form-related or relatively more context-related. For example, the same C-unit *Could you check the line?* could potentially be labelled as a 'question' (more form-related) or 'request-service' (more context-related). We could also say that 'question' would be a more general (**generic**) label, whereas 'request-service' would be more specialized to a particular type of service dialogue.

In this project we have aimed at an intermediate position on this form-related ↔ context-related scale. We are aiming to create a **generic speech-act categorization scheme** which could apply to many different types of task-oriented dialogue. In fact in this project we have agreed to annotate three different kinds of dialogue. We cannot afford to use labels which are too specific

to a given task. Also, the task of speech-act categorization has been undertaken by a combination of automatic and manual procedures. The automatic part of this assignment depends heavily on the recognition of formal aspects of each C-unit – viz. key words and syntactic structure. This consideration explains the nature of the speech-act labels which we have used, and which are listed and explained below.

The normal assumption is that a single C-unit has **one** speech-act category only. But there are occasions where it is difficult to avoid the conclusion that two labels should be attached to the same C-unit. In telephone service dialogue, a regular example considered bivalent in this way is illustrated by a common type of utterance where the service provider says *I'll just check on that for you*. The function of this utterance on the one hand is to signal that speaker x has the intention to perform some action (***informIntent***). On the other hand, the function is to signal that a gap in the dialogue is likely to ensue, and that speaker y should hold the line. Hence a second value, ***hold*** is added, and in the original version of the corpus this was captured using the compound speech act label ***informIntent-hold***.   Apart from this relatively common compound label, only one further compound label was allowed in the annotation scheme, viz. ***thank-bye*** (see 6.38)***.*** In the NXT version of the corpus we used the SECONDARY-ACT attribute for the additional information, for example:

```
<segment type='decl' sp-act='reqInfo' id='le752150.decl.13' secondary-act='abandon'>
```

In the major part of this document that now follows, we present an alphabetical list of speech-act categories, together with brief explanations of criteria for label assignment.

5. A classified list of speech-act categories

By way of introduction, Table 1 below gives an overall list of the speech-act categories used, with an indication of their function in the dialogue:

Table 1 – Classified List of Speech Act Labels	
accept	*mainly responding*
ackn(owledge)	*mainly responding*
answ(er)	*mainly responding*
answ(er)Elab(orate)	*mainly responding*
appreciate	*mainly responding*
bye	*interpersonal management*
complete	*dialogue control*
confirm	*responding / initiating*
correct	*dialogue control*
correct-self	*dialogue control*
direct	*mainly initiating*

directElab(orate)	*mainly initiating*
echo	*dialogue control*
exclaim	*expressive*
expressOpinion	*expressive*
expressPossibility	*expressive*
expressRegret	*expressive*
expressWish	*expressive*
greet	*interpersonal management*
hold	*dialogue control*
identifySelf	*dialogue control*
inform	*mainly initiating*
informIntent	*mainly initiating*
informIntent-hold	*dialogue control*
init(ialize)	*dialogue control*
negate	*mainly responding*
offer	*mainly initiating*
pardon	*dialogue control*
raiseIssue	*mainly initiating*
refer	*mainly initiating*
refuse	*mainly responding*
req(uest)Direct	*mainly initiating*
req(uest)Info(rm)	*mainly initiating*
req(uest)Modal	*mainly initiating*
selfTalk	*external to dialogue goals*
suggest	*mainly initiating*
thank	*interpersonal management*
thank-bye	*interpersonal management*
thirdParty (talk)	*external to dialogue goals*
unclassifiable	(unspecified)
uninterpretable	(unspecified)

Note: To simplify, the letters in brackets above are omitted.

Table 1 gives an alphabetical list of the speech-act categories, which are to be explained further

in Section 3 below. In the second column, these speech act types are classified in terms of

a more general set of function-oriented categories. Brackets show parts of the labels

which are omitted (for brevity) in the speech-act categories. The tags are classified in

terms of a small set of more general categories. For example, most common speech acts

can be classified as MAINLY INITIATING (e.g. a request for information) or MAINLY

RESPONDING (e.g. an answer to such a request). Hence, such speech acts commonly co-

occur in pairs – classically called 'adjacency pairs'. But on the other hand, there are also

other categories such as EXPRESSIVES, which are not particularly prone to initiate or

respond. Figure 2 shows the relationship between these categories in the form of a tree

diagram.

```
                        ┌─────────────┐
                        │ a. dialogue │
                        └─────────────┘
                       /               \
          ┌──────────────┐          ┌──────────────┐
          │ b. external to│          │ c. dialogue  │
          │ dialogue goals│          │   proper     │
          └──────────────┘          └──────────────┘
                              /          |          \
              ┌──────────────┐  ┌──────────────┐  ┌──────────────┐
              │ d. directly  │  │ e. dialogue  │  │ f. interpersonal│
              │  pursuing    │  │   control    │  │  management  │
              │  dialogue    │  └──────────────┘  └──────────────┘
              │   goals      │
              └──────────────┘
              /      |       \
  ┌──────────┐  ┌──────────┐  ┌──────────┐
  │ g. mainly│  │ h.       │  │ i. mainly│
  │ initiating│ │ expressive│ │ responding│
  └──────────┘  └──────────┘  └──────────┘
```
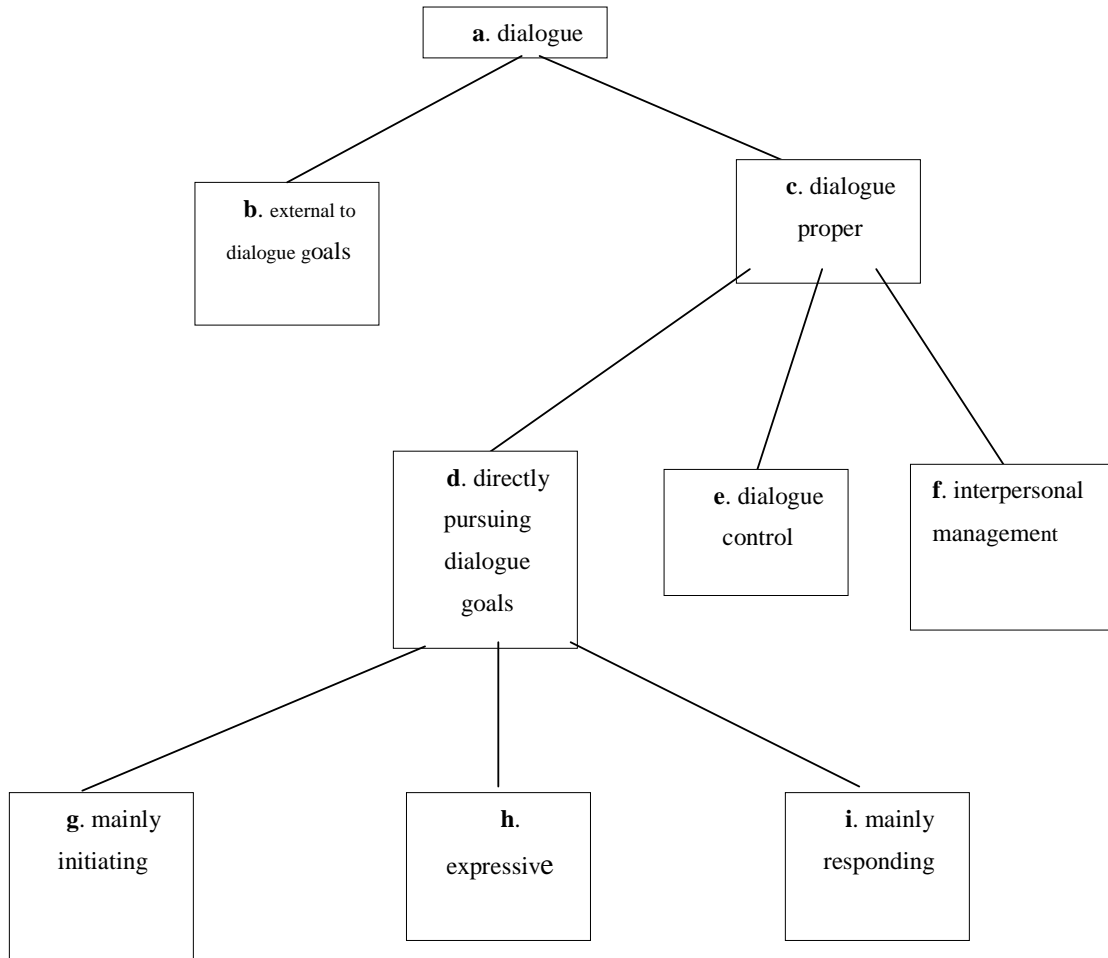
Figure 2: Major categories of speech acts in dialogue

In the following sections we provide a glossary of all the speech-act categories used in the annotation. In practice, the boundaries of speech-act categories are fuzzy, and a major task of the annotation scheme has been to limit the fuzziness as far as possible, so arbitrary rules of assignment sometimes have to be employed. We have used the guideline that there is a hierarchy of rule applications corresponding to the above hierarchy of speech-act categories. For example, a category in the "**b** external to dialogue goals" class overrules one in the "**e** dialogue control" class, and one in the "**e** dialogue control" or "**f** interpersonal management" class overrules one in the "**g** mainly initiating" or "**i** mainly responding" class. To give two examples:

(i) The speech act category **(b)** *thirdParty* is assigned to a C-unit *What did you say love?* (addressed by the caller to a third person) which might otherwise be assigned **(e)** *pardon.*

(ii) The speech act category **(e)** *hold* is assigned to a C-unit *Hold on!* which would otherwise be assigned **(g)** *direct*.

A way of explaining the logic of (i) is to say that if an utterance is addressed to a third party, this is to be marked first and foremost as external to dialogue goals: i.e. the category ***thirdParty*** is assigned. As this C-unit is peripheral to the dialogue, no more precise label needs to be assigned. (In developing models of dialogue, it would be reasonable to exclude third party dialogue, and so further analysis would be unnecessary.) Similarly, the logic of (ii) is to say that if an utterance has a dialogue control goal such as asking the other speaker to hold the line (i.e. the speech act category is ***hold***), whether this is put in the form of a directive like *Hold on!* or an information statement like *I won't keep you a moment* is of secondary importance, and can be ignored.

6. Definitions of Speech-act Labels: a Glossary

The following section presents a glossary of the speech-act labels in Table 1:

6.1 accept

A positive response (e.g. *yeah, yes, okay, right, That's fine*) to a initiating speech act such as a ***suggest***, ***offer***, ***expressRegret, informIntent*** or ***direct***. The speaker does not just acknowledge that the previous speech act has been heard and understood, but also accepts, complies with or agrees with the force of that speech act. The dialogue management speech act ***hold*** can be followed by an ***accept****,* too. E.g.

(1) *A: Could you just hold on a second.*
     *B: Sure. (=**accept**)*

(2) *A: Could you try for me*
     *B: Certainly. (=**accept**)*

Acceptances of offers can look like requests, using the same formulae that in other contexts would mark an utterance as a ***direct***: e.g. *A: Shall I cancel your booking?* can be accepted by:

     *B: Yes **please**.   (=**accept**)* OR:

     *B: **Would you mind**? (=**accept**)*

These responses are not annotated as ***directs***, although they are clearly very similar to requests in their intention. (But if the preceding turn were annotated as ***reqDirect****,* then ***direct*** would be appropriate for the response.) The negative counterpart of an ***accept*** is a ***refuse***.

Note [a]:

A 'minimising' response following an apology (labelled **expressRegret**), such as *That's alright, Not to worry*, is treated as an **accept**. In effect, the speaker by saying this accepts the apology. On the other hand, such a response following an expression of thanks (labelled **thank**) is treated as an acknowledgement (**ackn**).

Note [b]:

Accept is also used occasionally for "agreement" to an assertive speech act such as an **inform** – i.e. where the label **ackn** would be too weak, because the responder is doing more than just signalling understanding.

Note [c]:

After a speaker refuses or declines an offer, a minimising utterance classified as an **accept** commonly follows: e.g.

> A: Shall I put you through to directories?
>
> B: No. (=**refuse**).
>
> It doesn't matter. (=**accept**)

It may seem contradictory to allow an **accept** to follow its apparent opposite, a **refuse**. However, in this context, *It doesn't matter*, or a similar minimising utterance, signals acceptance of the other speaker's intention (in making the offer) to be helpful. On the other hand, the **refuse** is refusing or declining the offer.

6.2 ackn

An acknowledgement is a casual positive response (e.g. *yeah, yes, okay, right*) to a declarative C-unit, or to a fragment C-unit which is assumed to have a declarative/informative/assertive intention. An **ackn** is a backchannel (but see Section 1 above), signalling that the speaker is following or taking on board what the other speaker is saying. E.g. *A: It's the latest train you can get. B: Yeah*. An **ackn** can be seen as a weaker equivalent of an **accept**.

Note [a]:

After a **confirm**, the minimal response is considered an **ackn** rather than an **answ**.

Note [b]:

The label **ackn** can apply to a negative response e.g. *No*, where speaker x is giving a backchannel to a negative statement uttered by speaker y. E.g.: *A: There's nothing happening. B: No.*

Note [c]:

A special variant of **ackn** occurs where a speaker responds to a thank-you by using a minimising expression such as *you're welcome*, or *it's okay*. (However, a minimising response to an apology is deemed to be an **accept**.)

6.3 answ

An **answ** is a positive, negative, or other more contentful response to a question (as a request for information). E.g.

Do you have a railcard? **Yeah / No**.

When does it arrive? **At 17.30.**

Note [a]:

A NOTE ON *YEAH, YES, OKAY, RIGHT*: These response forms can count as an **accept**, **ackn** or **answ** (and **init** is also possible). In broad terms, **answ** is used as an answer to a question, while **ackn** is used as a backchannel responding to a declarative (statement). **Accept** is used where neither **ackn** nor **answ** seems appropriate. The utterance to which an acceptance responds is initiating: it puts forward an agenda (or a proposal or proposition) of some kind, to which the speaker is signalling acceptance or compliance. (**Accept** is also the positive counterpart of **refuse**.) .

Note [b]:

If the response to a question is not a direct answer, but does in some degree provide information, the label **inform** is used instead of **answ**. (E.g. *A: What's the time? B: I don't know*. Or *A: What town's the code 01494 for? B: Ah, you'll have to ask Directory Enquiries for that. = **inform***) This may especially occur in those cases where the response initiates a digression from the current topic initiated by the other speaker. E.g.:

*A: For which journey do you wish to purchase a ticket? (=**reqInfo**)*

*B: Well, I've got a problem. (=**inform**)*

Note [c]:

Despite Note [b], an indirect response to a yes-no question is labelled **answ** where it can be explained in terms of an ellipsis of *yes* or *no*. For example:

    A: Is that Southampton? B: Portsmouth (= No, Portsmouth).

    A: Are you leaving soon? B: In half an hour (= Yes, in half an hour)

6.4 answElab

Naturally enough, an answerElab is an elaboration of an **answ**. It may occur if an answer has already been given in a previous C-unit and the speaker is providing additional information, e.g. *Yes, a debit card*, where *Yes* is the **answ** *a debit card* is the **answElab**; or *No, I don't*, where *I don't*, although it adds no further information, clarifies or reinforces the *No*. Similarly, in *I do, yes*, the *yes* is an **answElab** following and reinforcing the **answ** *I do***.** The fact that it adds no further information to the **answ** does not disqualify it from being an **answElab**. Note, further, that the elaboration can actually continue for a few utterances, even spanning a few turns in the case of partial information, such as sequences of numbers for a credit card, where there are often intervening echoes from the other speaker. In this case, a sequence of **answElab**s is used. **AnswElab** is often equivalent to the high-level category **inform**, except that it occurs in response to a question. Here is an example (with fictitious numbers):

A1: What's your credit card number?

*B2: 2154   (=**answ**)*

*A3: 2154   (=**echo**)*

*B4: 7938   (=**answElab**)*

*A5: 7938....   (=**echo**)*

6.5 appreciate

This speech-act category applies where speaker x responds appreciatively to a previous turn in which speaker y has indicated something from which speaker x is presumed to benefit. *Appreciate* normally labels short, somewhat formulaic C-units such as *(That's) great / wonderful / lovely / marvellous*. Also, **appreciate** often accompanies other short speech acts such as **ackn** or **thank**.

Note:

An **appreciate** is sometimes bifunctional in fulfilling the role of an acknowledgement while also expressing appreciation. In such cases, the single label **appreciate** is used in preference to **ackn**.

6.6 bye

Saying farewell at the end of the dialogue. Normally the word *bye* itself is used, sometimes accompanied by other words such as *Okay then...* If the farewell word is repeated, as in *bye bye*, this still counts as a single speech act **bye**. (See also **thank-bye**.)

6.7 complete

Speaker x co-operatively completes an utterance started by speaker y. This happens rarely in task-oriented dialogues, although it is quite common in conversation. E.g.:

   *B: third of Oc... A: **October***

The **complete** may be said in a questioning or declarative manner, but this does not effect the way the C-unit is labelled. The preceding turn, which is left incomplete by its speaker, should be labelled in the way which would have been appropriate if the speaker had completed it him/herself.

Note: Very occasionally there is a conventionalized interrogative use of an incomplete statement. For example:

   A: And your address is? (=**reqInfo**)

   B: Ninety-nine Pemberley Avenue, Milton Keynes (=**answ**)

This could be analysed as a sequence of an incomplete utterance followed by a **complete**, but the interrogative function is so conventionalized that the first C-unit is not felt to be incomplete. Hence *A's* and *B's* utterances are respectively classed as **reqInfo** and **answ**.

6.8 confirm

The speaker repeats and/or summarises some information that has been previously discussed, or which is otherwise assumed to be given or 'retrievable' from shared knowledge – e.g. a recap of information about the departure date or time – at the same time giving the hearer a chance to correct it. This frequently occurs in summaries by the service provider. But, unlike *echo*es, a

confirm is not an absolute verbatim repetition of what has been said before. More than one *confirm* may occur in a sequence, where the summary consists of a sequence of C-units.

    Confirms are important in service telephone dialogues such as the Trainline dialogues, where the agent needs to confirm the details of the transaction, so that no mistake (such as sending the wrong train tickets) will occur. This can also be characterized as a speech act of **grounding**, i.e. one establishing or confirming common ground to be shared by speaker and hearer. *Confirm* is not, on the other hand, used to label another kind of utterance which in ordinary parlance might be said to "confirm" – for example, when speaker x asks speaker y to repeat information already given.

    Note [a]:

    There are cases where *confirm* is applicable even though the information has not previously been mentioned. E.g. *A: Third of October. B: Saturday*. Here the agent confirms the day of travel by adding **retrievable** information, viz. the day of the week. The *inform* speech act – the alternative – would not be appropriate here.

    Note [b]:

    On the other hand, the label *confirm* is not used where the utterance takes the form of a question, or where it contains a tag question. E.g. *You're coming next week are you?* Generally, this is classed as a *reqInfo* instead, and the other speaker responds with an *answ*.

    Note [c]:

    *Confirm*s are poised between 'initiating' and 'responding' speech acts. They recapitulate what has been said or implied, but on the other hand, they also look forward to a response from the other speaker, who is expected to signal the correctness of the confirmation. In this sense they both **confirm** and **seek confirmation** from the interlocutor. The interlocutor's response, by the way, is classed as an *ackn* (not an *answ*).

6.9 correct

This label might describe more explicitly a 'correct misspeaking' speech act. It means that speaker x is correcting something which speaker y said wrongly: E.g.

    *A: And you're taking the 12.35 to Birmingham International (=**confirm**)*

    *B: **Birmingham New Street**. (=**correct**)*

    *A: Sorry Birmingham New Street. (=**correctSelf**)*

The second C-unit uttered by A is classed not as a *correct*, but as a *correctSelf* (see the next section).

6.10 correctSelf

See *correct* above for an illustration of this label. *CorrectSelf* is not used for a normal disfluency phenomenon, where the speaker corrects what he or she just said in the same turn – resulting in a false start: e.g. *I want to go to Birmingham Internation {#} Birmingham New Street.* This kind of

repair phenomenon does not involve a new C-unit (see Section 1), but just results in the addition of a 'repair' within an existing C-unit. But, if the repair results in a new C-unit in its own right, as in the example under **correct** above, this is treated as a **correctSelf**. Note that a **correctSelf** may occur as a new C-unit in the same turn as the error: e.g. *I want to go to Birmingham International {#} Sorry, **Birmingham New Street**.*

6.11 direct

A speech act with the communicative intention to bring about a response by the addressee. The response should be not just a verbal response (as in answering a question – see **answ**) but should include some physical or mental action or the maintaining of a physical or mental state of affairs – e.g. booking a ticket: *I'd like to book a ticket* or *Would you book that for me please* - or a negative instruction such as *Don't forget to collect your tickets from the booking office*. A **direct** can be conveyed by an imperative, a question, a declarative, or a fragment. Its force can vary from strong to weak: e.g., from a command or instruction, to a request or an entreaty. These are all directive speech acts, although they vary a great deal in the amount of indirectness or politeness they employ:

> *Could you put me through to emergency.*
> *Yes, dial the same number.*
> *I want the faults line, please.*

The politeness marker *please* is a tell-tale signal for a **direct**. However, it is far from infallible, because requests for information also often contain *please*, as well. Also, a response to an **offer**, labelled an **accept**, can contain *please*.

> Note [a]:
>
> It is important the distinguish **direct**s from **suggests** – the latter place no obligation on the hearer to do anything, and tentatively propose an action or decision either by the hearer alone, or by the speaker and hearer combined. **Suggests** imply benefit to the addressee rather than the speaker. They include giving advice.
>
> Note [b]
>
> A **direct** is sometimes a response to a **reqDirect**, and in this function can be a fairly minimal utterance, such as *Yeah* or *Yes please.* E.g.:
>
>> *A: Do you want to return on the same day?*
>> *B: Yeah. / Yes please. / Please.* [=**direct**]
>
> A minimal negative response could also be a **direct**, e.g. *No thanks*, as a response to *Do you want a forward-facing seat?* In response to an **offer**, on the other hand, *No thanks* would be classified as **refuse**.
>
> Note [c]:
>
> A telephone dialogue often begins with an utterance like *I wonder if you can help me please?*. At this stage in the dialogue, the nature of the request (whether it's a request for information, or a directive), is not clear. Thus later it may turn out that the caller has asked for verbal information, or else for an action like checking a line

15

or dialling a telephone number. Because of this uncertainty, we label this general request *direct* (rather than *reqInfo*), using the most general directive label.

6.12 directElab

An elaboration of a *direct*. This speech act immediately follows a *direct* which is a positive or negative response to a *reqDirect*. E.g.:

> *A1: Do you want (me) to book that seat? (=reqDirect)*
>
> *B2: Yeah, (=direct) but I'd like it non-smoking please. (=directElab)*

Or:

> *A1: Do you want me to put you through?*
>
> *B2: Yes please, (=direct) to directory enquiries (=directElab)*

You will see that *directElab* is closely parallel to *answElab* in responses to requests for information.

6.13 echo

A response in which the speaker simply echoes or 'parrots' something the other person said in a preceding turn. What is echoed can be the whole of, or just a part of, the preceding turn. The function of an *echo* is generally to make sure that what speaker y said has been correctly heard and decoded by speaker x. Like *confirm*, it is a speech act of **grounding**. This may be important, e.g., in verifying that personal details such as speaker y's postcode or credit card numbers have been correctly noted.

> Note [a]:
>
> An *echo* is functionally similar to a backchannel (labelled *ackn*). But, unlike a *ackn*, an *echo* also has the function of verification of the message.
>
> Note [b]:
>
> *Echo* is used where the current utterance consists entirely of material repeated from the earlier turn. Where speaker x repeats what speaker y said **and adds something else**, this is not an echo. For example, in the following, *B*'s utterance is counted as an *inform*, or a *confirm* if it is clear that *B* is seeking to confirm what has already been stated:
>
>> *A: Arriving at 7.50.   B: 7.50 p.m.*
>
> Exceptions to this "exact repeat" rule are not even granted where the utterance was obviously in intention a repetition, but where there was a slight discrepancy of expression: e.g.
>
>> *A: the sixth October.   B: of October. (= confirm)   A: Yes.*
>
> However, in one case, we do grant an exception to the "exact repeat" rule. This is where it is clear that speaker x did intend to echo what speaker y said, but made a mistake (for example, through mishearing what speaker y said). It is the speaker's assumed <u>intention</u> to echo that is more important here :
>
>> *A: 0 2 5 3 (= inform)*
>>
>> *B: 0 2 5 4 (= echo)*

*A: No, 0 2 5 3 (= **correct**)*

Note [c]:

Although an ***echo*** has to consist entirely of repeated material, note that the preceding utterance it echoes may have contained additional material not repeated: e.g. *A: The code is 0 1 5 2 3.   B: 0 1 5 2 3 (=**echo**).*

6.14 exclaim

This is a speech act whose primary purpose is to express emotion – typically either pain (*Oh no! Jesus! Oh God! Pity. What a nuisance*) or pleasure (*Wow! It's so fantastic*). An ***exclaim*** should not be confused with ***appreciate*** (see 3.5), where the pleasant emotion is directed towards what someone else is doing or has done. Adjectives such as *Excellent. Great. Lovely*, uttered as responses to some service the other person has provided, are examples of ***appreciate***.

Note [a]:

What can be an ***exclaim***, as well as a ***question*** or a ***pardon.*** In the case of an *exclaim*, the word *What!* may be imagined as followed by *!* rather than *?*, and as expressing an emotion such as disappointment or outraged incredulity.

Note [b]:

The word *oh* is an ***exclaim*** (not an ***ackn***) if it occurs on its own in a C-unit. Otherwise, it can come at the beginning of a larger C-unit, which may itself be an ***exclaim*** as in *Oh dear!*, or may be a different speech act. E.g. *Oh I see* is likely to be an ***ackn***. In deciding whether to split ***oh*** off from what follows, the relevant criterion is: Does *oh* have the force of a separate exclamatory speech act? This is more likely if a pause follows it.

Note [c]:

More peripheral cases of ***exclaim*** signal not so much emotion as recognition. *Oh* sometimes falls into this category. A highly peripheral case is an utterance such as *There you are* – for example, where the telephone operator signals recognition of an event taking place at the moment of speech, as in: ***There you are*** *– it's ringing for you*. We call this an *exclaim*, really for lack of an alternative.

6.15 expressOpinion

In this case (in contrast to ***inform***), a typically <decl> C-unit expresses an evalution or judgement, rather than providing information about the world. ***ExpressOpinion*** speech acts often begin with, or contain, expressions like *I think, I guess, I bet, My view is*. This speech act is not so common in task-oriented dialogues, but occurs, for example, in the Trainline data where the customer 'hazards a guess' about the time of a train: *I think there's one about 5.30.* Sometimes ***expressOpinion*** is realised by an interrogative or fragmentary form, rather than by a declarative form. For example: *Isn't it a pity they've sold them all* or *Pity they've sold them all* or *Awful day isn't it.*

Note: ***ExpressOpinion*** should only be applied where the default speech act ***inform*** is clearly inappropriate. This occurs, in particular, where no definite new information is offered by the utterance. For example, in *I*

think there's one about 5.30, the speaker has no definite information, but is merely putting forward a statement which may or may not prove true.

6.16 expressPossibility

Like ***expressOpinion***, ***expressPossibility*** typically applies to an eventuality which may or may not prove true. Most examples of this speech act contain one of the modals *may, might,* or *could*. E.g.:

(1) *I'll put you through to directory enquiries,* **they might be able to help you**.

(2) *There could be a fault.*

On the other hand, the modal *can* never expresses 'possibility' in this sense. For example, the second part of (3) is a ***suggest***, since it puts forward the plan of some action by the hearer (or the speaker and hearer combined):

(3) *I'll put you through to our faults department,* **you can have a word with them.**

6.17 expressRegret

There is no 'apology' speech act in the current annotation scheme, but ***expressRegret*** is used instead, as a more general speech act whose function is to express regret, whether or not this means that the speaker is accepting blame. C-units – other than <dm>s – that contain the phrases *I'm/we're sorry, I/we regret* or simply *pardon* or *sorry* are likely to be labelled as ***expressRegret***, except where Notes [a] and [b] indicate otherwise.

> Note [a]:
>
> *Sorry* and *pardon* are often used interrogatively, asking the other speaker to repeat what he or she just said. These utterances are classed as a separate speech act, known as ***pardon***.
>
> Note [b]:
>
> The following type of utterance is not classed as an ***expressRegret***, but as an ***inform***: *I'm afraid there's no connection for that service*. Here the regretful expression *I'm afraid* is considered incidental to the main point of the utterance, which is to convey information. On the other hand, *Sorry {#} there's no connection for that train* is split into two speech acts: *Sorry* is an ***expressRegret*** and the rest of the utterance is an ***inform***.
>
> Note [c]:
>
> *Sorry* (or *I'm sorry*) as an ***expressRegret,*** when it occurs at the beginning of a turn, is sometimes a kind of apology for a problematic situation, or even for the fact that the speaker is talking at all! E.g. *Sorry, I meant to call the Gas Board* (at the beginning of a dialogue). This is treated as an ***expressRegret***, not an ***init***. In such a case, the C-unit label is <frag> or <decl> rather than <dm>.

6.18 expressWish

If a declarative contains an expression of the speaker's volition, but is not identifiable as a directive or an ***inform***, it should be labelled as ***expressWish***. This is most commonly illustrated by utterances beginning *I want/wish* or *I'd like/prefer* (also *we want, we'd like*, etc.).

ExpressWish is used as a last resort, whereas other labels, especially ***inform*** and ***direct*** are preferred if they can be appropriately be applied. For example, a customer who is asked *For which service do you want to purchase a ticket* may reply *I want a ticket from London to Stafford* (optionally adding *please*). It is clear that the agent is seeking to know what service can be performed (***reqDirect***) and the customer is replying by giving the details of that service (***direct***). But in *I just wish the trains were more punctual* it's clear that the caller is not attempting to get the service-provider to do something about it.

6.19 greet

This is an opening greeting formula. It may be a single word or phrase such as *Hello, Hi,* or *Good afternoon.* It may precede or follow a self-identification of the speaker. E.g.

A: ***Good afternoon*** *(=**greet**) Virgin Trainline Julie speaking (=**identifySelf**)*

There is sometimes a use of *Hello*, etc, in the middle of a phone-call, where speaker x wants to re-establish contact with speaker y after some kind of interruption of the connection. We treat this 're-greeting' as just a variant of ***greet***, even though it might be argued that its function is not so much greeting as checking that the channel is open.

> Note:
>
> *Excuse me,...* at the beginning of a dialogue is classed as a ***greet***, not an ***expressRegret***. It is a dialogue-opener.

6.20 hold

The speaker tells the hearer to hold the line, or else implies that the hearer should hold the line while something else is done. This is usually a kind of directive, but is not labelled '***direct***' because it is an important part of the mechanics of dialogue control, for which formulaic expressions are often used. E.g. :

Just hold the line a moment.   Bear with me a minute please.

Wait a minute.   Hang on while I check the times.

If you wouldn't mind holding a minute.

> Note:
>
> Notice that ***hold***s are often followed by a minimal response from the other speaker – e.g. *Yeah, Okay*. If the ***hold*** has a kind of directive force (as in usual), these are labelled ***accept***, rather than ***ackn***.

6.21 identifySelf

This label is only used if a speaker identifies him/herself or the company or institution he/she represents, as an opening move in a dialogue. The speech act is often virtually obligatory in telephone service-providing dialogues. It is highly formulaic, and normally only occurs at the beginning of a dialogue (usually after the ***greet***). In the BT data, however, it may also occur

when the operator contacts another service provider for information or transfers the caller, so that in effect a second dialogue begins in the middle of the same telephone call.

> Note:
>
> If a speaker requests information about the addressee's name in the middle of a dialogue, this does not lead to an ***identifySelf***. Thus responses to: *What's your name again?* or *Could you spell you name please?* or *Who's that speaking?* are labelled simply as ***answ***.

6.22 inform

This is normally conveyed by a <decl> or (less frequently) a <frag>. In fact in the case of a declarative C-unit, ***inform*** is the most neutral label, which is generally applied if no other label is appropriate (but see ***raiseIssue***). Typically speaker x has the goal of informing speaker y about something speaker x believes that speaker y did not know or was not aware of before, generally without this having been elicited. E.g.:

> *The last train leaves at 16 50.*
>
> *Someone left the phone off the hook.*

A 'partial inform', where part of the message is new information and part has already been given, is also included under this speech-act type. On the other hand, if the <decl> recapitulates or summarises something that has been **expressed** or **taken as understood** earlier in the dialogue, giving an opportunity for the addressee to signal assent or dissent, this is not an ***inform***, but a ***confirm***.

The borderline between ***inform*** and other speech act labels associated with declaratives can be difficult to determine. Labels with potential overlap with ***inform*** include not only ***confirm*** but ***expressRegret***, ***expressWish***, ***expressPossibility***, ***expressOpinion***, ***informIntent***, ***raiseIssue***. ***Inform*** has ended up being a flexible label, used where some element of **conveying information** or **making the addressee aware** is present. For example, after a longish period in which the telephone is ringing, the operator may say to the caller: *I'm sorry, **there's no reply***. The part of this example in bold (*there's no reply*) is labelled ***inform***, although it is fairly certain that the caller has already reached that conclusion. The argument here is that, although the information content of the utterance in context is presumably very low, the operator is making the caller explicitly aware of something, and so (as a borderline case) this is labelled as an ***inform***.

> Note [a]:
>
> An informative answer to a request for information is not an ***inform***, but an ***answ***. E.g.
>
> *A: Good afternoon. **Can I help you?*** (-- This is an ***offer***)
>
> *B: Yes. Hello. **My phone line has gone dead**.* (– This is an ***inform***.)
>
> *A: What kind of fault is it?* (-- This is a ***reqInfo***)
>
> *B: Well, **my phone line has gone dead.*** (-- This is an ***answ***.)

Note [b]:

Where a speaker, in responding to a *confirm*, actually offers some new information in addition to what would have been conveyed by 'Yes' or 'No', this is considered to be an *inform*; e.g.:

> A: *You just need the credit card's address(?)* (This is a *confirm* when uttered by the customer)
>
> B: *We need the full credit card details.* (This is an *inform*)

Note [c]:

*Inform*s sometimes occur in sequences. For example, where a speaker is giving a long piece of information which subdivides into several C-units, this is labelled with a separate *inform* for each C-unit. Also, *inform*s can be used to label digressions, where the speaker goes off at a tangent, but is nevertheless conveying (some) new information. E.g. *I'll have to call you back about that.* (On other digressive speech acts, see *unclassifiable*, *raiseIssue*).

Note [d]:

In a complex utterance, where a subordinate clause is given separate C-unit status, the subordinate clause should also be labelled as an *inform* if it simply adds to or qualifies the information in the main clause. E.g. *if the train arrives more than an hour late (=inform), you can reclaim part of the fare (=inform).* (Here the semantic function of the subordinate clause, as conditional, is signalled by the mode attribute-value 'condition' in the C-unit label, rather than the speech act attribute.) However, where the subordinate clause is subdivided into two coordinate constructions, it is not further subdivided into two *inform*s: e.g. *if the train arrives late and you miss your appointment (=inform), you can reclaim part of the fare (=inform).* On the other hand, in other cases the *if*-clause can be appropriately labelled *suggest*: e.g. *If you phone up the faults line, the engineers will handle it.*

Note [e]:

In this annotation scheme, there is no speech act label for a repetition, where the same speaker reiterates the same piece of information more than once. Hence where this happens (for example, where the addressee fails to hear or understand what has been said, we use the *inform* tag for the repetition as well as the original utterance. The rationale behind this is that, although the information has already been uttered once and in that sense is not 'new' information, the information has not yet become 'common ground' between speakers x and y, and hence the repetition can be regarded as another attempt to convey the same information 'anew'. Telephone dialogues can be full of redundancy, and there are occasions where the same piece of information is offered as many as six times in the same dialogue. Here the speaker may repeat not just to ensure that the information is properly received and understood, but to emphasise the need to take the information seriously and to act on it. Nevertheless, the same label *inform* is used for each repetition.

6.23 informIntent

The speaker informs the addressee of something the speaker (possibly with the help of others) intends to do. E.g.

*Okay, **I'm going to book you on the 10.30**.*

I'll put you through to our faults department.

Let me just check up on that.

21

Let's have a look. (Depending on context, this can be a **suggest** or a **informIntent**. But in our service dialogues, *Let's...* will often mean 'I intend to do such-and-such right away'.)

Usually this speech-act label means that the service provider is going to do something on behalf of the customer. This speech act is similar to an offer, except that it normally has a <decl> form and does not 'consult' the addressee's wishes at this point. It is more like an *undertaking* than an *offer*. This is logically a subcategory of **inform**, but it has a special role in service dialogues, as illustrated by the examples above.

> Note [a]:
>
> Statements with *I want* or *I wanna* are not **informIntent**s, but typically **direct** or **expressWish**.)
>
> Note [b]:
>
> Some **informIntent** moves have the opening expression *I'm trying to...*. Unlike those illustrated above, these are usually spoken by the customer rather than the service-provider, and imply that there is some difficulty.
>
> Note [c]:
>
> An **informIntent** cannot refer to an action already done. For example, *That's booked for you* refers to a completed action by the speaker. This is labelled simply as an **inform.**

6.24 informIntent-hold

In this dialogue-controlling compound speech act, the speaker tells the addressee to hold the telephone line while the speaker does something (e.g. consulting a database or timetable). E.g. *Just bear with me while I check that.* However, when the operator hands the caller on to another telephone line, saying something like *I'm putting you through*, this is not treated as an **informIntent-hold**. Even though there may be a pause after such a move, the pause is due to another operator's delay in answering, and the speech act does not constitute a request to hold the line. *(I'm) putting you through* is classed as an **informIntent**.

As noted above, the NXT version of the corpus uses two attributes on segment elements for this information:

```
<segment sp-act=informIntent' secondary-act='hold'>
```

6.25 init

This speech act is associated with discourse markers such as *Now, Well*, and *So,* which indicate an 'initialization' (or initiation) of a new part of the dialogue – e.g. beginning a new topic or sub-goal of the dialogue. (Notice that *now* and *well* also have other functions as adverbs of time and manner, so that potentially they could be misidentified. For example, *Now?* and *Well?* might occur as one-word-question, expressing incredulity or exasperation.)

<u>Note:</u>

Okay or *ok* should be labelled **init** (not **ackn** or **accept**) where its function is to signal that a new stage of the dialogue follows, rather than to signal a response to an earlier turn. A particular case to be labelled **init** is where *okay* is used as a 'pre-closure' signalling that the dialogue is about to come to an end. *Okay* in such a case is often spoken with an interrogative intonation, and means roughly 'I'm just checking that you are satisfied with the outcome of this dialogue, before going on to say goodbye.' In the absence of intonation evidence, we base the decision on interpretation of the context. Where the *okay* is followed by a **bye,** this is likely to indicate that the *okay* is actually 'initiating' the ending of the dialogue.

6.26 negate

This label applies a 'no' answer, such as *no, no way, I'm afraid not*. But other labels, such as **answ**, are also frequent in that negative function. There are however occasions where **negate** is the only label one can reasonably use: that is, where the other response labels **answ, ackn,** and **refuse** are for various reasons inappropriate. This is particularly so where a speaker negates, denies, or disagrees with the proposition of a previous declarative, as in A: *the train leaves at 10 45*. B: **no, (=negate)** *it leaves at 10 55*.

The **negate** label can also be used where a <decl> C-unit is used to deny or disagree with something the previous speaker stated. E.g.:

A: *You ring up every night and ask this.* (**raiseIssue**)

B: *I don't ring up every night.* (**negate**)

Negate is used in effect as a 'last resort' label, where no other label is appropriate for a negative response.

6.27 offer

A speech act in which the speaker proposes a future action the speaker can undertake for the hearer's benefit. It is assumed that the hearer has the right to accept or decline the offer. E.g. *Would you like me to put you through? Shall I repeat that? Do you want me to book you on the 10.34?* This is distinct from an '**informIntent**' where the speaker does not consult the wishes of the hearer at this point, but simply declares an intention to do something. Equally, an *offer* is not the same as a *suggest*, which proposes action involving the hearer as an agent. Standard responses to an offer are **accept** (a positive response) and **refuse** (a negative response).

Note [a]:

There is sometimes a fine line between an **offer** and an **informIntent** (which is characterized as an 'undertaking' rather than an offer). Compare:

1. *I can book it for you.* (**offer**)

2. *I can put you through to directories.* (**offer**)

3. *I'll book the super-advance for you.* (**informIntent**)

*4. I'll book the super-advance. (**informIntent**)*

The *for you* is a tell-tale sign of an offer. The *I'll...* is a tell-tale sign of an **informIntent**. If they are combined, as in 3 above, we have something which is a 'hybrid' of an offer and an **informIntent**. For the sake of having a consistent convention, we treat this as an **informIntent**, because it does not seem to allow the hearer an option of refusal.

Note [b]:

There are also borderline cases between an **offer** and an **inform**. A service-provider may say:

I've got one [i.e. a discount seat booking] *from Taunton to Crewe.*

Although the motivation of offering to book this journey may be behind the speaker's utterance, there is no mention here of a future action of benefit to the hearer – and so this does not qualify as an offer: it is just labelled **inform** instead.

Note [c]

Generalized **offer**s are quite common at the beginning of a dialogue: *How can I help you (today)? Can I help you?* etc.

6.28 pardon

A general request for repetition, where speaker x could not hear or understand what speaker y said. E.g. *Pardon? What? Eh? Sorry? I beg your pardon? What did you say?*. (Note: on *Sorry* as a pardon vs. *Sorry* as an expression of apology, see 6.17.) The speech-act **pardon** does not apply to 'echo questions' where speaker x asks for a clarification, through repetition, of **part** of the previous turn: e.g. *A: The train arrives at Warrington Bank Quay at 10.50. B: **It arrives at <u>what</u> time?** In this case B's utterance should be labelled as a **reqInfo**, because the speaker is signalling the need for an extra piece of information.*.

Note:

Since the information requested has already been given, it might be thought that this should not be labelled **reqInfo**. However, the important point is that B has not understood part of what A said, and so this is in effect a request for information new to the speaker.

6.29 raiseIssue

This is normally an initiating move in the dialogue, where a speaker starts a new topic which does not function as a regular part of the transactional purpose of the dialogue, but where nevertheless the issue raised by the speaker in not irrelevant to that transactional purpose, and in the mind of the speaker presumably needs to be discussed as a pre-requisite or adjunct to that purpose. E.g. *I hope you're not going to cut me off* (spoken by a caller to the operator): the issue of 'cutting off' is clearly a concern of the caller, relevant to achieving the caller's purpose in phoning the operator. But it doesn't actually contribute to the goal of the dialogue as such. Another example is *You ring up every night and ask me that*, spoken by the operator, which is a kind of complaint or accusation about the caller's repetitive behaviour, and which seems to

demand an explanation from the caller. In general, complaints are a subclass of ***raiseIssue***. Perhaps a useful point is that with ***raiseIssue***, the speaker has a different agenda from the current agenda of the addressee.

RaiseIssue utterances are normally in declarative form, but they differ from ***inform*** in that they do not convey any new information that was previously unknown to speaker y, except that they inform speaker y that something is a matter of concern for speaker x. Similarly, they differ from ***confirm*** speech acts, which summarise or restate what has been said or what has been assumed to be common knowledge between the speakers. ***RaiseIssue***, like ***unclassifiable***, is a label used as a 'last resort', when more regularly occurring speech acts such as ***inform, confirm,*** or ***expressOpinion*** do not apply. Thus if a speech act could be classified either as ***inform*** or as ***raiseIssue*** (i.e. where the speaker is offering information at the same time as raising and issue), we choose ***inform***.

6.30 refer

The basic idea of ***refer*** is that the speaker is referring to some piece of common ground between speaker and addressee.

This label applies typically to a fragment such as a prepositional phrase, a noun phrase, or an initial adverbial clause, when it occurs alone, or perhaps as part of a larger sentence, and refers to an element of information relevant to the task, such as place or time information. E.g. *To Birmingham International. Tuesday 9th October*. The label ***refer*** is not used where the fragment concerned is directive or interrogative (although in our data this can be discovered only by looking at the context), or where it is an answer to a question: for these, the normal labels are ***direct***, ***reqInfo*** (or ***reqDirect***) and ***answ*** respectively.

> Note [a]:
> ***Refer*** also applies where a <frag> which is a fronted topic or 'prefatory phrase' occurs before a following more substantial C-unit (often one with a finite verb). For example: ***That handset {#} it doesn't work***.
> Note [b]:
> Another type of ***refer*** is the presuppositional specification of time found in an initial adverbial clause:
> ***When I contact you***, (= ***refer***) *can we actually fix a date?* (= ***reqInfo***)

6.31 refuse

This is a negative response to a speech act such as a directive, offer, proposal, or suggestion, where the speaker indicates non-compliance with or non-acceptance of what the hearer proposes. ***Refuse*** in this respect is the negative counterpart of ***accept***. E.g. *Do you want to book that now? No thanks*. A ***refuse*** is not always impolite in tone: many refusals are cases where a speaker 'politely declines' an offer or invitation.

It may seem strange for a refusal to be immediately followed by an acceptance. However, this is the way we label a response such as the following, where speaker B declines an offer, immediately adding a remark such as *it doesn't matter* to signal that the offer was 'accepted in good part':

A: *Would you like me to put you through?*

B: *No (= **refuse**), it's okay/alright (= **accept**)*

6.32 reqDirect

The speaker requests a directive or instruction from the hearer. This is a common speech act towards the beginning of service dialogues, where the service-provider wants to know what service the customer is seeking. E.g. *For which train do you wish to purchase a ticket?* It may also recur further on in a dialogue again, when the initial information gathering phase has been completed and the operator is now requesting an explicit instruction for the particular action to be taken, e.g. *Do you want me to book this for you now?*.

A request that does not elicit the wishes, intentions or instructions of the other speaker should be labelled *reqInfo*, even though it appears to have the underlying motivation of requesting instructions. E.g.: *Do you need any other service?* Similarly, in this context the response will be an *answ*.

A response to a *reqDirect*, on the other hand, is a *direct*, even if this response is a simple 'yes' or 'no'. The 'yes' or 'no' here amounts to an instruction for the other speaker to do something, or to refrain from doing something. For example:

A: *Do you want me to check the number? (= **reqDirect**)*

B: *Yes please. (= **direct**)*

Note [a]:

We do not use *reqDirect* for an utterance such as *A's* below, where the operator is asking for further information, after the caller has requested an early-morning wake-up call.

A: *And that's to the number you're calling from now, it is? (=**reqInfo**)*

B: *Yes (=**answ**)*

Although this could be seen as a request for more precise instructions, it is easier and more straightforward to treat is as a request for further information.

6.33 reqInfo

The speaker asks the hearer to supply information. This is the most common and neutral speech act function of questions. E.g. *How old are you? Do you have a railcard?* There is a potential confusion between *reqInfo* and *direct*, because both may be described as 'requests', and both may be accompanied by politeness markers such as *please*. But the distinction is made on the

basis that **reqInfo** specifically seeks to elicit a linguistic (verbal) reply, while **direct** seeks to elicit a response which is defined more generally, to include non-linguistic (physical or mental) behaviour.

Another potential confusion is with the speech act **reqDirect**. Note that a question such as *and how many people's travelling?* is considered a **reqInfo** because it doesn't explicitly ask for a directive, such as *how many people would you like me to book for?* and therefore doesn't qualify as a **reqDirect**. An important (though rather arbitrary) criterion for distinguishing between **reqInfo** and **reqDirect** is that presence of an explicit expression of the wishes of the hearer signals the **reqDirect** interpretation. In the absence of such a signal, we choose the **reqInfo** interpretation. E.g.:

A: When would you like to return? (=reqDirect)

B: After six o'clock. (=direct)

A: Is that on Monday the fifteenth again? (=reqInfo)

B: Yeah. (=answ)

Note [a]:

A <decl> which ends with a tag question is typically labelled as a **reqInfo** – e.g. *That's in the morning is it?* This is because such utterances normally elicit a response which is functionally similar to an **answ** following a yes-no question. However, occasionally such tag questions are not intended as questions, but may (for example) be **expressOpinion** – e.g. *It takes rather a long time, doesn't it?* In this case, question intonation would not be used, and it is clear the speaker is sharing an opinion, rather requesting information. A 'yes' response to this is an **accept** (i.e. agreement with the sentiment expressed) rather than an **answ**.

Note [b]:

Notice that **reqInfo**s can be conveyed by a wide range of types of C-units. Most are in the form of questions – e.g. *How much does it cost? Can I (be a pain and) ask you....* – others are in the form of statements or imperatives – e.g. *I'd like to know when..., I wonder if you'd mind telling me ... Just repeat your postcode, please.* Notice also (as in the last example) that **reqInfo**s, as well as **reqDirect**s, can contain a *please*.

6.34 reqModal

This label is used with a speech act which is a request for permission, or for suggestions, or for advice: e.g. *Can I have a word with the operator? Should I contact the police?* For permission, the modal verbs used in such cases of *can*, and (less often) *may* or *could*. For suggestions or advice, the modals *shall* and *should* are sometimes used. However most questions containing modal verbs, even these modal verbs, have other speech-act labels: **reqInform**, **reqDirect**, or **direct.** The **reqModal** label is used only 'as a last resort', when these other labels are not applicable.

6.35 selfTalk

This occurs where the speaker opts out of the dialogue (so to speak), and addresses a remark to himself/herself. E.g. *Where did I put that pen*. The grammatical form of the self-addressed remark is not relevant: it could be a <decl>, <q_yn>, <q_wh>, <imp>, or <frag>. The relevant point is that it is not part of the communication between the different speakers, speaker x and speaker y. A useful diagnostic of a *selfTalk* speech act is that the other speaker does not respond to the force of the self-talk. Thus although *Where did I put that pen* poses a question, the other speaker does not respond to it with an *answ*.

6.36 suggest

In this category are included (often tentative) proposals made about the future. *Suggest*s propose some future action or decision by the hearer or by the speaker and hearer together: E.g.

> *Let's check the numbers again.*
> *You could take a later train, if you wanted.*
> *Why don't you telephone our service department?*
> *Maybe we'd better do it now.*
> *How about Tuesday afternoon?.*

Suggests stand both for making suggestions and offering **advice**: they can, therefore, be fairly 'certain' in tone, as in *You need directories on 192*. This recommends that the addressee take an action for the **addressee's own benefit**. Notice *suggests* are distinct from '*offers*' like *Let me check the times for you.* or *Shall I book that for you?* or *Would you like me to connect you?* where some kind of helpful action by the **speaker** is being proposed.

> Note:
> An *if*-clause is sometimes labelled as a *suggest*. E.g.
> > *If you phone the fault line* (=*suggest*) *they'll sort that out for me* (=*inform*).

6.37 thank

A speech act thanking the other person. E.g. *Thanks. Thank you* (*very much*). *Thank you for the advice*. When one of these expressions comes on the end of a longer C-unit, as in *No thanks,* or *I'll go for the cheaper one thanks*, it is not given the *thank* label. Some label such as *refuse* (in the first case) or *informIntent* (in the second case) is applied.

6.38 thank-bye

The compound speech act *thank-bye* is used if **thank** and **bye** are run together in a single utterance without a pause or other break: *Thanks bye bye*, or (less commonly in the opposite order *bye thanks*). This is a closing move sometimes made by a customer after a call to a service-provider. The word *Cheers* at the end of a phone call sometimes appears to combine both functions of thanking and saying goodbye, and may therefore be labelled **thank-bye**.

This is another case in which the NXT version uses the SECONDARY-ACT attribute:

```
<segment type='frag' sp-act='thank' secondary-act='bye'>
```

6.39 thirdParty

This speech act label is a 'blanket label' to cover all utterances addressed by one of the dialogue speakers to a third party, or (rarely) by a third party to one of the dialogue speakers (e.g. when a wife checks up with her husband on what he wants when booking train seats). Such 'third party dialogue' is regarded as peripheral to the dialogue annotation task, and so it is unnecessary to decide how far normal speech act categories such as **inform** or **reqInfo** are applicable to utterances involving third parties. This would be difficult or impossible, in any case, because the responses of the third party speaker are generally not provided in the transcription. The label **thirdParty** label can thus be used for all such utterances.

6.40 unclassifiable

With **uninterpretable** – see below - this is a 'last resort' choice of speech-act category, where no other more specific label can be reasonably assigned. Unlike **uninterpretable** speech acts, **unclassifiable** ones **are** interpretable. This label can be used, for example, for a throw-away remark or joke which does not contribute to the transactional nature of the dialogue, or where the speakers engage in conversation or back-chat about topics which are irrelevant to the purpose of the dialogue (e.g. chat about the weather). But if such a throw-away remark can be assigned a more explanatory label such as **inform** or **raiseIssue**, then this ought to be preferred to **unclassifiable**.

6.41 uninterpretable

A speech act which cannot be assigned any other value, because its interpretation is unclear. This could be because it is incomplete (i.e. the speaker didn't finish saying what he or she intended to say), or because it is seriously disfluent, or because the transcriber couldn't make sense of what was said. Abandoned C-units (i.e. those which the speaker fails to complete) are often labelled

uninterpretable. (The mode label **abandon** is assigned at the same time.) On the other hand, an incomplete C-unit can often be assigned to a regular speech act category, even if its total meaning cannot be deciphered. For example, *Could you possibly...* is obviously 'trying to be' a *direct*, and so should be labelled as **direct**.

If, through some gap or fault in the transcription or the XML coding, a turn is empty, i.e. contains no words uttered by a speaker, this is labelled ***uninterpretable***, on the grounds that it is better to assign some label than to leave the speech-act attribute empty. This situation occurs especially at the end of some BT dialogues where a turn contains no words except for some words or symbols in curly brackets – particularly at the end of a phone call: e.g. '{end_ok}'.

7. Supplementary speech act labels: *-disc* and *-cont*

In addition to the speech act categories in Section 6 above, there is need for two supplementary labels which are hyphenated to other speech act labels to signal an interruption: **disc** ('discontinued') and **cont** ('continuation'). These are designed to deal with the rare occasions when one speaker x interrupts another speaker y, in such a way that x's turn cannot be treated simply as a backchannel (see 1.1); thus y's turn continues after the interruption by x, in fact resuming the same speech act that y started before the interruption. For example:

A: *Do you want a forward-facing seat (=**reqDirect-disc**)*

B: *Yes thanks (= **answ**)*

A: *or a backward-facing one?(= **reqDirect-cont**)*

In cases like this, the two separated parts of the same C-unit have to be labelled as two separate C-units. But, to show this interruption, the first part is labelled (say) ***reqDirect-disc***, and the second part is labelled ***reqDirect-cont***. (We here assume, by way of exemplification, that the speech act category is ***reqDirect***. Other labels with these supplements can also occur: e.g. ***inform-disc*** and ***inform-cont.*** Any speech-act label or the form ***X-disc*** should be followed, after an intervening piece of dialogue, but ***X-cont***.)

In the NXT version of the corpus, the *–disc* and *-cont* parts of the values were replaced by DISC and CONT attributes on the relevant segments. So ***reqDirect-disc,*** for example, would be:

```
<segment sp-act= 'reqDirect' disc='yes'>
```

8. Splitting turns into C-units

The unit 'sentence', which is commonly regarded as the smallest syntactically independent unit of written text, does not apply to spoken language. C-units are used instead. In dialogue, C-units are independent pieces of dialogue, which are not syntactically included in a larger unit. They

may be (a) an independent clause (with subject, verb, object, etc.), (b) a non-clausal unit, that is, a piece of language capable of being syntactically recognized as a word, phrase etc., but without the markers of a finite clause – in particular, a finite verb, (c) a dependent clause – including a non-finite clause – and (d) a discourse marker/particle such as *well, okay, right*, *yes, no*. Since discourse markers and particles can easily attach themselves prosodically to preceding and following C-units, we also allow a fifth type of C-unit: (e) one of the types (a)-(d) augmented by one or more discourse markers and particles. This will be allowed where the discourse markers or particles are not separated (e.g. by a pause) from the preceding or following C-unit.

Fillers such as *um* between yeses/nos/discourse markers and other C-units are kept with the following unit.

Occurrences of *Yes/No* that are repeated (e.g. *no no no*) are combined into one utterance rather than treated as different C-units. A similar case is *ok alright* – where the two consecutive <dm>s express the same kind of response, and are therefore run together into a single C-unit.

9. Disfluencies in relation to splitting

9.1 Pauses and overlapping speech

Pauses and overlapping speech are marked respectively *{#}* and [...] in the examples at (A) below. These can be considered potential forms of disfluency, but as such they do not affect the splitting of C-units. For example, *It's {#} gone dead* is a single C-unit, with the form label <decl>. We recognize that this is a disfluent pause (rather than one which comes at the end of a C-unit) because the syntax of the clause *It's gone dead* is interrupted in mid-flow by the pause. Note that disfluencies were extensively marked up in the jointly annotated section of the corpus.

9.2 Repeats

Repeats are one kind of disfluency which raises issues of where to split a turn into C-units. A disfluent C-unit is not split if the speaker retraces his/her steps and 'repairs' an utterance by repeating the same word (sequence) or completing a previously unfinished word and then goes on to finish the utterance. These are varied examples of repeat disfluencies of this kind:

> *and i can't get any s… {#} satisfaction*
> *what's {#} what's it doing*
> *[i i] didn't*
> *i didn't] i didn't get the number*
> *all we can dg… keep doing is keep trying*

31

Notice that sometimes a repeating disfluency is complex, with more than one repeat combined:

{have { y you} {have have} you} checked that number recently.

Here there is embedding of one repeat in another, as is shown by the curly brackets *{ }*.

There may also be cases where we would normally use a different C-unit label, such as for a *yes*,
but where this does not occur initially and does not have an accepting or acknowledging force,
but only a phatic one, e.g. *we'll just **yeah** we'll just [...].* In such cases, we again do not allow a
split, as the disfluency can still be classified as a 'repeat'.

9.3 False starts

Another kind of disfluency occurs at the beginning of a C-unit, where the speaker starts to say
something, then re-starts, with some change in the form of the utterance:

we're just trying to we're trying to get something sorted;

well well for eg… well for five o'clock[1].

In cases like this, we do not split the false start from the remainder of the utterance, so long as
the false start belongs to the same form category. For example, above *we're just trying to* and
we're trying to get something sorted are both declarative in form. So we treat them as comprising
a single C-unit with the form category <decl>.

The same non-splitting decision applies in the following cases:

(1) ***i've got i been*** *me telephone is causing me an awful lot of trouble*.

(2) ***can you*** *can i have the number for the international operator please*;

Example (1) begins with two false starts: *i've got* and *i been*. But both of these are in <decl>
form, like the following utterance *me telephone is causing me an awful lot of trouble*. In
Example (2), the speaker begins with a false start in <q_yn form>, and also continues with a yes-
no question *Can I have....* Hence here again, we avoid splitting.

On the other hand, if the false start has a different syntactic category from the continuation, there
has to be a split between them. For example:

I'd like to er | *would you mind calling back later*.

Here a declarative false start is followed by an interrogative continuation, and so the sequence
should be split into two C-units at the place indicated by |. The speech act label will be
expressWish for the first C-unit, and ***direct*** for the second. The first C-unit is also given the
mode value "abandon", since it is left incomplete.

---

[1] assuming that *eg…* here stands for *example*

Splitting should also occur if there is a backchannel from the other speaker at the point where the false start finishes: e.g. (*sorry*) in this example:

 i had to go for a long way for to make (*sorry*) | *i had to go for a long way to make that call.*
Both C-units here are ***inform***s, but they are treated as separate units.

9.4 Reformulations in the middle of a C-unit

However, no split should occur if there is a disfluency (in the form of a retrace-and-repair sequence) in the middle of an utterance. For example, in the following, the part in bold is similar to a false start, except that it begins and ends in the **middle** of an utterance. If the bold section were deleted, the whole would make good sense as a C-unit, and therefore no split is needed.

> *it's a number **in theg**… in this country*;
>
> *is there **any other number that i** any other way i can get through to them*;
>
> *there's no way i can **d…** return that is there*;
>
> *and then **they w**… they're able to help you.*

9.5 Prefaces like *I mean, you know*, and *I wonder*

Note that cases like the following are not treated as false starts:

   a) ***I mean*** *will someone come....*

   b) ***You know*** *he hasn't called back...*

   c) ***I wonder*** *would you mind if....*

These are more like prefatory formulae introducing the following speech act, and are treated as part of the same C-unit as what follows. The form category in these cases is what would be appropriate if the preface *I mean* etc. were removed. That is, for a) it is <q_yn>, for b) it is <decl>, and for c) it is again <q_yn>.

9.6 *Thank-bye*

At the end of a dialogue, speakers sometimes combine ***thank*** and ***bye*** speech acts as a way of bringing the transaction to an end. Note that the utterance *thank you bye* or *thanks bye bye*, etc., does not need to be split into two separate C-units. A compound speech act *thank-bye* is used if the two speech acts are run together without a break, and as noted earlier, this is split in the NXT version into a speech act and a secondary act using two attributes.