Guidelines for Dialogue Act and Addressee Annotation Version 1.0

October 13, 2005

Contents

Chapter 1

Introduction

The AMI Project is developing technology for browsing recorded meetings. As part of this effort, we are recording many hours of meetings and annotating the recordings by hand for many different kinds of things that might help a meeting browser organize recordings and present them to users. Annotating data by hand is slow and expensive, but once we have annotations for some meetings, we can figure out how to get them automatically for new ones. The meetings that we are annotating have already been transcribed, so there is a written version of what the speakers said.

1.1 What is dialogue act annotation?

This manual provides instructions for annotating the AMI Meeting Corpus for dialogue acts. Dialogue act annotation is about marking up the transcription according to speaker intention — that is, what kind of thing each person is trying to achieve by what they say. A very simple dialogue scheme might have you separate questions (in which the speaker is trying to get some information from another participant) from statements (in which the speaker is trying to give another participant some information) from suggestions (in which the speaker is raising some possible course of action), and put everything else that is neither question, statement, nor suggestion into a "bucket" to throw it away. Our scheme is essentially like this, but with more different types of acts to identify. An essential part of dialogue act coding is finding where one dialogue act finishes and another one starts; this is called "segmentation".

An important thing to keep in mind throughout is that dialogue act annotation is about function (what the person means) and not form (how they say it). The form of an act can give hints about the function, but it does not completely determine what the function is. For instance, questions often start with a "WH" word and sound like a question because the speaker's pitch rises at the end, but many suggestions start with "Why don't we...", and some people's voices go up at the end of almost everything they say. Suggestions in particular tend to come

in many different forms, since people who aren't too sure that the group will like their suggestion tend to hide it by making it sound like a question, such as "Can you take the minutes today?", or even a statement. Similarly, "no" shows disagreement after "Shall we have the project manager's presentation now?", but agreement after "We don't want to do that, do we?" We always want dialogue acts coded by function, not form. Often the function is clearer if you listen to them rather than relying on the transcript.

It is important also to pay attention to the videos. The intention of a dialogue act may be expressed by a combination of speech and gestures or drawings, for instance if someone says "It's like..." while making a hand gesture showing the shape and size of the object under discussion. By listening to the speech only we may take this as an incomplete dialogue act without a clear intention. But when you look at the speaker's movements the function of the utterance is clear and complete.

One last point — although most dialogue acts involve some speech, it is entirely possible for someone to have exactly the same effect as a spoken dialogue act without saying anything, for instance, by nodding in response to a question. It's important to remember this when we come to the part of the annotation that involves showing how different acts relate to each other, but we are only annotating dialogue acts that leave some trace in the transcription of what was said.

1.2 The meetings and how they are transcribed

Many of the meetings that we are annotating aren't from real groups, but have been collected by having four people role-play a design team from a small company, RealReactions. In these meetings, there are four participants, taking four different roles: a project manager (PM), user interface specialist (UI), marketing expert (ME), and industrial designer (ID). The teams design a new kind of remote control from start to finish over a series of four meetings. In the four meetings, the same person takes the same role every time, but won't necessarily always be sitting in the same place.

The rest of the meetings that we are annotating are either from real groups doing their normal work or from other role-playing exercises that are more self-explanatory (for instance, in some of them, a real research group plans a simplified version of a real move to a new building). These meetings usually have four participants, but sometimes they can have three or five. For all of these meetings, the participants are just labelled A to C, D, or E.

In the annotation tool, you will see the transcription laid out in a way that is designed to be easy to read. Each line of the transcript first identifies the speaker and then the words that they say. There are some things that aren't words shown in the transcript; $ means laughter, % shows a cough, and # means some other noise. The transcriptions contains some punctuation, but this isn't something you'll want to take too seriously. Punctuation usually reflects the transcriber's views on form, but it doesn't necessarily tell you the function, and

at any rate, you will not necessarily agree with the transcriber's choice. Using the tool, you can both read the transcription and listen and watch the recording; if you select a line of the transcript by left clicking on the speaker id and press ctrl-right-click, you will hear just that section of the recording.

In small meetings like these ones, sometimes one person will speak for a long time, and sometimes more than one person talks at the same time. The transcription breaks up each person's speech into whatever chunks the transcribers thought were handy and shows them in the order in which they start. You might be tempted to think that where the transcribers broke the transcription into lines is important, but it isn't - they simply broke them wherever it was convenient for them to do so. For instance, the transcription might be broken right in the middle of a statement just because someone else says "uh-huh" there. For this reason, don't take the line breaks seriously when you are deciding how to segment a meeting into dialogue acts. You could need to join several transcription lines into one act, or you might need to break one transcription line into several acts.

There are two main steps in dialogue act annotation: segmenting the transcription into separate dialogue acts, each of which conveys one speaker intention, and classifying those intentions. At the same time as these steps, there are a few other things to annotate as well. Although these are not usually part of the way dialogue act annotation is done, we are including them because it saves time to have to mark them while you are familiar with the meeting and how you have divided it into acts. We will describe these after the main steps.

Chapter 2

Segmentation

Dialogue act annotation is about the type of intention the speaker has. Intuitively, each time a new intention is expressed, you should mark a new segment.

It is possible for a speaker, for instance, to ask two different questions in a row without anyone else speaking. You may be tempted to mark the two questions as a single act, but you should not, since each segment should contain a single speaker intention. This is true not just for questions, but for all act types.

It can be tricky to know whether you are dealing with one segment or two of the same type by the same speaker, especially when the speaker is giving information. The following tests will be helpful:

- Think whether or not the speech that might be a second segment adds any new information. For instance, if someone says "No, it's not", the "it's not" is not a separate segment, since it rephrases the same information as the "No".

- Lengthy pauses or conjunctions that introduce whole new clauses such as "so", "because", and some uses of "and", "but", and "or" can be hints that a new segment is starting. If you encounter a conjunction and aren't sure whether to break the segment at that point, think about whether the first half requires the second half to be complete; for instance, if you split an "if" clause from the "then" that follows, neither segment will express a complete intention, suggesting that they should really be combined into one segment.

- Watch out for places where the speaker changes from talking to one person to talking to someone else or the whole group, or the other way around. In these cases, there would be two intentions, and therefore two segments. However, keep in mind that the speaker's intention is the deciding factor.

For example, if one person says all of what's in the examples below, the pipes (|) indicate the segment boundaries.

Example 1

D	And then you have the numeric pad in the dark blue at the bottom, | and on the right-hand side you have the access to the menu on the T_V_, | and on the left-hand side you have the the the ability to turn off the voice recognition. | So this is pretty much what we had on the white board the last time.

Example 2

D	so in terms of function, you have to be able to turn the T_V_ on and off, volume and channel control, menu control voice recognition control, | and we have incorporated the LCD screen on the flip panel as part of the design, | if we figure out it is too expensive, well then you just take it off.

Example 3

D	And part of the thing is a lot of people say that they don't like something that is too light | because they do not feel like they they have enough control over it | So I mean maybe this is excessively heavy | but I think it needs to have some weight | it needs to feel like you're holding something.

It is also possible in theory for one speaker to start a dialogue act and for another to finish it — such as when someone can't think of a word, so someone else fills it in. However, we want all of our acts only to contain transcription from a single speaker. In these cases, mark a separate segment for each speaker, giving each of the segments the same type. This is the only case when a single intention will be split over more than one act.

If there are cases where you are in doubt about what to do, use two segments, not one.

Chapter 3

Classification

As well as segmenting a meeting into dialogue acts, the annotation requires you to classify each dialogue act according to the kind of intention the speaker conveys by the act. Our dialogue act classes fall into the following groups:

- Classes for things that aren't really dialogue acts at all, but are present to account for something in the transcription that doesn't really convey a speaker intention.

- Classes for acts that are about information exchange.

- Classes for acts about some action that an individual or group might take.

- Classes for acts that are about commenting on previous discussion.

- Classes for acts whose primary purpose is to smooth the social functioning of the group.

- A "bucket" type, OTHER, for acts that do convey a speaker intention, but where the intention doesn't fit any of the other classes.

We address each of these groups in turn, first explaining what the classes in the group have in common and then defining each class. At the end of all of the groups, we describe possible confusions among the types and how to resolve them.

3.1 Special classes to allow complete segmentation

When you've finished annotating a meeting, we want everything in the transcription to be in a dialogue act segment, with nothing left over. This would ordinarily cause some difficulties, since not everything in the transcription really conveys a speaker intention. We get around this by using some special classes

for segments that aren't really dialogue acts but let us mark up everything. These special classes are as follows:

3.1.1 BACKCHANNEL

In backchannels, someone who has just been listening to a speaker says something in the background, without really stopping that speaker. Backchannels signal that what the speaker has just said presents the person who utters the backchannel with no difficulty that they want addressed, so that the speaker can continue with their next intention. Some typical backchannels are "uh-huh", "mm-hmm", "yeah", "yep", "ok", "ah", "huh", "hmm", "mm" and, for the Scottish speakers in the data recorded in Edinburgh, "aye". B ut backchannels can also repeat or paraphrase part or all of what the main speaker has just said. Backchannels are always fairly short, since if they were very long they'd be likely to interrupt the main speaker, but they can be more than one word, for instance, if the speaker is repeating a complete phrase from the main speaker, or if the speaker strings together several of the shorter expressions.

In the examples below, the dialogue act in question is always written in capital letters.

Example 4

D	Right away I'm making some kind of assumptions about what what information we're given here
B	MM
D	thinking, 'kay trendy probably means something other than just basic,
B	YEAH
D	something other than just standard...

Example 5

D	So they w all work actually function together but I have different remote controls for each of them.
B	MM-HMM
D	So it's sort of ironic...

Example 6

A	...depends on how much you can cram into that price
B	MM-HMM
D	OKAY
A	Um. I think that's the main factor.

TEMP REPETITION EXAMPLE REQUIRED HERE

3.1.2 STALL

Sometimes people start speaking before they are really ready, or keep speaking when they haven't figured out what to say, just to try to get or keep the attention of the group. When they do this, they often use special sounds called "filled pauses", which are transcribed as "uh", "um", or "mm" depending on how they sound, but they might also use words that don't commit them to much, like "OK", "yeah", "like", or "but". Sometimes what they say will have some content, but will be a valid start to what they're really trying to say, and they will repeat it more or less verbatim as part of their proper dialogue act once they get going properly.

Example 7

A	I MEAN - | I mean we should make the remote control red.

Example 8

C	SO - | then we have sample sensor

Cases like this are likely to have hyphens showing that there was some disruption to what the speaker was saying at this point, although they might not, and not every hyphen indicates this behaviour. In the second example, listening is the key to knowing that the "so" isn't part of the eventual act.

Although this behaviour is so natural we often don't even hear it unless we are listening carefully, it's more of a problem for computers being able to tell which dialogue act someone is making. What happens at the beginning and end of a dialogue act can be crucial for this (recall that many questions start with a "wh" word and end with the speaker's pitch going up), but this behaviour gets in the way of a computer using that information. For this reason, we want you to separate out the parts of the transcription where the speaker is doing this into STALL segments. STALL isn't really a dialogue act, since the speaker doesn't really convey an intention in these segments (apart from, in some sense, "hold on, I want to speak"). It's just a convenience for us. Only mark as STALL this kind of material if it is when a speaker first starts speaking after being silent for a while or it is between two of their intentions that you would be placing in

two different segments anyway; getting the segmentation into intentions correct is more important than identifying this behaviour. Also, it would be tedious to mark every filled pause that is between acts as STALL, since there are very many. If the only transcription associated with the STALL is a filled pause, transcribed exactly as "uh", "um", or "mm", then you can include it in the nearest dialogue act; we'll put them in separate STALL segments automatically after you are done.

Example 9

PM	SO UM | we want to do a new remote control

Example 10

PM	uh maybe we can change the colours

In the last example the filled pause is not segmented as a STALL because there is no other text associated with it.

3.1.3 FRAGMENT

The FRAGMENT category is for things in the transcription that aren't STALLS or BACKCHANNELS, but do not convey a speaker intention.

One kind of FRAGMENT is when the speaker started saying something, but didn't get far enough to express the intention they started on before they either changed their mind and said something else, or stopped altogether.

Example 11

C	I WANT - | what if we were to make the remote control shiny?

Example 12

A	I THINK THE - |
C	The remote shouldn't be too heavy.

Example 13

B	YEAH I THINK THROUGHOUT THE PROJECT WE'LL PROBABLY | I mean depending on the the things that we include in the remote control I mean I think there'll be a price per button or per function that add on there | IF WE HAVE UH

The last example has two separate FRAGMENTs with another act in between.

Another kind of FRAGMENT is for when a segment is needed to cover some part of the transcription, such as laughter, which isn't really speech, or which is unclear enough that it hasn't been transcribed but marked as a gap in the transcription. It isn't necessary to separate out laughter, coughing, noises, and so on, when they occur just before, just after, or in the middle of a dialogue act, but sometimes they occur by themselves, so that part of the transcription can't easily be covered by some other, meaningful segment. In these cases, use the FRAGMENT class.

You may find other kinds of FRAGMENTs that we haven't realized will happen in the data. Treat FRAGMENT as a kind of mini-bucket just for things that don't convey a speaker intention but are neither BACKCHANNELs nor STALLs.

3.2   Acts about information exchange

Acts in this category can either express information or attempt to elicit information from others.

For our purposes, information includes statements that are objectively true about the world (such as "snow is white" and "Sanyo makes remote controls") as well as, in the scenarios that elicit meeting behaviour such as the one for designing a remote control, things that are set by the materials the participants receive (such as "The marketing report says brown is the new black.") However, people also often make subjective statements ("cats are very friendly animals"), and sometimes present as information things that are just plain wrong (such as "Titanium is relatively cheap", when it isn't). We aren't interested in the actual truth or subjectivity of what the participants say — in all of these cases, they present what they say as information, and it should be classified that way. We also aren't interested in whether the information is new to the group members or tells them things they already know (the exception to this being when someone repeats what a previous speaker just said as a sign that she has understood them).

3.2.1   INFORM

The INFORM act is used by a speaker to give information.

Example 14

PM	WE WANT TO DO A NEW REMOTE CONTROL | IT HAS TO BE ORIGINAL, TRENDY, AND USER FRIENDLY

Example 15

PM	THE BUDGET FOR THIS PROJECT IS THREE THOU-SAND FIVE HUNDRED EUROS | AND DEADLINE IS AT THE END OF THIS DAY

Example 16

PM	...do you have any other information for us at this stage
ME	NO
PM	no, okay

Example 17

ME	...BECAUSE UH NOBODY HAS COLOURFUL REMOTE CONTROLS | THEY'RE ALWAYS BLACK

3.2.2   ELICIT-INFORM

The ELICIT-INFORM act is used by a speaker to request that someone else give some information. The act doesn't have to make clear who is meant to give the information.

Example 18

UI	any of these things we need to address in the next meeting | or should we at least set some basic things that we need to prepare for
PM	YEP SO MARK ARE THERE ANY CONSTRAINTS OR INFORMATION FROM THE MARKETING SIDE | I mean | WHAT SORT OF WHAT SORTA PRODUCT ARE WE ENVIGIT- ENVISAGING
ME	well I think from the research that we've done so far what we would uh what we would like to have is something that would work for a variety of different devices but still maintaining simplicity

Example 19

PM	...DO YOU HAVE ANY OTHER INFORMATION FOR US AT THIS STAGE
ME	no
PM	no, okay

3.3   Acts about possible actions

This group of act classes is about expressing possible actions that the group, some individual in the group, or some person or group in the wider environment could do, such as the person running the recording equipment or people from the organization to which the group belongs (e.g. for the remote control design teams, researchers from the marketing department). Meeting participants can discuss many different kinds of actions, from ones concerning the room (e.g., opening doors) and how they conduct their business (e.g., having presentations, breaking for lunch) to actions to complete any tasks they have come together to discuss. Actions, in this context, are not always physical: they can also refer to a variety of steps the group take towards the completion of their goal, e.g., advancing ideas about what the remote control should look like, or proposing new features. A dialogue act can be about a possible action without that action ever happening, or even anyone indicating that it will. For instance, people quite often discuss actions and decide against them, or decide for them, but change their minds later. They also sometimes discuss actions because they shouldn't be taken. All of these cases should be classified as being about possible actions.

As with information exchange, acts in this category can either express possible actions or attempt to elicit the expression of possible actions from others.

3.3.1 SUGGEST

In a SUGGEST, the speaker expresses an intention relating to the actions of another individual, the group as a whole, or a group in the wider environment. Sometimes, SUGGEST can take the form of a question, especially when the speaker is not sure that the group will accept the idea or suggestion.

Example 20

	[...prior discussion about who should talk next. PM is giving instructions how to organise this...]
PM	SO MARK'S THE DEFAULT | AND IF HE'S TALKING MAYBE JOHN CAN FILL IN THE NOTES WHILE MARK IS GOING ON | so just does anyone else have any thing that they've prepared on their role or information on the project for present at this meeting...
ID	MAYBE THERE ARE A FEW THINGS THAT WE CAN CLARIFY BEFORE WE GET ON TO THE

Example 21

UI	WELL ALSO WE NEED TO SET UP UH SOME SORT OF FEEDBACK BETWEEN JOHN AND VINCENT AND THE MARKETING AREA | BECAUSE THEY HAVE TO COME UP WITH IDEAS | AND THEN WE HAVE TO TEST THEM WITH THE CUSTOMERS SO
PM	yeah
ID	ok

Example 22

	[...prior discussion about the colour of the remote control the group has to design...]
ID	Yeah | EVEN WE CAN CHANGE COLOURS | no?

3.3.2 OFFER

In an OFFER, the speaker expresses an intention relating to his or her own actions.

Example 23

UI	you need to give me your ideas — AND THEN I NEED TO SEE WHETHER THAT WOULD SELL IN THE MARKET PLACE

Example 24

PM	Yeah uh	the agenda of the meeting is opening	THEN I'M GOING TO TALK ABOUT UH THE PROJECT MANAGE-MENT

3.3.3 ELICIT-OFFER-OR-SUGGESTION

In a ELICIT-OFFER-OR-SUGGESTION, the speaker expresses a desire for someone to make an offer or suggestion. This can either be about something specific or a more general attempt to elicit an act about a possible action, such as "What should we do next?"

Example 25

	... the group discuss a range of features and try to see whether there are any further ideas...
ID	Should we maybe make a decision about what features we actually want to include, 'cause we've thrown a lot of features onto the table, but
PM	Yeah.
ID	Do we actually want to incorporate all of them	OR HAVE WE MISSED ANYTHING?

Example 26

UI	YOU NEED TO GIVE ME YOUR IDEAS — and then I need to see whether that would sell in the market place

This last example is a harder case than the others - the idea here is that the PM wants management to make a suggestion to the group about the design.

3.4 Commenting on Previous Discussion

In these acts, the participants contribute to the discussion by commenting on what has been said or done so far. As usual, there is an act for attempting to elicit comment on the previous discussion.

3.4.1 ASSESS

An ASSESS is any comment that expresses an evaluation, however tentative or incomplete, of something that the group is discussing, where the something could be another dialogue act or something apparent from the working environment, like slides or, in the remote control design trials, the playdough remote control mock-up. There are many different kinds of assessment; they include, among other things, accepting an offer, expressing agreement/disagreement or any opinion about some information that's been given, expressing uncertainty as to whether a suggestion is a good idea or not, evaluating actions by members of the group, such as drawings. Assessments themselves can be assessed in further acts, and if the thing being assessed is a dialogue act, it doesn't have to be from a different speaker, since people can assess their own contributions to the discussion.

An ASSESS can be very short, like "yeah" and "ok". It is important not to confuse this type of act with the class BACKCHANNEL, where the speaker is merely expressing, in the background, that they are following the conversation.

Example 27

	... discussion of an offer to research past designs and report back...
PM	THAT WOULD BE GREAT. | so if you find out from the technology background, | THAT WOULD BE GOOD.

Example 28

	... discussion of the price for the remote control...
ME	'TIS BIG NUMBER.

Example 29

ME	It has to be nice looking
ID	YEAH

3.4.2 COMMENT-ABOUT-UNDERSTANDING

COMMENT-ABOUT-UNDERSTANDING is for the very specific case of commenting on a previous dialogue act where the speaker indicates something about whether they heard or understood what a previous speaker has said, without doing anything more substantive. In a COMMENT-ABOUT-UNDERSTANDING, the speaker can indicate either that they did understand (or simply hear) what a previous speaker said, or that they didn't. This class is quite limited, since if the speaker does anything beyond comment on whether or not they understood, the act counts as something else. For instance, if they ask for clarification, that would be some type of eliciting act (depending on what they need clarified), or if they express a reaction to something they understood, that would be an ASSESS.

Example 30

	B is drawing her favourite animal on the whiteboard
B	Yes, that's a rabbit. That's my favourite one.
D	A WHAT?
A	A rabbit.
D	A RABBIT, OH YEAH.

3.4.3 ELICIT-ASSESSMENT

In an ELICIT-ASSESSMENT, the speaker attempts to elicit an assessment (or assessments) about what has been said or done so far. Sometimes a speaker seems to be making a suggestion and eliciting an assessment about it at the same time. In these cases, look at the information presented: if it is new information then it will be a SUGGEST (see above), if it implies previously known information then it will be ELICIT-ASSESSMENT.

Example 31

	… the group discuss a range of features and try to see whether there are any further ideas…
ID	Should we maybe make a decision about what features we actually want to include, 'cause we've thrown a lot of features onto the table, but
PM	Yeah.
ID	DO WE ACTUALLY WANT TO INCORPORATE ALL OF THEM | or have wee missed anything?

Example 32

	... discussion of several designs...
UI	I WANTED FEEDBACK | I think we need to rate these | BUT WE'LL SEE WHAT YOUR UH PERSONAL PREFERENCES ARE

Example 33

PM	*[...]* so, | MISTER MONEY, WHAT'S YOUR OPINION ON THIS REMOTE CONTROL? |
ME	we gonna try to measure how good it is instead of just talking about it

3.4.4 ELICIT-COMMENT-ABOUT-UNDERSTANDING

In an ELICIT-COMMENT-ABOUT-UNDERSTANDING, the speaker attempts to elicit a comment about whether or not what has been said or done so far has been understood, without further asking for assessment of anything in the discussion.

Example 34

	B is describing an idea about the remote control functions
B	like three mentals states, yeah, you know what I mean, we can just make it uh
	[other participants acknowledging]
B	controlled by a brain, | HUH?

3.5 Social acts

Social acts affect interpersonal relationships in the group — they make the group or some individual in the group feel better or worse about himself or about the group as a whole.

Most dialogue acts will have some social overtones, since even the choice of words can reveal positive or negative feelings towards the other group members. It is especially common to overlay something to make the group feel positive about an individual or the group as a whole when criticizing their work, since otherwise people tend to react badly. Only classify an act as social if it does not also fit in one of the other groups — that is, only if it is intended solely to improve or harm the interpersonal relationships within the group, without content that bears more directly on the task the group is performing.

3.5.1 BE-POSITIVE

BE-POSITIVE includes any social acts that are intended to make an individual or the group happier. For example, little acts of politeness like greeting one another or saying "please", "sorry", and "thank you" smooth social functioning in the group, as do things like good-natured jokes, positive comments about someone's appearance or intelligence, and expressions that say they are doing a good job.

To the great detriment of society, one common way in which groups are positive about themselves is by being negative about other groups or about people outside the group. Because this improves relations in the group itself, it counts as being positive.

Example 35

PM	..so I think thats it for now if we all head of and well meet again in about an hour or so
UI	okay
PM	THANKS

Example 36

ID	... I guess the main constraint then is seeing as we pretty much got no constraints on functionality the main constraints gonna be price | um so uh I mean | are we talking in terms of material price
PM	yeah
ID	UM SORRY SORRY TO COMPLICATE THINGS

Example 37

	...one participant pulls a funny face...
UI	GOOD EXPRESSION! [laugh]

3.5.2 BE-NEGATIVE

BE-NEGATIVE includes any social acts that express negative feelings towards an individual or the group. As for BE-POSITIVE, there are many possibilities for BE-NEGATIVE acts; among other things, they include hostile comments, jokes if the point is to run down someone, and expressions of frustration or withdrawal. Acts do not have to be outwardly aggressive to be negative; for

instance, simply demonstrating boredom during someone else's presentation will decrease the group's social functioning. People can be negative about others in the group, the external environment, or themselves.

3.6   The bucket class, OTHER

The OTHER class is a bucket for all proper dialogue acts - where the speaker is conveying an intention - that don't fit any of the other classes. We don't give any examples for the OTHER class because it could contain all kinds of things. For instance, this category includes self-addressed speech, where a participant speaks to herself, as if mumbling or thinking aloud (eg. "Now, where am I?" while looking at the slides half-way through a presentation), usually uttered in a subdued fashion, as they are not really meant for anyone else to hear.

It could be that as you code, you will begin to recognize classes of acts that we haven't defined in the manual because we didn't expect them to occur frequently enough that we should bother having you classify them separately. It's always possible we've misjudged the data, though. If you discover that you're using OTHER frequently, and often for the same kind of thing, note it briefly on the discussion Wiki. We're very unlikely to change the annotation manual, but your comment will help us when we analyze the data.

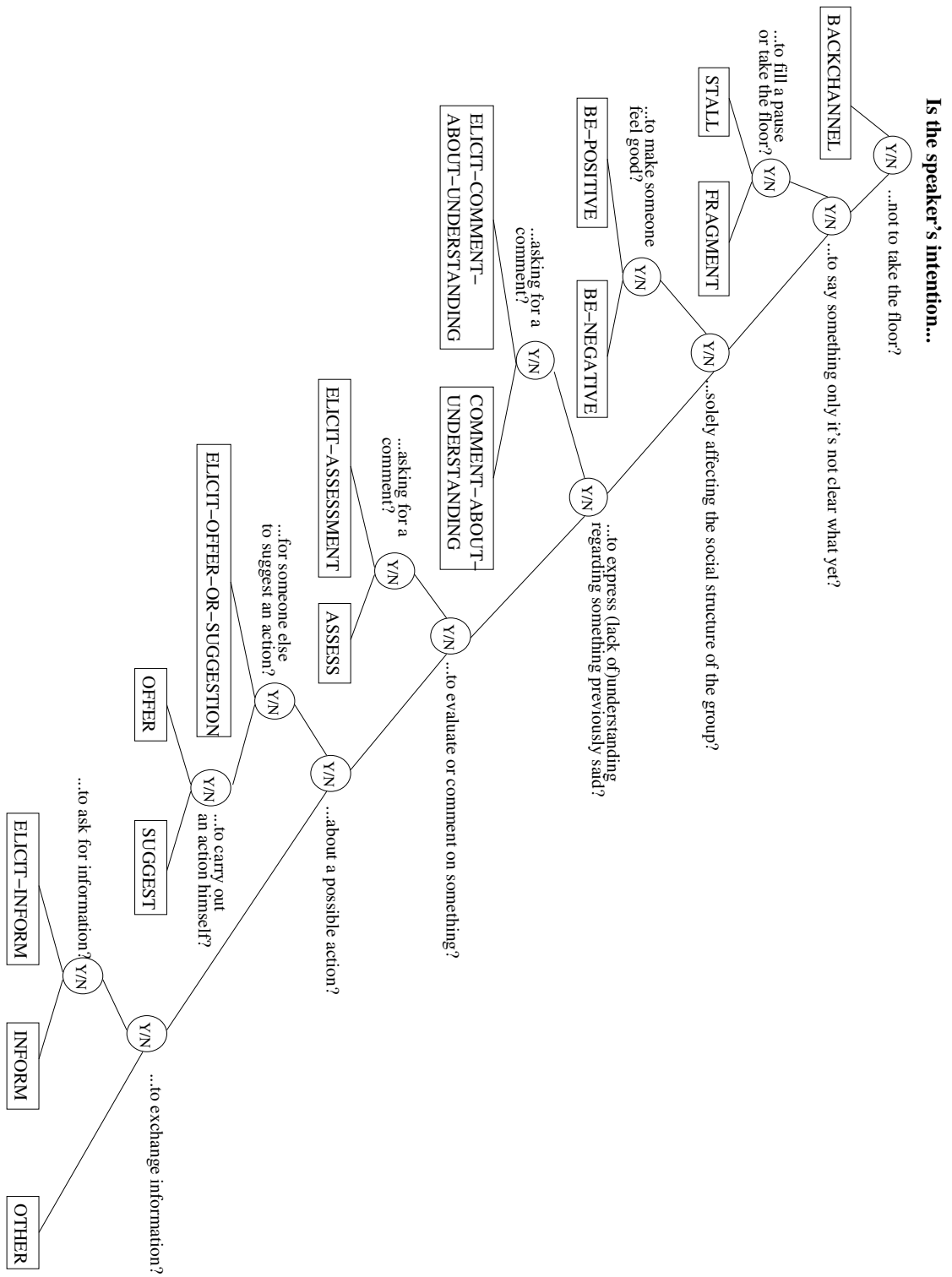## 3.7   A decision tree to help with classification

You need to read the complete manual in order to apply our dialogue act classes correctly. However, many annotators find a "decision tree" with questions to answer that end in a choice useful even if such a tree is always a simplification. Use the one in figure 3.1 as a memory aid, and not as a substitute for the complete manual.

3.8   Other Steps

In addition to segmenting the transcription into dialogue acts and classifying them according to their intention, our annotation includes marking some simple relationships between dialogue acts, indicating special cases in which an act is addressed not to the entire group but to someone in particular, and differentiating acts that are about the work that the group is doing from "reflexive" acts about how the group should do the work.

3.8.1   Describing Relationships

When people speak, they are often responding to something someone has said, something someone has done, or something around them. As you classify each dialogue act, you should consider whether or not it is a response to something, and if it is, indicate that relationship by adding a link. For instance, consider

Is the speaker's intention...

...not to take the floor?

...to say something only it's not clear what yet?

BACKCHANNEL

...to fill a pause
or take the floor?

STALL

FRAGMENT

...to make someone
feel good?

BE–POSITIVE

BE–NEGATIVE

...solely affecting the social structure of the group?

...to express (lack of)understanding
regarding something previously said?

...asking for a
comment?

ELICIT–COMMENT–
ABOUT–UNDERSTANDING

COMMENT–ABOUT–
UNDERSTANDING

...to evaluate or comment on something?

...asking for a
comment?

ELICIT–ASSESSMENT

ASSESS

...about a possible action?

...for someone else
to suggest an action?

ELICIT–OFFER–OR–SUGGESTION

OFFER

...to carry out
an action himself?

SUGGEST

...to exchange information?

...to ask for information?

ELICIT–INFORM

INFORM

OTHER

22

Figure 3.1: Schematic decision tree for the dialogue acts scheme.

the case of two dialogue acts, the first of which is a question (an ELICIT-INFORM), and the second of which is the answer (an INFORM). The answer occurs in response to the question, so you need to link the two together, using the question as the source of the link, and the answer as the target. This is the most usual kind of relationship, but there are other possible acts that could be related to the question, too. Refusing to answer the question, asking for it to be repeated, and saying that one doesn't know the answer are all possible responses, and therefore would require annotation.

(In the tool, the window that allows you to link two dialogue acts together says "adjacency pairs", but if you've read about adjacency pairs for your coursework, don't take this labelling seriously - you won't be marking all adjacency pairs as classically defined, and not all the relationships you will be marking are adjacency pairs.)

There will be cases when you may want to mark a relationship, but will not be able to identify a dialogue act as the source. This can happen for several reasons. The source may be something that isn't expressed verbally, such as for assessments of physical objects:

Example 38

	...upon picking up a whiteboard pen and stepping up to the whiteboard for the first time...
C	Okay	VERY NICE	alright

Or the source could be more than one previous dialogue act, as in assessments of entire discussions.

Example 39

C	Should we maybe make a decision about what features we actually want to include,	'CAUSE WE'RE THROWING A LOT OF FEATURES ONTO THE TABLE.

In these cases, just leave the source of the relationship unspecified.

In some sense, BACKCHANNELs are always in response to what the main speaker at that point is saying. There are so many BACKCHANNELs that it would be tedious to annotate the relationships this implies, and we can figure them automatically, so don't bother to annotate them. We also don't need you to annotate relationships that have BE-POSITIVE, BE-NEGATIVE, STALL, or FRAGMENT as the source or target. However, you should annotate relationships that have the more substantive acts as targets, including those that fall into the OTHER bucket class. As with ELICIT-INFORMATION and IN-FORM, you can expect to find relationships that have ELICIT-ASSESSMENT or ELICIT-COMMENT-ABOUT-UNDERSTANDING as source and ASSESS

or COMMENT-ABOUT-UNDERSTANDING as target, and between ELICIT-OFFER-OR-SUGGESTION and either OFFER or SUGGEST. You will also find relationships where the target is an ASSESS and the source is just about any other substantive act, and among various OFFERs, SUGGESTs, and IN-FORMs, as the group discusses what to do.

As well as indicating the source and target of a relationship, you need to classify the type of the relationship as POSITIVE, NEGATIVE, PARTIAL, or UNCLEAR.

POSITIVE means the target supports the intention of the source, for instance, by reacting positively to it, accepting or agreeing with it, indicating it has been understood, or providing what the source is attempting to elicit (see above examples);

NEGATIVE means it rejects the source, for instance, by presenting an objection to it, countering the source with an alternative the speaker prefers, or refusing to provide what the source is attempting to elicit;

Example 40

A	Mm. So, *some kind of idea uh with um um cellular phone with a a screen that will tell you what,* no .
C	NO, NO SCREENS | it's too complex.

PARTIAL means it partially supports the source but rejects it in some aspects, for instance by agreeing with part of a suggestion or providing part of what the source is attempting to elicit; and

Example 41

	[C is drawing on the whiteboard]
D	A kind of snake? *A cobra?*
C	Yeah, uh | NOT REALLY, | a small cobra.

UNCERTAIN means it expresses genuine uncertainty about the source, for instance, by saying that speaker is unsure whether or not a suggestion is a good idea or whether some information is true, or by expressing an inability to provide what the source is attempting to elicit.

Example 42

C	We can adapt only one switch, suppose here like we can make two switches and if I'm left-hander I use this switch to follow the main operations.
B	*I mean if it's less than three uh then we can make it uh like a...*
D	THREE BUTTONS, YOU MEAN?

3.8.2 Indicating Special Addressing Conditions

A speaker may address a dialogue act to the whole group of participants present in the meeting, or to a particular subgroup of them, and also to one single participant in particular. Your task is to mark the addressee(s) **for each dialogue act** that is not a BACKCHANNEL, STALL, FRAGMENT or OTHER. The last category includes cases of self-addressed speech, where a speaker mumble to himself or thinks aloud without really addressing anybody. Examples of this would be utterances like "Oops!" (after spilling water on the table), "Now, where am I?" (while trying to find the right slide on the laptop), "I have to get up" (while realizing that he cannot give a presentation from his seat). You do not have to make any distinction between addressing the entire group or addressing a subgroup of participants. In both cases, you should mark dialogue acts as addressed to a group.

Group vs. Individual

In a group discussion, many of the dialogue acts will simply be addressed to the group as a whole. However, at times a speaker shows by verbal or non-verbal behavior that he singles out one participant as the intended receiver of the dialogue act he performs. In such cases, only the participant receiving the primary attention of the speaker is the addressee.

There may be several reasons for a speaker to address an utterance to one specific participant:

1. the speaker expects a reaction, response or an action to be performed from the addressed individual (which does not mean that the addressee takes it that way)

Example 43

PM to ME		Can you go to the next slide?

Example 44

ME to PM	From from your side uh, you're gonna have to go back the management and s- be more s- precise

Example 45

PM to ID	Yeah but but end of the day, you're the sales guy, so I will come back and sit on your head because uh you are going to give your sales projection, okay.

Also, the speaker may explicitly announce to the addressee that he intends to ask him a question before the speaker actually asks the question. For example, "I have a question" can be paraphrased as "I have a question for you".

2. the speaker provides a *direct response* to a previous speaker who requested some information or opinion to be provided primarily to him.

Example 46

PM to ID	what's mean exactly, advanced chip on print? What's the meaning of that?
ID to PM	I think it's um um a multiple uh chip design um and it's uh maybe printed on to the circuit board.

Example 47

UI to ID	Why was plastic eliminated as a possible material?
ID to UI	Because um it gets brittle, cracks -

However, A will sometimes continue addressing B with the dialogue acts that follow such a direct response. In these cases it is not always easy to define whether a dialogue act is addressed to an individual or to the group. Other cues like gestures and body language, or knowing what role

each participant plays in the meeting may help define the directionality of the dialogue act in such ambivalent cases.

On the other hand, sometimes it is obvious from the content of the request and from the conversational context that the previous speaker intended that the selected individual or someone from the group should provide a response to the whole group.

Example 48

PM	to	So what do you think about uh the design?
ID		

Example 49

PM	to	Do you have any other information for us at this stage?
ME		

Example 50

		[the group is making decision what type of functions they would like to have.]
ID	to	Do we need a power button at all?
Group		

Responses to these types of questions are mostly addressed to the group.

3. the speaker provides an evaluation of what a previous speaker has said or done.

Example 51

UI	to	I think if you re- if you use really good quality wood, then it
Group		might work, but you can't just use...
ME	to	No y- no no no
UI		

27

Example 52

PM	to	Okay, tha- that would be great, so if you find out from the
ID		technology background, okay, so that would be good.
ID	to	sounds good
PM		

In other instances, however, such evaluations can be addressed to the whole group: in a group discussion, participants -sometimes simultaneously- may comment on decisions that they took together, or an idea that has just been suggested.

4. the speaker expresses himself in a positive or negative way towards an individual. For example, the speaker may apologize for something that he said or did to the one that is being addressed (e.g. interrupting), or make rude comments to that individual.

If a speaker is talking about another participant in the third person, perhaps indirectly addressing him, that participant should not be marked as the addressee. For example, if PM says to the group *"So, I ll invite uh Christine to discuss about uh the functional design"*, he is addressing the whole group, despite the fact that he is indirectly asking Christine to begin her presentation.

Sources of information

The speaker may explicitly address a single individual by, for example, using vocatives (*Christine, can you tell us about industrial design*). In most cases, the addressing is not explicit and you should use different sources of information regarding speaker verbal and non verbal behavior as well as the conversational situation, to identify who is being addressed.

The following list contains criteria that you should consider in identifying who is being addressed:

- **Content of dialogue act**- Sometimes the content of a dialogue act can provide enough information to identify the addressee(s) of the dialogue act. For example, *"Okay then, uh, let's move to Agnes"* and *"And finally in this meeting we have to decide together about the conceptual design"* are suggestions addressed to the group, whereas *Yeah but but end of the day, you're the sales guy, so I will come back and sit on your head because uh you are going to give your sales projection, okay"* is addressed to the marketing expert. On the other hand, the content of a dialogue act can give a clue that an individual is addressed, without refering to that individual. For example, *Do you have any other information for us at this stage?* indicates that an individual is addressed, but the other sources of information has to be considered to identify which individual it is.

- **Conversational situation**- In a number of cases, the person addressed will be the person who last spoke, the person whose speech was the stimulus for the present response. It may also happen that two participants start exchanging information about a certain issue, addressing each other. Two participants may also form a side conversation.

- **Meeting context**- the particular activity taking place can provide clues as to the addressee(s)
  During presentations, most of the dialogue acts are addressed to a group, unless there is a distinctive cue, that indicates that an individual is addressed. For example, the speaker may comment or refer to something that an individual has said previously, giving primary attention to that individual, or he makes a side remark, like *"I have added your figure on the slide"*.

- **Gestures and postures**- Looking at the single participant, or pointing at the single participant sometimes may indicate who is being addressed. But, it was not always necessary to believe that because A looks at B while saying something, he is addressing B. In group discussions a speaker's gaze may be directed to the person who last spoke, or to the person whose line of work is related to what is said or the person who coordinates the discussion i.e. PM, or simply at the person who is sitting in front of the speaker. The speaker's gaze can also function to monitor listeners, to see how they take up his words. If a speaker is monitoring one participant, it does not necessarily imply that he is addressing this participant in particular; he can be addressing the whole group. Therefore, you should not rely on gaze information only to identify the addressee. Gaze can be a supporting criterion but usually not a primary one. The same holds for pointing: the gesture can be used as a non-verbal reference to a participant, who is not actually the addressee.

Example 53

P1	to	As she *[pointing at P2]* has just said.
P3		

- **Domain knowledge**-The project manager will probably not ask the user interface designer how many of the tested people would like to have speech control.

Tag set

- Individual: A, B, C, D

- Group: the Group tag is used if more than one participant is addressed

- Unclassifiable (UNC): - annotators cannot determine who is being addressed by the speaker.

3.8.3 Marking Reflexive Acts

Most of the time during meetings, the participants are simply going about the business of whatever task they are doing (in the case of the scenario meetings, working out the design of the remote control). However, most groups will at one time or another step back to discuss not the task, but how they as a group are approaching the task. For instance, someone might suggest not that the remote control fascia should be red — a task level suggestion — but that the group should put off deciding what colour the remote control fascia should be until after they have determined the material from which it should be made. Utterances like these, which show someone from the group reflecting on how the group is carrying out its task, are rather special, because at least some people think that more of them make a group more likely to succeed, and because they can give good signposting that will help an outsider understand the meeting. For this reason, we want you to mark dialogue acts that are reflexive.

A dialogue act should be marked as reflexive if the content of the act is about how the group carries out the task. This could mean an act about how to have the meeting, such as a question asking whether or not there is time for the participants to introduce themselves, or a suggestion that the meeting should start with a presentation; it could mean an act about how to make a decision, for instance, a statement pointing out that voting doesn't always lead to the best solution; or it could mean an act about how to divide up the work, such as committing the project manager to performing budget calculations during a break between meetings. Any proper dialogue act class — that is, everything that isn't marked as BACKCHANNEL, STALL or FRAGMENT — can count as reflexive.

If you think about it carefully, there is a sense in which all STALLs are reflexive, since they express the intention that everyone else should be quiet so that the speaker can say something. That is, they are about how to have the meeting right at that point. That's true for all STALLs, so you need not mark them as reflexive.

All reflexive acts are about how the group should carry out their task. This doesn't mean that all non-reflexive acts are used by the group to carry out their task, though. Sometimes groups just chat aimlessly, for instance, about what they had for lunch. We don't have any special way of marking these other kinds of activities - you only need to mark the reflexive acts and leave anything else that happens unmarked.

Example 54

A	I'M FIRST GOING TO DO AN OPENING | THEN WE GET USED TO ONE ANOTHER | AND WE SPEAK ABOUT THIS TOOL WE'RE GOING TO DESIGN | AND TRY TO MAKE A PROJECT PLAN | SOME DISCUSSION | AND THEN WE TALK ABOUT THE NEXT MEETING. |

Example 55

D	| MEETING'S OVER? | GREAT. |

3.9   Possible confusions

There are always some possible confusions between classes in any scheme for classifying dialogue acts. Here are some we've thought of. Seeing whether or not you understand why we think there is a possible confusion and why we describe the difference between the classes as we do is a good test of whether or not you understand the coding scheme.

3.9.1   FRAGMENT versus STALL

You can tell FRAGMENTs from STALL cases by the material they cover. If the speaker continues after, and the material being coded has the same meaning and the same form as what follows, apart from minor rewordings, code it as a STALL. Otherwise, code it as a FRAGMENT.

3.9.2   FRAGMENT versus OTHER

Both FRAGMENT and OTHER are buckets, of sorts. FRAGMENT is for things that don't really convey a speaker intention, and so aren't properly dialogue acts, and OTHER, for things that do convey a speaker intention, and therefore are properly dialogue acts, just not ones that fit one of the classes we've defined.

3.9.3   Actions versus information exchange

It's easy to miss accidentally that dialogue acts are about possible actions, since if they are expressed indirectly, they can take the same forms as acts about information exchange. When you decide whether an act is about information exchange or about actions, the important thing to remember is that everything concerning what some person or the group does counts as an action. For instance, for the remote control design meetings, the group has to decide upon a colour for the remote control. "The remote control could be green" is about an action - that is, it is a suggestion that the group make the remote control green

- whereas "titanium came up tops on the marketing report" would be about information exchange, unless it was clear that by saying this, the speaker meant that the remote control should be made of titanium, in which case it would be an indirect suggestion about the action of making the remote control of titanium. This distinction is subjective, but listening will help you decide what the speaker's intention is. ELICIT-OFFER-OR-SUGGESTION tends to be fairly clear. A useful test for OFFER and SUGGEST is to think about whether an act could be rephrased starting "I/we/she/he/you could/should/will..." and still express the same intention.

3.9.4   BE-POSITIVE versus ASSESS

There can be a fine line between expressing an assessment and just being positive, especially if what's being assessed has nothing to do with what the group is trying to achieve, but is something personal. For instance, consider the case of commenting positively on someone's appearance - this is usually done for the social impact, and it's hard to imagine how such comments could contribute to the overall discussion unless there were a lengthy side discussion of such matters. Positive comments solely for social impact are not restricted to side matters, however; for instance, after a presentation, someone may well say "great job!", and this will not necessarily bear on the discussion that follows - in fact, it's more common when it hasn't been a particularly great job, and therefore the presenter needs bolstering.

If, in your judgment, the point of a comment is solely for the social impact, and not for the sake of contributing to the group's discussion of whatever the comment is about, then it counts as a BE-POSITIVE, not an ASSESS. This really is a judgment call, since many forms of speech could go either way. For instance,

Example 56

ID	I BET NOKIA COULDN'T MAKE A REMOTE CONTROL AS GOOD AS THIS

could be an ASSESS if the speaker is using his contribution to express the opinion (in a jokey, and clearly positive way) that the remote control design is a good one in the context of a discussion of the design, but it's more likely to be intended solely to make the group feel happier.

3.9.5   BE-POSITIVE versus BE-NEGATIVE

One thing to be careful about is jokes at someone else's expense. In some groups where social functioning is very smooth, people will mock-insult each other as a way of bolstering the group. For instance, for recording purposes the group participants are required to wear head microphones. "That's a very interesting

hairdo you've got there Mike" would be an example of BE-POSITIVE, even though it could be taken as a negative comment about Mike's appearance. When referring to the playdough models in the remote control scenario, "Come on, a two-year-old could have made that!" could also be positive. Jokes also, of course, are sometimes taken the wrong way by their targets. What matters is your judgment about the speaker's intention in making the joke.

3.9.6 SUGGEST versus all types of elicitation

There are three kinds of acts that attempt to elicit acts from the other meeting participants - ELICIT-INFO, ELICIT-ASSESSMENT and ELICIT-OFFER-OR-SUGGESTION. In a sense, all of these are suggestions - they suggest that the meeting participants perform the act of speaking with a particular intention. Remember to use the more specific ELICIT acts, and reserve SUGGESTION for acts that suggest other kinds of actions.

3.9.7 STALL versus reflexive OFFER to speak

When speakers aren't ready with the content of what they want to say yet, instead of using the kind of stalling behaviour that the STALL category covers, they might explicitly announce that they wish to say something (for instance, "Hold on, I want to say something"). They can also attempt to reserve ahead of time a longer speaking slot than normal by telling the group, for instance, there are three points they want to make. Although the effect of such behaviours is much like what a STALL does, they are quite different in character from the hesitations and restarts covered by the STALL category, and can cause further discussion (about scheduling) separate from the content that the speaker wants to produce. Because of this, they should be treated as reflexive OFFERs to perform the action of speaking, just like any other suggestions for how to get through the discussion.

3.9.8 Instances of "okay" and similar forms like "yeah"

"Okay" can be tricky to code because it can occur in many different contexts: as a BACKCHANNEL that does not attempt to interrupt another speaker but happens in the background; as an ELICIT-ASSESSMENT, particularly if said as if a question; as an ASSESS that, for instance, accepts the truth of some information; as an OFFER in response to a specific SUGGESTION; since it isn't very contentful, as a STALL while the speaker thinks what to say; and probably as other things besides. You need to take into account both what others have just said and the intonation before you decide what to do. The same holds for other, similar forms like "yeah".