Abstract

This document explains an annotation schema for interactions among two or more participants that applies across modalities and types of interaction. It also serves as an annotation manual. The goal is to provide a single annotation that can document the communicative flow for many types of interaction. It relies on a reduced set of dialogue act types, groups together into a single Dialogue Function Unit (DFU) a sequence of dialogue acts of the same type that have a common intention (analogous to a plan), and places a special emphasis on the analogue of adjacency pairs, which we capture as links between DFUs. The initial version of the annotation manual was developed through analysis of examples from three contrasting corpora: Enron email threads, Switchboard telephone conversations, and the Dover trial transcripts. This version (v 2.0) has been updated as a result of annotation of 21 thousand words of a library-patron telephone transaction corpus, and 18 thousand words of Enron email threads.
1 Introduction: Goal of this Annotation

There has been much work on dialogue annotation. The prime area of interest has been the annotation of dialogue acts (DAs), to label the communicative function of each utterance. The repertoire of DAs used within any given labeling schema is usually related to Austin and Searle’s writings on speech acts [1, 5]. In addition, based on the 1970s work of Sacks and Schegloff [4], DAs are seen as forming adjacency pairs which represent a limited set of types (such as question-answer).

Many dialogue act annotation languages in use in computational linguistics have developed a large set of dialogue act types (but cf. [2]). The goals in applying these annotation languages to corpora are to collect quantitative data on the types of dialogue acts speakers use in various types of interaction, to enrich our understanding of dialogue, and to determine what dialogue acts a dialogue system should understand and produce. Very little work has been done on documenting the distribution of adjacency pairs, or relying on such data to guide dialogue strategy for spoken dialogue systems.

Our annotation follows in this tradition, but we concentrate on issues which have not received much attention. Our focus is primarily on the adjacency pair, specifically, on the relation of an initiating utterance in an adjacency pair to the responding utterance, and to cases where one member of an adjacency pair is missing or is an unanticipated type of dialogue act. We say that the initiating DA has a forward link (or flink) to the responding one, which in turn has a backward link (or blink). As a corollary to our focus on the adjacency pair, we aim to minimize the repertoire of dialogue act types we use.

In the standard case, when a discourse evokes a flink, it will eventually be paired with a blink evoked in a subsequent part of the discourse. An example would be a typical adjacency pair of a question and a response. However, when a dialogue utterance evokes a flink, a corresponding blink is not obligatory, and sometimes does not occur. Similarly, a blink can occur independent of a pre-existing flink. We are interested in understanding these three types of contexts.

1. How can we break up a turn in order to identify a unit consisting of one or more DAs where the unit evokes a flink? We call these units Dialogue Function Units or DFUs.

2. How can we capture the potential response to a DFU, independent of whether the expected response occurs? How can we capture what part of the prior discourse a DFU responds to, independent of whether the prior discourse has generated an obvious expectation for a response?

3. How can we relate the DFUs to the evolving structure of the interaction, and
in particular to its “outcome” - such as whether a transaction is completed successfully. What role does DFU structure play in accounting for the distribution of flinks and blinks?

We can look at relations among DFUs that consist of blink/flink pairs, DFUs that stand alone, and the overall structure of the interaction.

We are investigating these questions by examining contrasting types of interaction that have distinct types of outcomes, as well as distinct constraints on when and where openers and responses can occur. Our first annotation manual arose from an analysis of a small corpus exemplifying the first three types listed below. Concurrent with releasing this second version of the manual, our focus has shifted to annotating a larger number of interactions in the third and fourth categories below.

- switchboard conversations (telephone speech; non-task oriented; sequential DFU pairs);
- court transcripts (speech; task-oriented interaction with institutionalized rules for turn taking; DFUs embedded within DFUs);
- email threads from the Enron corpus (email; task-oriented interaction; concurrent DFU pairs).
- library patrons’ telephone calls to request recorded books (telephone speech; task-oriented; sequential DFU pairs with some embedding).

Our approach is partly inspired by an earlier annotation method designed to capture overlapping content in summaries [3]: the granularity is not given a priori but determined by the text(s) to be annotated. If a passage appears to be one part of a DFU pair, then it is a viable DFU, and we can also give it a short description which the annotator uses as justification for his or her choice.

2 Motivation

The goals of our annotation effort are to enable research into the following questions:

1. What are the differences in the organization of interactions that vary in modality, genre, etc.? What are the similarities in the DA to DA links across these modalities and genres?
More specifically, for spontaneous face-to-face dialogue among peers, adjacency pairs are generally sequential, turns can be quite simple, the interaction is mixed-initiative, and implicit norms govern turn taking. For other types of interaction the norms can be quite different, e.g., for adjacency pairs to be non-sequential (email), and for institutionalized norms to govern turn taking (court transcripts). For complex turns, or for shifts in focus, a higher level of structure is required, along with predictions about what’s open to the interlocutor and with what level of effort.

2. To develop models that can identify the interactions in extended dialogs even in modalities where they are not simple to detect. This is useful for information extraction and summarization.

3. To develop a predictive model of what gets responded to how, meaning when interlocutor1 says X, what range of behaviors are more likely for interlocutor2. This model should also cover modalities and genres where adjacency pairs are not simple to detect.

3 Corpora

3.1 Corpora for Initial Investigation

Before beginning to design our annotation, we reviewed several dialogs and selected three that contrast in ways that we want our annotation to capture. The three dialogs we selected to annotate first are:

1. Enron email thread (ETh)
2. Switchboard discussion of judicial system and sentencing (SW)
3. Court transcript of the Dover trial on intelligent design (available from the ACLU Pennsylvania website\(^1\) (DT)

The SW discussion is spontaneous conversation, but the topic was “assigned” for the purposes of collecting speech data, and was not generated by the interlocutors. As a consequence, the dialogue topics have a wandering quality, and the overall dialogue lacks the focus that the other two interactions have. This is reflected also in certain types of utterance functions we observe that do not occur in the highly focused task-oriented ETh and DT interactions. For example, there are

---

\(^1\)http://www.aclupa.org/legal/legaldocket/intelligentdesigncase/dovertrialtranscripts.htm
non sequiturs, spin-off topics, completions of the other party’s utterance, and re-statements. Two other key differentiators are first that the interlocutors have equal freedom in the kinds of utterances they can make, and that the SW interaction is spontaneous (unplanned), real-time dialogue; thus, DAs follow each other in sequential adjacency pairs. This affords an opportunity for spin-off topics and other completions that the other interactions do not afford.

The DT interaction differs from both ETh and SW in having explicit “rules of the game” regarding who can initiate an adjacency pair (the lawyer questions the witness, not vice versa), and other rules for the interaction (such as what can be said, what can be objected to, etc.). It is far less spontaneous; both the lawyers and the witnesses prepare strategies and stances in advance, as well as specific speeches to make. However, we observed a very interesting sequence in which the trial attorney attempts to get a witness to say something that the witness refuses to agree to; thus, while the attorney has overt control over the interaction, the witness can use the “rules of the game” in a sense. Also, the witness is occasionally allowed to volunteer information. Edwina Rissland has spent a decade learning about and studying legal reasoning and argumentation, and has some papers and references she could share regarding the types of conventional “moves” lawyers use in constructing arguments (e.g., something she calls “slippery slope” which has to do with establishing where the boundaries are in deciding about a given instance, e.g., where the law clearly states that the state should not support one religion over another, and this can be used to bar a crèche display at a courthouse or other public building; however, what if the crèche is an ornament on a Christmas tree and is too small for the details to be discerned?).

The ETh interaction is less one-sided and argumentative than the court transcript, and more focused than SW. Like DT, it is task-oriented, but unlike DT, it is relatively spontaneous, and does not have “explicit” rules. Here, the different status of the two interlocutors is not so apparent as in DT; either party has freedom to initiate a wide range of DA types, but presumably, there is a status difference that is reflected less overtly in the content and form of the interaction. The most salient difference between an email interaction and the two face-to-face verbal interactions (SW and DT) is that the adjacency pairs are not sequential.

We developed our first annotation manual using the ETh interaction because it was the most complex and non-linear in the structure of the DFU pairs, and probably the most open-ended in terms of which “openers” (e.g., questions) the interlocutor chooses to respond to. If we can account for what’s going on in ETh at this level, then we can probably also account for DT and SW.
3.2 Phase 2 Corpora

For the next phase in refining our annotation scheme, we decided to focus on a set of ETh threads, and to select a task-oriented telephone-based dialogue corpus as a contrasting case.

We have suspended efforts on annotating more court transcripts; the court transcript we were using (DT) presented numerous unanticipated annotation complexities for our annotators.

We have also suspended efforts on annotating the switchboard corpus, due to its artificiality. Speakers were recruited, did not know each other in advance, and were assigned topics to discuss.

A telephone bandwidth speech corpus has obvious utility to the spoken dialogue community, so we have elected to annotate transcripts from the human-human real-world subset of the Loqui corpus (LCS1 for Loqui Corpus Subset 1), which is under development by one of the co-PIs.²

4 Annotation Scheme

This section contains the description of the annotation scheme and of the annotation procedure. All examples refer to the Enron thread shown in section 11.

4.1 Dialogue Acts

What action does this DFU perform with regard to the dialogue that it is part of? How does the utterance of this DFU change the worlds (cognitive states) of the speaker/writer and of the hearer/reader? Dialogue Acts (DAs) are always about the intended effect, not about the linguistic form. For example, a #1 can be a syntactic question (Where is Petaluma?), a statement I wonder where Petaluma is), or an imperative (Tell me where Petaluma is!), but in all cases the DA is the same.

• Inform: This DFU conveys information. This covers many different types of information that can be conveyed:

  1. Answers to questions Request-Information. A blink is obligatory to the Request-Information.

²This subset of the Loqui corpus consists of 82 anonymized calls made by patrons of the Andrew Heiskell Talking Book and Braille Library to librarians on Reader’s Advisor duty. Most borrowing (and other) transactions between patrons and the library are handled by telephone; items are shipped and returned by mail. The Loqui corpus will be released at the end of the Loqui project, circa 2010.
2. Elaborations on previous DFUs (by the same speaker/writer, or by another). A blink is obligatory to that DFU (which should also be an Inform).

3. Signaling that a request for action (Request-Action) has been completed. A blink is obligatory to the Request-Action.

4. A DFU in an email which signals that there is an attachment (or which points to information elsewhere, for example in a URL), and the attachment is the actual information. The attachment is not annotated, but the DFU indicating that it is included is an Inform.

Apart from the cases listed above, an Inform does not carry obligatory blinks or flinks with it.

- **Commit**: This DFU commits the speaker/writer to performing a (linguistic or non-linguistic) task.
  
  A blink is obligatory if this is in response to a previous Request-Information (“good question, I will find out the answer and get back to you”) or Request-Action (“sure, I will make coffee, no problema”).

- **Accept**: In this DFU, the speaker/writer accepts a proposal or a task (linguistic or non-linguistic) assigned or offered in a previous DFU.

- **Reject**: In this DFU, the speaker/writer rejects a proposal or a task (linguistic or non-linguistic) assigned or offered in a previous DFU.
  
  A blink is obligatory to the Request-Information (“I will not answer that question”) or Request-Action (“unfortunately, you will need to make your own coffee”) that is being responded to.

- **Backchannel**: This DFU is about the process of communication itself. There should be no blink, even in the first two cases below, where the backchannel seems to be signalling more than that the speaker/writer is merely attending. We have two types of Backchannel. A “plain” Backchannel, and a backchannel with grounding behavior; see section 7 below. Use Backchannel in these cases:
  
  1. The speaker/writer acknowledges the understanding or receipt of another participant’s communication, but no question has been explicitly or implicitly asked by the other participant.
  2. The speaker/writer signals that a previous communication was not received or understood.
3. The speaker/writer signals that the speaker/writer is listening/reading.

Note that some Backchannel communications are followed by a request for clarification; these requests are separate DFUs which should be labeled Request-Information.

- **Request-Information**: This DFU obliges the hearer/reader, or opens an option to the hearer/reader, to provide information (either facts or opinion), either in the dialogue or through another form of communication.

  Request-Information establishes a flink (which can remain unresolved if no one answers the request, or, normally, is paired with an Inform blink).

- **Request-Action**: This DFU obliges the hearer/reader, or opens an option to the hearer/reader, to perform some non-communicative action, i.e., an action that cannot in itself be part of the dialogue. Examples of such actions include writing a report (which may subsequently be sent by email), mowing the lawn, or making coffee.

  Request-Action establishes a flink (which can remain unresolved if no one answers the request, or, normally, is paired with a Commit blink, or an Inform blink, where the Inform conveys the completion of the task).

- **Perform**: These are classic performatives, such as swearing and firing: the action is achieved in the utterance itself (“You’re fired”, “I swear that I did not commit the crime”). Note that not all uses of performative verbs such as swear are necessarily performative. Perform also includes jokes or anecdotes, where the goal is not the conveying of information, but entertaining the hearer/reader (for example, by making him or her laugh). There is a fine line - sometimes, an anecdote both entertains and informs.

  A blink or flink is not obligatory.

- **Conventional**: These are greeting, introductions, expression of thanks, etc.

  A blink or flink is not obligatory.

- **NoDA**: This is used for text which does not have a dialogue act, such as signatures.

  There cannot be a blink or flink.

Often times, in email threads (but less so blogs or spoken dialogs) there is an act of communication which did not happen in the email thread itself. For example:
M1.1 It was good talking on the phone to you just now.
M1.2 The answer to your question is that Enron cannot pay anything now.

Here. M1.2 is clearly an Answer, but there is no question in the email thread. If this happens, add Offline to the speech act:

M1.1 It was good talking on the phone to you just now.
Conventional: previous phone call
M1.2 The answer to your question is that Enron cannot pay anything now.
Inform-AnswerOffline: Enron cannot pay

4.2 Links between Dialogue Acts

Each DA is assigned one or more links. This step is best done in the following order:

1. Determine whether there is a forward link or flink: does this DFU set up an expectation in the dialogue that the reader/hearer perform a certain action? This action can be a dialogue act (respond to a question) or a non-dialogue action perform a task).

A flink is simply added as Flinkname in a new line after the DFU, where name is the name of the DFU, namely $n.m$ if it is a single segment, or $n_1.m_1 - n_2.m_2$ if it is comprised of several segments, ranging from $n_1.m_1$ to $n_2.m_2$. Here is an example:

M2.6. What do you think the time frame is on this?  
[Request-Info: estimated time to complete July-August]  
Flink2.6

Note that depending on the DA you have assigned to the DFU, you have certain choices:

- Some DAs necessarily set up a flink. These are: Request-Information and Request-Action
- The NoDA never has a flink.
Some DAs may or may not have a flink - you need to judge from the content of the DFU. These are all remaining DAs: Inform, Commit, Reject, Backchannel, Perform, and Conventional.

2. Determine whether there is a backward link or blink: does this DFU relate to a previous DFU, by performing an action which responds in some sense to the previous DFU? A DFU can have both a flink and a blink.

A blink is added as Blinkflink-name in a new line after the DFU, where flink-name is the name of the DFU to which this blink links back. As before, it is written as \textit{n.m} if it is a single segment, or \textit{n_1.m_1 - n_2.m_2} if it is comprised of several segments, ranging from \textit{n_1.m_1} to \textit{n_2.m_2}. We continue the example from above (the flink example) with the corresponding blink:

\begin{itemize}
  \item M3.1. If I can get all of the information today, I can tell you this afternoon.
  \item M3.2. It doesn’t take long to create a calc sheets.
  \item [Inform: earliest date Enron pay Pasadena] Blink2.6
\end{itemize}

Note that the current DFU is 3.1-3.2, but the name of the blink is blink2.6, which shows that it is answering a previous question which was asked in DFU 2.6 (in the previous page).

If you add a blink, there are two options:

\begin{itemize}
  \item The blink corresponds to an existing flink. This is the case in our example that we have been using. No further action is needed other than noting the blink.
  \item The blink links back to a DFU which does not have a flink. This can happen when the previous DFU did not have an obligatory flink. In this case, please add the flink in the usual way to the original DFU, but call it \textit{sflink} for \textbf{secondary forward link}, as it is created only by a later blink.
\end{itemize}

Note that depending on the DA you have assigned to the DFU, you have certain choices:

\begin{itemize}
  \item Some DAs necessarily set up a blink. These are \textit{Commit} and \textit{Reject}.
  \item DA \textit{NoDA} cannot have any blink.
\end{itemize}
• For some DAs, you need to carefully understand the subtype of the DA (though you do not mark it) in order to choose whether or not to add a blink. This is the case for Inform and Backchannel DAs. Please consult the entries for these two DAs in Section 5.2.

• The remaining DAs may or may not have a blink - you need to judge from the content of the DFU.

Here is a complex example of a blink:

M5.2. On payment...
M5.3. We are having to forecast five days out.
M5.4. If I don’t know today, I cannot pay Pasadena until next Thursday
Inform: 5 day forecast rule
Blink2.7
Sflink5.2-5.4
...
M6.1 thats fine
Inform: ok
Blink5.2-5.4

Here, when first annotating DFU 5.2-5.4, it was identified as an Inform which is an answer to a previous question. So the DFU got a blink. But there was no reason to add a flink, as the reader is under no obligation to respond or to perform any other action. However, as the second part shows, the reader did respond, by letting the writer know that not paying till Thursday is okay. Thus, we get a blink from the second to the first part, and a sflink in the first part.

5 Basic Annotation Procedure

In this subsection, we lay out the basic steps involved in the annotation in some detail. You can find a summary of the annotation procedure in section 9.

5.1 Preamble: Thread-Level Annotation for Email Threads

The annotator starts out by reading through the entire thread (email, blog, or a series of conversations among the same people). Each file contains one thread, so there is no judgment to be made what a thread is. Then, choose the type of thread it is: What, in general, is the purpose, or content type, of this discussion thread?
or the Loqui subset, the thread type will be **External-event-planning**, where the event is to send library items to the patron, to place items on reserve, or similar types of library transaction.

- **Information-Fact**: The thread purpose is to convey or exchange information which is considered by all to be a fact.

- **Information-Opinion**: The thread purpose is to convey or exchange information which is considered by some or all to be opinion.

- **Information-Unsure**: Mark this general label if the thread purpose is clearly information, but you are unsure whether that information is fact or opinion.

- **External-event-planning**: Planning events that will take place outside of the email exchange, such as a meeting, or performance of a task.

- **Collaboration-on-information-product** Collaboration on a document or information. Mark this if the work will be done “inside” the email communication channel.

- **Social**: The main purpose of the thread is simply being social.

- **Other**: None of the above applies.

Then, enter in freehand a very short description of the topic of the thread. Use the list below as a guide (it may be useful for Enron in particular), but feel free to enter your own one-to-three word topic.

Social, Personal nonsocial, Internal project, Internal policy, Energy trading, Energy regulation, Business partnership, Legal advice, Human resources, Meeting minutes

### 5.2 Preamble: Dialogue Level Annotation for Loqui Dialogues

Add a tabular summary at the top of each DFU annotation for a Loqui dialogue that documents all the *book requests*, meaning each time a patron requests an item. There should be three columns headed “Lines”, “Request”, and “Successful”; use the following format:

```
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Line</td>
<td>Type of Request</td>
<td>Successful</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------</td>
<td>------------</td>
</tr>
<tr>
<td>8.4</td>
<td>Auth</td>
<td>n</td>
</tr>
</tbody>
</table>
```
The purpose of the table in the preamble is to summarize:

- Patron requests, not librarian requests
- Patron requests for reading material, not other types of requests
- Patron requests that are smoothly successful in that the librarian explicitly or implicitly appears to understand the request with no need for repetition by the caller, clarification by the caller, or any grounding or remediation

For each book request, identify the line number in column 1, identify the type of request in column 2, and in column 3, indicate whether the librarian appeared to understand the request on first hearing it. The types of requests are RC (for a request by RC number), Title (for a request by book title), Auth (for a request by author), Other (for a request for an item in some other fashion), and any combination of the above, e.g., Auth, AuthOther, AuthTitle, Other, RC, Title, TitleAuth and so on.

To reiterate, the contents of each column should be:

- **Column 1:** Line number in which the caller initiates the first request
- **Column 2:** Type of request (e.g., RC, Auth)
- **Column 3:**
  - “y” if the librarians response is an explicit or implicit indication that the requested item is being looked up, e.g., “okay, got it” or “okay”
  - “n” if the librarians response is an implicit or explicit indication of a need for confirmation or clarification
  - put “n/g” for “no with grounding” if it is correct repetition (grounding behavior only, not clarification) with no prosodic or other indicator that the librarian is attempting to elicit confirmation from the caller
5.3 Identify Utterance Units

The next step in the annotation consists in identifying utterance units. These are roughly tensed clauses that are an independent orthographic sentence (i.e., a sentence as delimited by a period\(^3\)), or that are introduced by coordinating conjunctions. A clause unit should never be longer than an orthographic sentence. The exact definition of a clause unit is not critical for the sake of this annotation.

Each clause unit is assigned an id. For the Eth, if \(m\) is the sequential position of the email message and \(n\) is the sequential position within the email of the utterance, then the label for the utterance unit is \(M_m.n\). For example, M1.2 is utterance unit 1 in message 2. For LCS1, the FINISH

5.4 Assign a Dialogue Act

Conceptually, each utterance unit is assigned a simple dialogue act (DA) \((\text{Request}, \text{Inform}, \text{Suggest}, \ldots)\) from a very small set of DAs. (Note that in practice, this step happens in conjunction with the step described in the next subsection, see section 6.) The choice of DA is augmented with a freehand description of the function or content. The DA information is written in square brackets below the utterance unit. It is of the following form:

\[
\text{Tag: Description}
\]

Here, \(\text{Tag}\) names the DA type (e.g., \text{Request}, \text{Inform}, \text{Suggest}), and \(\text{Description}\) is a natural language description of the content of the request or suggestion etc. The \(\text{Tag}\) should come from the set of tags presented in Appendix A.

Here is an example:

M1.2. I have completed the invoices for April, May and June.

\text{Inform: of invoice completion}

6 Identify Dialogue Function Units

The utterance units are grouped into discourse function units (DFUs) according to the following simple rule: if two adjacent utterance units have the same DA annotation (i.e., the same tag and the same description), then they are part of the same

---

\(^3\)Note that in transcriptions of informal spoken text, periods are used rarely; the court transcripts contain periods because the speech is more formal in nature, and is transcribed by a court stenographer.
DFU. This rule is pretty clear concerning the actual DA tag: it needs to be identical. But the description is freehand, so it is up to the annotator to decide whether two utterance units are in the same DFU or not. To minimize disagreement, we use the following heuristic:

**DFU Extent Heuristic:** two utterance units are part of the same DFU if they have the same DA tag, and if they can be given a coherent description which is not simply the conjunction of the descriptions of the two units.

We illustrate this heuristic with an example:

M1.2. I have completed the invoices for April, May and June  
*Inform:* invoice completion
M1.3. and we owe Pasadena each month for a total of $3,615,910.62.  
*Inform:* debt
M1.4. I am waiting to hear back from Patti on May and June to make sure they are okay with her.  
*Inform:* pending approval from Patti

Here, we have made each utterance unit its own DFU, each with different DA annotations. While this is reasonable, there is also a reasonable interpretation under which these three inform acts are in fact related to the same content, namely the writer's activity on invoices. Since such an interpretation is also reasonable and leads to a larger DFU, we prefer it:

M1.2. I have completed the invoices for April, May and June  
M1.3. and we owe Pasadena each month for a total of $3,615,910.62.  
M1.4. I am waiting to hear back from Patti on May and June to make sure they are okay with her.  
*Inform:* activity on invoices (invoice completion/debt/pending approval from Patti)

7 Specific Annotation Guidelines for the Loqui Corpus

The following FAQs have came up when new annotators looked at the following excerpt:
1. **Backchannel vs. Inform**

Sometimes the confusion between backchannel and inform comes up in what Herbert H. Clark referred to as “grounding” behavior, where speakers indicate to each other what their current state of understanding is, in order to establish the degree of “common ground” (e.g., in his 1996 book *Using Language*), so we have recently added something we want to see in the annotations, namely a backchannel that exhibits “grounding.” You would put “[Backchannel: grounding, ...]” but see below for more detail.

18 caller#2: oh okay
umm <B>
five nine zero six nine
**Inform:** book request number one
Blink17
19 librarian#1: <P> one second
<P> five nine zero six [os nine eo]
**Backchannel:** grounding, librarian repeats callers number
Sflink19
20.1 caller#2: [os yeah eo]
**Inform:** number was correct
Blink19
20.2 # <B> five nine zero eight six
**Inform:** book request number two
21 librarian#1: <BN> <P> eight six
**Backchannel:** grounding, librarian repeats part of caller’s number
22 caller#2: five nine zero nine zero
**Inform:** book request number three

In this excerpt, the librarian#1 has two turns where he or she repeats part of the RC number. This would be labeled as “Backchannel: grounding, ...” where you fill in your own description at the ellipsis. In the first case below, in 20.1 the caller explicitly confirms that the librarian was correct; here, because the caller confirmed the “grounding” behavior, the backchannel at 19 would get an Sflink. Otherwise, you would not have one, as in 21 and 22, where after the backchannel in 21, the caller just goes on with the next request.

28 Librarian # 1: unfortunately it doesn’t have
a number
Inform-Finish: (26.3, 28) book does not have a number
Sf
29 Caller # 2: oh
Inform: caller seems disappointed
Blink
30 Librarian # 1: so I can’t put it on your reserve list yet
Inform: librarian cannot put book on caller’s reserve list yet

Since the caller is doing more than indicating a state of continued attention in line 29, this is an [Inform].

2. Backchannel vs. Inform vs. Request-Information

- **Backchannel**: Speaker provides non-lexical indication of continued attention (“uh huh”)
- **Inform**: Speaker’s utterance is propositional (“I’m listening”)  
- **Request-Information**: Speaker’s utterance is an open proposition (“You’re listening?”; “Am I listening?”)

3. Request-Information vs. Inform

75.2 librarian#1: every cassette has four sides
76 Caller#1: every cassette has four [os sides <B> eo]
. . .
80 Caller#1: well I would [os think two sides eo]
82 Caller#1: play on one side play on the other side
83.1 librarian#1: [os no no eo] no
83.2 every cassette will have four sides

In line 82, the caller says what she believes to be the case about how to play cassettes, but in line 83, the librarian responds, “no no no,” and corrects her. In order to be able to label these DFUs, it once again depends on the prosody. Here, the caller uses statement intonation, so it is **Inform**, but if the caller used question intonation, it would be **Request-Information**.
4. Why is 3.1 not an Inform?

1 librarian#1: yes how can I help you
   Request-Information: how can librarian help
   caller
Flink1
2 caller#1: <B> may I order some books please
   Request-Information: can caller order books
   Blink1
Flink2
3.1 librarian#1: okay

If you listen to the librarian#1, her prosody clearly indicates that the “okay” is not an answer to the previous question. Rather, “okay” serves as a discourse cue and thus merits its purpose as backchannel.

5. Hold On

(a) When a librarian asks a caller to hold on (e.g. “One moment please” or “Hold on a second”) it is typically a Request-Action. Because it occurs so often, there may be a temptation to label it Conventional in this corpus.

(b) When a librarian asks a caller to hold on, and the caller responds “okay,” this response is a Commit?

8 Specific Annotation Guidelines for the Enron Corpus

INSERT OWEN’S NOV 9-10 EXAMPLES HERE

9 Summary: Annotation Instructions

1. Annotate one thread at a time.

2. Read through all the messages in the thread one first time.

3. Then read each email, and break up the sentences into groups which correspond to similar dialogue acts. We call these groups of sentences DA groups. You may need to break up a sentence. Number each sentence (or sentence fragment, if you have broken one up) by \( n.m \), where \( n \) is the message number and \( m \) is the sentence (or sentence fragment) number.
4. After each group of sentences with the DA group, write down the DA and paraphrase its content.

5. Then write down whether there is a backward link and to what, or whether a forward link is created necessarily (by these DAs: inform, request, ...)

6. If you create a blink to something which did not have an flink, you need to add an sflink (secondary forward link). If you want to create a blink to what is only a part of a previous DA group, you need to break up that DA group into 2 or more parts. Be sure to relabel all blinks that refer to the whole group (they get several blinks, normally). Be sure also to fill in the DA for each new DA group (the DAs should all be the same), paraphrase the content of the utterance, and create new flinks if needed.

7. Now add content links backwards for all DA groups which have as their DA inform.

10 Labeling

10.1 Thank you

- When a caller gives a strange response “Thank you” to the librarian’s request for her to hold on, how should it be labeled? Should it be an [Inform] or [Conventional: thank you]?
  In this case, it should be [Conventional: thank you].

- When a librarian says “yeah” in response to the callers “thank you very much,” should it be called [Conventional]?
  If it appears that the librarian means “you’re welcome” by saying “yeah,” then it should be [Conventional].

- When some says, “Thank you,” in response to something, can there be a Sflink after the previous statement and a Blink after the [Conventional: thank you]?
  There are two options. If “thank you” is uttered as a direct response to something, then there should be a sflink-blink pair as in the following case:

  Example:

  31.3 I just ordered these 5 books you requested [Inform: order is complete]
However, if the prosody and the situation of “thank you” suggests that it is being used more generally, like at the end of a dialogue (which is often the case with these dialogues), then there shouldn’t be a sflink-blink pair.

*Example:*

63 yes I checked everything to make sure
[Inform: librarian assures that the order is complete]
64 uh thank you ever so much
[Conventional: thanks]
65 you’re very welcome
[Conventional: thanks]

10.2 Oh

- When a caller says “oh” with an intonation that suggests she is disappointed by the news that the book she wants doesn’t have a number, is this considered [Backchannel] or [Inform] with a Blink that connects to the previous statement?

  If the intonation of “oh” indicates apparent disappointment, then it should be considered as [Inform], but it would be a vague [Inform].

10.3 Other Expressions

- Is “Andrew Heiskell Library may I help you” a [Conventional] greeting, an [Inform], or a [Request-Information]?

  Since the purpose of this conventional greeting is to ask what the caller would like to achieve during the phone call, it should be a [Request-Information].

- How should “I will be right back” be labeled?

  In the Loqui corpus, the expression “I will be right back” is a common phrase denoting the librarian’s brief absence, while checking the status of a holding (for example), and subsequent return. Therefore, it should be considered as [Conventional].
10.4 NoDA

- What should be done when someone tries to say something but is unintelligible? Should it be considered as a [NoDA]?

If the utterance doesn’t seem to be relevant to the conversation, then it should be a [NoDA]; however, in the case that the other person understands what was said and responds to it, even if the annotators cannot figure out what is said, then it could be labeled as [Inform: (not clear)].

- What is the difference between [NoDA] and <NU>?

[NoDA] is a label assigned to a DFU as part of DFU annotation, while <NU> is a tag used during the transcription process, so part of the difference is that they are totally different levels of representation.

10.5 Inform-Start, Inform-Continue, and Inform-Finish

1. When is Inform-Start, Inform-Continue, and Inform-Finish: (#,#,#) necessary? Consider the following example:

5.2 may I have your last name
Request-Information: librarian requests patron’s last name
Flink5.2
6 caller#1: <B> <LASTNAME> <LN-INITIAL-1> <LN-INITIAL-2> <LN-INITIAL-3> <LN-INITIAL-4> [os <LN-INITIAL-5> eo]
Inform-Start: caller gives last name and spelling
Blink5.2
7 librarian#1: [os <LN-INITIAL-1> eo] <LN-INITIAL-2> <LN-INITIAL-3> <LN-INITIAL-4> Backchannel: grounding, librarian repeats all but last initial of spelling
8 caller#1: <LN-INITIAL-5> <LN-INITIAL-6> <LN-INITIAL-7> <LN-INITIAL-8> [os <LN-INITIAL-9> <LN-INITIAL-10> <LN-INITIAL-11> eo]
Inform-Continue: Spelling of last name continued from last mentioned initial
Blink5.2
9 librarian#1: [os <LN-INITIAL-5> <LN-INITIAL-6> eo] <LN-INITIAL-7> <LN-INITIAL-8>
When the librarian asks for the caller’s last name, it takes more than one try for the librarian to be completely satisfied with the information. For instance, the librarian exhibits the usual backchannel-grounding behavior of repeating what the librarian has heard from the caller. Therefore, the completion of DFU 5.2 does not occur until the line number 10, which is why Inform-Start, Inform-Continue, and Inform-Finish: (#,#,#) are necessary to show that this specific DFU spans for more than one line of DA.

10.6 Book Request Table: Preamble to Loqui Annotations

1. A librarian agrees to locate a book, but then it turns out that she can’t find it because the title given by the caller was incorrect. Since it was the caller’s fault to have given a wrong title, should this request then be considered successful on the librarian’s part?

   In the case that the librarian completes what the caller has requested, then the request is successful even if the caller accidentally gives an inaccurate RC number.

2. If a request is not regarding a title but something else (i.e., removal of an author from a favorite list), should this DFU be put in the table?

   These other requests should be noted, but they should not go in the same tables as the book requests.

3. During the dialogue the librarian gives no indication as to which of the 10 requested catalog numbers were successfully logged. At the end of the conversation, we learn that seven of the requested items have been ordered, and therefore, can we assume that three were unsuccessful? Is there a way to indicate such a phenomenon?

   A “successful” request has nothing to do with whether a book was ordered or not, but whether the request was understood by the librarian the first time it was actually said in a conversation.
11 Example Thread

-----Original Message Number 1-----
From: Parker, Megan
Sent: Wednesday, November 28, 2001 8:38 AM
To: Ward, Kim S (Houston)
Subject: City of Pasadena

M1.1. Kim:
[salutation]

M1.2. I have completed the invoices for April, May and June
M1.3. and we owe Pasadena each month for a total of $3,615,910.62.
M1.4. I am waiting to hear back from Patti on May and June to make sure
[Inform: invoice completion/debt/pending approval from Patti]
sflink-1.2-1.4: acknowledgment of accomplishment (actualized by subsequent

M1.5. Do you want me to pay Pasadena on Friday for these months M1.6. or
you want me to hold off until I finish July and August?
[Request-info-either/or: commission to pay Pasadena or to delay paying
Pasadena] flink-1.5(commission to pay Pasadena now) or
flink-1.6(commission to pay Pasadena after Jul-Aug)

M1.7. Again, I do not have all of the information for July and August,
M1.8. so I cannot give you any numbers.
M1.9. If I go by what is currently in the system as a guide, Pasadena would
M1.10. I need to forecast the money today,
[Inform: of lack of numbers/Pasadena debt to Enron/forecast deadline]

M1.11 so please let me know what you would like to do.
[Request: answer to [or(M1.5,M1.6)]
flink1.5(commission to pay Pasadena now) or flink1.6(commission to pay Pasadena after Jul-Aug)

M1.12. FYI-I will be out of the office all next week.
[Inform: of future absence period all next week]

M1.13. Megan
[signature]
-----Original Message Number 2-----
From: Ward, Kim S (Houston)
Sent: Wednesday, November 28, 2001 8:42 AM
To: Parker, Megan
Subject: RE: City of Pasadena

M2.1. Thanks you so much for working on these as fast as you have
[Execute performative: thanks-content to Megan completing invoices fast]
Blink1.2-1.4 (accomplishment)

M2.2. there was a priority in getting them out.
M2.3. However, as you know, the less cash outlay that we have right now
[Inform: of high priority for completing invoices fast, (re) enron to mi]

M2.4. Lets wait and try to offset the amount we owe them with what they
M2.5. so I say wait until you get July and August done.
[Suggest action: Suggest: offset debt to Pasadena w/ amount due from Pas]
Blink1.6

M2.6. What do you think the time frame is on this?
[Request info: estimated time to complete July-August]
Flink2.6(estimate time to complete July-August)

M2.7. And when would be the earliest that we would make the payment?
[Request info: earliest date enron pay Pasadena]
Flink2.7(estimate date enron pay Pasadena)

M2.8. Thanks,
[Thanks]
-----Original Message Number 3-----
From: Parker, Megan
Sent: Wednesday, November 28, 2001 8:56 AM
To: Ward, Kim S (Houston)
Subject: RE: City of Pasadena

M3.1. If I can get all of the information today, I can tell you this afternoon.
M3.2. It doesn’t take long to create the calc sheets.
   [Inform: earliest date enron pay Pasadena]
Blink2.6

M3.3. I understand from Janine that you or Patti can provide me with the detail I need.
M3.4. If necessary, I can come pick it up.
M3.5. I have sent Patti a list.
   [request for action: Kim gives Megan detail Megan needs]
Sflink3.3-3.5

M3.6. For payment, we have to forecast the money two days out.
M3.7. So, if I know today, I can pay on Friday.
   [inform: earliest date enron pay Pasadena]
Blink2.7

M3.8. Megan
   [salutation]

-----Original Message Number 4-----
From: Ward, Kim S (Houston)
Sent: Wednesday, November 28, 2001 9:03 AM
To: Parker, Megan
Subject: RE: City of Pasadena
M4.1. Patti is the one with the details,
M4.2. I’m just the deal maker
M4.3. and don’t have access to any of the systems.
M4.4. All I know is what fixed priced baseload deals we have.
M4.5. I have no idea what flows.
M4.6. I hope Patti can help you soon.
[reject: has no info, Patti does]
Blink3.3-3.5

M4.7. Let me know if you don’t hear from her today.
[request for info: has Megan not heard from Kim]
Flink4.7

M4.8. Thanks,
[thanks]

M4.9. Kim
[signature]

-----Original Message Number 5------
From: Parker, Megan
Sent: Wednesday, November 28, 2001 11:38 AM
To: Ward, Kim S (Houston)
Subject: RE: City of Pasadena

M5.1. I’ll check with Patti later today.
[inform: will get info from Patti later today, which will enable her to]
Blink4.7

M5.2. On payment...
M5.3. We are now having to forecast five days out.
M5.4. If I don’t know today, I cannot pay Pasadena until next Thursday.
[inform: 5 day forecast rule]
Blink2.7
Sflink5.2-5.4

M5.5. Megan
-----Original Message Number 6-----
Subject: RE: City of Pasadena
X-From: Ward, Kim S (Houston)
X-To: Parker,
X-Time: 17:42:33 GMT

M6.1 thats fine
[inform: ok]
Blink5.2-5.4

M6.2- we can’t pay them anyway
[inform: no money]

Appendices

A Tags used in speech transcription

- <B>: breathing noise
- <BN>: background noise
- <BSTN>: background caller
- <CHKL>: chuckle
- <CGH>: cough
- <CT>: clear throat
- <G>: garbled (completely unintelligible, not clear if speaker articulated any words)
- <IR>: incoming call ring
- <NS>: not sure (for use by person doing the transcription to mark spots to return to and review again)
• <NU>: non-understanding (speaker probably said something specific, but recording is not clear enough, or speaker spoke too quickly or softly)

• <P>: long pause

• <ST>: self-talk (similar to NU, but speaker was clearly speaking to herself)

• <TN>: telephone noise, when librarian can be heard working at keyboard during call

• <TP>: typing noise, when librarian can be heard working at keyboard during call

• <VB>: voice in background

References


