Meeting Recorder Project: Dialog Act Labeling Guide

ICSI Technical Report TR-04-002
February 9, 2004

Rajdip Dhillon
Sonali Bhagat
Hannah Carvey
Elizabeth Shriberg

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
09 FEB 2004		**00-00-2004 to 00-00-2004**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Meeting Recorder Project: Dialog Act Labeling Guide	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **International Computer Science Institute (ICSI),Berkeley,CA,94704**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Same as Report (SAR)**	**132**	

ACKNOWLEDGEMENTS

We especially thank Jeremy Ang for processing our data and Chuck Wooters for providing us with the TableTrans software. We are also grateful to Don Baron and Chris Oei for their assistance in preparing data for annotation. We are thankful to Ashley Krupski for her annotation efforts; Barbara Peskin and Jane Edwards for their assistance in using the corpus; and Dan Jurafsky and Andrei Popescu-Belis for supplying us with their input.

This work was supported by a DARPA Communicator project, ICSI NSF ITR Award IIS-0121396, SRI NASA Award NCC2-1256, SRI NSF IRI-9619921, SRI DARPA ROAR project N66001-99-D-8504, and by an ICSI award from the Swiss National Science Foundation through the research network IM2. The views represented herein are those of the authors and do not represent the views of the funding agencies.

TABLE OF CONTENTS

INTRODUCTION

This labeling guide is adapted from work on the Switchboard recordings and the accompanying manual (Jurafsky et al. 1997). The Switchboard-DAMSL (SWBD-DAMSL) manual for labeling one-on-one phone conversations provided a useful starting point for the types of dialog acts (DAs) that arose in the ICSI meeting corpus. However, the tagset for labeling meetings presented here has been modified as necessary to better reflect the types of interaction we observed in multiparty face-to-face meetings.

This guide consists of five major sections: Quick Reference Information, Segmentation, How to Label, Adjacency Pairs, and Tag Descriptions. The first section supplies definitions for terms used throughout this guide and contains the correspondence of the Meeting Recorder DA (MRDA) tagset, which is the tagset detailed within this guide, to the SWBD-DAMSL tagset. This section also contains the entire MRDA tagset organized into groups according to syntactic, semantic, pragmatic, and functional similarities of the utterances they mark. The section entitled “Segmentation,” as its name indicates, details the rules and guidelines governing what constitutes an utterance along with how to determine utterance boundaries. The third section, “How to Label,” provides instruction regarding label construction, the management of utterances requiring additional DAs or containing quotes, and the use of the annotation software. The section entitled “Adjacency Pairs” details how adjacency pairs are constructed and the rules governing their usage. The section entitled “Tag Descriptions” provides explanations of each tag within the MRDA tagset.

Two appendices are also found within this guide. The first provides a labeled portion of a meeting and the second contains information regarding tags used for a select number of meetings.

With regard to the examples from meeting data found throughout this guide, it must be noted that the start and end times for each utterance within the examples do not reflect the most recent time alignments. However, the start and end times are accurate to a point which allows for them to be located within their corresponding audio files without difficulty.

SECTION 1: QUICK REFERENCE INFORMATION

1.1 Terminology

Below is some rudimentary terminology used in dialog act labeling:

utterance: a segment of speech occupying one line in the transcript by a single speaker which is prosodically and/or syntactically significant within the conversational context

speech: a group of successive utterances or successive portions of an utterance

turn: the period during which a speaker has the floor

label: the entire set of DAs and/or other tags applicable to an utterance

dialog act (DA): the tag or sequence of tags pertaining to the function of an utterance or portion of an utterance. Each DA contains at least one general tag and may contain one or more specific tags, depending upon the nature of the utterance

tag: the individual component(s) of a DA or label

general tag: the tag which represents the basic form of an utterance (e.g., statement, question, backchannel, etc.)

specific tag: the tag which represents the function or a characteristic of an utterance and is appended to the general tag (e.g., accepting, rejecting, acknowledging, rising tone, etc.)

disruption form: the tag which represents a disruption or otherwise indiscernible utterance

1.2 Mapping Meeting Recorder DA (MRDA) Tags to SWBD-DAMSL Tags

The following table shows the correspondence between Switchboard-DAMSL (SWBD-DAMSL) dialog tags and those used to label Meeting Recorder DA (MRDA) data. The tags within the table are ordered according to the categorical structure within the SWBD-DAMSL manual, with tags unique to the MRDA tagset being inserted in accordance with this categorical structure. The SWBD-DAMSL categories are not explicitly marked within this table in order to avoid confusion with the categories of the MRDA tagset.

Tags listed in *italics* are based upon SWBD-DAMSL tags but have had their meanings altered for the purposes of the MRDA data. Tags in **boldface** are not in the original SWBD-DAMSL manual but have been added to accurately characterize the MRDA data. Tag titles in **boldface** correspond to names of MRDA tags. All other tag titles correspond to names of SWBD-DAMSL tags.

Additionally, the reasoning behind why certain SWBD-DAMSL tags are not used in the MRDA tagset is found in Appendix 2. Explanations regarding the presence of tags unique to the MRDA tagset are found in Appendix 3.

TAG TITLE	SWBD-DAMSL	MRDA
Uninterpretable	***%***	***%***
Abandoned	%-	%--
Interruption	not marked	**%-**
Nonspeech	x	x
Self-talk	t1	t1
3rd-party-talk	t3	t3
About-task	t	t
About-communication	c	not marked
Statement-non-opinion	*sd*	*s*
Statement-opinion	*sv*	*s*
Open-option	oo	not marked
Yes-No-question	qy	qy
Wh-Question	qw	qw

Open-Question	qo	qo
Or-Question	qr	qr
Or-Clause	qrr	qrr
Rhetorical-Question	qh	qh
Declarative-Question	d	d
Tag-Question	g	g
Action-directive	*ad*	*co*
Offer	*co*	*cs*
Commit	cc	cc
Conventional-opening	fp	not marked
Conventional-closing	fc	not marked
Explicit-performative	fx	not marked
Exclamation	fe	fe
Other-forward-function	fo	not marked
Thanks	ft	ft
Welcome	fw	fw
Apology	fa	fa
Topic Change	not marked	**tc**
Floor Holder	not marked	**fh**
Floor Grabber	not marked	**fg**
Accept	*aa*	*aa*
Accept-part	aap	aap
Maybe	am	am
Reject-part	arp	arp
Reject	*ar*	*ar*
Hold before answer/agreement	h	h
Signal-non-understanding	br	br
Continuer	b	b

Rhetorical-question continuer	bh	bh
Acknowledge-answer	bk	bk
Mimic other	m	m
Repeat	not marked	**r**
Collaborative completion	2	2
Reformulate/summarize	bf	bs
Assessment/appreciation	ba	ba
Sympathy	by	by
Downplayer	bd	bd
Correct-misspeaking	bc	bc
Misspeak Self-Correction	not marked	**bsc**
Understanding Check	not marked	**bu**
Defending/Explanation	not marked	**df**
"Follow Me"	not marked	**f**
Yes answers	*ny*	*aa*
No answers	*nn*	*ar*
Affirmative non-yes answers	na	na
Negative non-no answers	ng	ng
Other answers	no	no
Expansions of y/n answers	*e*	*e*
Dispreferred answers	nd	nd
Quoted Material	q	not marked
Hedge	h	not marked
Continued from previous line	+	not marked
Humorous Material	not marked	**j**
Rising Tone	not marked	**rt**
Nonlabeled	not marked	**z**

1.3 Meeting Recorder DA (MRDA) Tagset

The categorization scheme for the Meeting Recorder DA (MRDA) tagset differs from the scheme employed for the SWBD-DAMSL tags seen. The reasoning behind this is that, in the process of adjusting the definitions of previously established SWBD-DAMSL tags and creating new tags to assist in adequately assessing the MRDA data, the resulting MRDA tagset could not be appropriately characterized when placed in direct relation to the SWBD-DAMSL tagset, given the nature of the data for which the MRDA tagset was employed. Consequently, the tags are not organized on a dimensional level, but rather the correspondences for the MRDA tagset are listed on the tag level. Descriptions of the individual tags within the MRDA tagset are found in Section 5.

Group 1: Statements
- s Statement

Group 2: Questions
- qy Y/N Question
- qw Wh-Question
- qr Or Question
- qrr Or Clause After Y/N Question
- qo Open-ended Question
- qh Rhetorical Question

Group 3: Floor Mechanisms
- fg Floor Grabber
- fh Floor Holder
- h Hold

Group 4: Backchannels and Acknowledgements
- b Backchannel
- bk Acknowledgement
- ba Assessment/Appreciation
- bh Rhetorical Question Backchannel

Group 5: Responses

Positive
- aa Accept
- aap Partial Accept
- na Affirmative Answer

Negative
- ar Reject
- arp Partial Reject
- nd Dispreferred Answer
- ng Negative Answer

Uncertain
- am Maybe
- no No Knowledge

Group 6: Action Motivators

- co Command
- cs Suggestion
- cc Commitment

Group 7: Checks

- f "Follow Me"
- br Repetition Request
- bu Understanding Check

Group 8: Restated Information

Repetition

- r Repeat
- m Mimic
- bs Summary

Correction

- bc Correct Misspeaking
- bsc Self-Correct Misspeaking

Group 9: Supportive Functions

- df Defending/Explanation
- e Elaboration
- 2 Collaborative Completion

Group 10: Politeness Mechanisms

- bd Downplayer
- by Sympathy
- fa Apology
- ft Thanks
- fw Welcome

Group 11: Further Descriptions

- fe Exclamation
- t About-Task
- tc Topic Change
- j Joke
- t1 Self Talk
- t3 Third Party Talk
- d Declarative Question
- g Tag Question
- rt Rising Tone

Group 12: Disruption Forms

- % Indecipherable
- %- Interrupted
- %-- Abandoned
- x Nonspeech

Group 13: Nonlabeled

- z Nonlabeled

SECTION 2: SEGMENTATION

Utterance segmentation is one of the most debated topics in discourse analysis. The function of dialog must always be considered when determining utterance boundaries. Lengthy utterances containing multiple conjunctions, speaker rambling, and floor-holding are just a few factors complicating the decisions regarding utterance boundaries. In order to segment transcribed speech into distinguishable utterances, the following factors are taken into consideration within the context of the conversation: syntax, pragmatic function, and prosody.

Prior to determining how to segment transcribed speech, knowledge of how utterance boundaries are marked within the transcript is necessary. There are two ways to mark utterance boundaries within the transcript. When a speaker trails off or is interrupted and consequently does not complete his utterance, an utterance boundary in the form of <==> is marked at the end of the corresponding utterance in the transcript. In Example 1 on the following page, speaker c2 does not finish his utterance (speaker c3 adds the remainder of c2's utterance shortly after) and an utterance boundary is signaled by the <==> in the transcript. If a speaker's utterance is complete, an utterance boundary in the form of < . > is marked at the end of the corresponding utterance in the transcript.

Returning to the factors involved in segmentation, in terms of syntax, utterance boundaries are primarily derived on a phrasal level. This is not to say that an utterance consists only of a noun phrase or a verb phrase, but rather that it is permitted for a complete utterance to consist only of a noun phrase, a verb phrase, or both. In Example 1[1], the noun phrase "jose" constitutes a complete utterance:

Example 1: Bmr010

280.000-284.762	c2	s.%--	and i did some training on - on one dialogue which was transcribed by ==
284.762-288.568	c2	s	yeah we - we did a nons- - s- - speech nonspeech transcription .
287.474-288.294	c3	s^2	jose.

Example 2 and 3 depict instances where verb phrases, "got it" and "wants to conserve" in Example 2 and "confused" in Example 3, behave as complete utterances:

1 Examples take a format in which the numerical values of the first column represent start and end times of utterances, the second column indicates the channel, the third indicates the DA, and the fourth presents the transcript.

Example 2: Bed011

114.007-116.680	c2	s	and um - i - i told it to stay on forever and ever .
116.680-119.347	c2	s	but if it's not plugged in it just doesn't obey my commands .
119.120-119.320	c1	s^bk	okay .
119.726-120.386	c2	s	it has a mind .
121.961-122.331	c1	s^bk	got it .
122.160-123.170	c4	s	wants to conserve .

Example 3: Bed003

2950.850-2957.110	c3	s	yeah the only like - possible interpretation is that they are - like - come here just to rob the museum or something to that effect .
2952.260-2953.830	c2	s^2	confused .

The pragmatic function of an utterance is also an important consideration for utterance boundary identification. Phrases or clauses that do not appear complete grammatically may actually form complete utterances on account of having unique functions within conversation. Although it may seem peculiar to segment utterances on a phrasal and clausal level, such a method of segmentation is utilized for the purpose of maximizing the amount of information derived from DAs.

Example 4 presents an utterance that appears complete grammatically, yet does not maximize the amount of information which can be derived from DAs.

Example 4: Bmr010

217.921-227.363	c6	s^cs	that uh - if we had something that worked for many cases before maybe starting from there a little bit because ultimately we're going to end up with some s- - kind of structure like that.

In Example 5, the same utterance from Example 4 is shown, however the utterance is segmented at the clausal level so that more information may be provided by the DAs that otherwise would not be present had the utterance not been segmented.

Example 5: Bmr010

217.921-222.161	c6	s^cs	that uh - if we had something that worked for many cases before maybe starting from there a little bit .
222.161-227.363	c6	s^df	because ultimately we're going to end up with some s- - kind of structure like that.

Syntax and pragmatic function are both taken into account when encountering conjunctions. Conjunctions such as "and," "or," "but," and "so" often behave as cues to locations where a string of clauses might be segmented into separate utterances. Rather than simply start a new utterance, a speaker might use one of these conjunctions as a connection between two complete utterances, as seen in a pre-segmented utterance in Example 6:

Example 6: Bmr020

595.187-608.363	c6	s	that's somewhat - that's somewhat subject to error but still we - we uh don did some ha- - hand checking and – and we think that - based on that we think that the results are you know valid although of course some error is going to be in there .

Example 7 depicts a correctly segmented version of Example 6:

Example 7: Bmr020

595.187-596.880	c6	s	that's somewhat - that's somewhat subject to error .
596.880-601.180	c6	s	but still we - we uh don did some ha- - hand checking .
601.310-604.837	c6	s	and - and we think that - based on that we think that the results are you know valid .
604.837-608.363	c6	s	although of course some error is going to be in there .

Caution must be taken not to segment utterances upon the appearance of conjunctions in every instance. Quite often, conjunctions are used to simply connect noun phrases or verb phrases that would not constitute separate utterances in the context in which they are used. In these cases, the utterance is not segmented at the conjunction.

Example 8 and Example 9 demonstrate instances when an utterance is not segmented upon the appearance of a conjunction:

Example 8: Bro014

238.387-240.098	c2	s^e	i mean it's like one little text file you edit and change those numbers .

Example 9: Bro014

302.417-305.275	c2	s	now h t k's compiled for both the linux and for um the sparcs .

On occasion, a speaker may have an extremely lengthy utterance with many conjunctive clauses and parentheticals. In such situations, each clause or parenthetical is segmented into a separate utterance. As with segmenting on a clausal or phrasal level, segmenting parentheticals in such a way allows for the maximization of information provided by DAs. In deciding how to segment such instances within transcribed speech, it is helpful to determine whether a speaker actually had the whole string of speech in mind or else unintentionally diverged from his original thoughts. Example 10 depicts a rather lengthy utterance prior to segmentation and Example 11 presents a segmented version of the same utterance.

Example 10: Bmr005

1012.960-1033.300	c4	s	but i - i mean - i think also to some extent its just educating the human subjects people in a way because there's if uh - you know - there's court transcripts there's - there's transcripts of radio shows i mean - people say people's names all the time so i think it - it can't be bad to say people's names it's just that i mean - you're right that there's more poten- - if we never say anybody's name then there's no chance of - of - of slandering anybody .

Example 11: Bmr005

1012.960-1019.350	c4	s	but i - i mean - i think also to some extent its just educating the human subjects people in a way .

1019.350-1025.740	c4	s^df	because there's if uh - you know - there's court transcripts there's - there's transcripts of radio shows i mean - people say people's names all the time .
1026.390-1028.940	c4	s	so i think it - it can't be bad to say people's names .
1029.270-1033.300	c4	s^df	it's just that i mean - you're right that there's more poten- - if we never say anybody's name then there's no chance of - of - of slandering anybody .

Prosody is also of considerable importance in detecting utterance boundaries. To take the prosody of an utterance into consideration is to take the aural cues such as the rise and fall of pitch, the energy level, and duration of the words of the utterance as well as the complete utterance into consideration. Utterances that appear complete syntactically, whether they are quite lengthy or consist of short phrases or clauses, may be incomplete prosodically. If the prosody of the end of an utterance consists of a pitch, energy level, or duration that is incongruent with that of a complete utterance, then that particular utterance is considered incomplete. General prosodic patters found within complete utterances and prosodic patterns specific to certain speakers are necessary factors in determining how to assess the prosody of a complete utterance.

Prosody is of use in determining whether an utterance is interrupted or abandoned. If a speaker begins trailing off in pitch and the energy level begins to decrease, the speaker's utterance is most likely to be marked as abandoned. Prosody can also help distinguish between floor grabbers and backchannels, as floor grabbers tend to have a higher energy level in contrast to the surrounding speech and backchannels do not.

Pauses also behave as signifiers to utterance boundaries. Oftentimes, the appearance of a lengthy pause indicates that the segment of speech following the pause constitutes a new utterance. If the portion of speech immediately preceding the pause is incomplete, that portion may either be an abandoned utterance or the beginning of an utterance of which the portion of speech following the pause is the end. If the former applies, and the portion preceding the pause is actually abandoned, a change in DAs, prosody, or both is an obvious signal that the pause is indicative of a boundary. However, if the latter case is applicable, no such drastic change in the prosody between the segment preceding and the segment following the pause will be present and both portions of speech are to comprise one utterance. To reiterate with regard to the latter case, an utterance boundary will not be marked at the pause. As a side note, it must be mentioned that some speakers tend to speak slowly in such a manner that their utterances are filled with frequent pauses. In such instances, pauses are not indicators of utterance boundaries unless the segment of speech following a pause is incongruent with the segment preceding.

As difficulty in determining utterance boundaries is encountered when considering the factors of syntax, prosodic function, prosody, and pauses, additional segmentation issues occasionally arise with the applicability of certain tags, namely <fg>, <fh>, <h>, <aa>, <ar>, <bk>, and <g>. Regarding <fg>, <fh>, and <h>, often the problem at hand is whether to segment an utterance in which a speaker utters a string of <fg>s, <fh>s, or <h>s, as seen in Example 12. If there exist significant pauses between each portion of the string of <fg>s, <fh>s, or <h>s, the utterance is segmented upon each pause and each resulting utterance is labeled appropriately as <fg>, <fh>, or <h>, depending upon its nature. However, if no such significant pauses exist, then the entire utterance remains intact and receives a suitable label. Additionally, it is far more difficult to judge if a pause actually signifies an utterance boundary within strings of <fg>s, <fh>s, or <h>s than within strings of fluent speech.

Example 12: Bmr012

1886.800-1891.3100	c1	s^cs	and then just sort of have that as the - and then you can have groups of twenty people or whatever .
1891.310-1892.080	c1	fh	and - and uh ==

As a general convention, unless an utterance is comprised solely of floor holders, it is not to end with a floor holder <fh>. In the case that a floor holder is found at the end of an utterance, it is split from the utterance and either receives its own line or is merged with the following utterance of the same speaker, depending primarily upon its prosody and its temporal proximity to the following utterance. If the length of the floor holder is incongruent to the length of the words of the following utterance, the floor holder is of a different intonation in relation to the following utterance, or a significant pause exists between the floor holder and the following utterance, the floor holder is not merged with the following utterance. If the floor holder is merged with the following utterance and the following utterance is not a floor holder, then it is permissible for the resulting utterance, which consists of a floor holder and another DA, to contain multiple DAs. Additionally, although a floor grabber and a hold do not occur mid-speech as a floor holder does, these tags may also be merged with the following utterance if deemed necessary and the resulting utterance will also contain multiple DAs. Section 3.3 specifies the manner in which utterances with multiple DAs are treated.

After splitting a floor holder from an utterance, it must be decided whether the portion which originally preceded the floor holder is complete or incomplete. Example 13 depicts an utterance ending with a floor holder and the same utterance is seen in Example 14 with the exception that the utterance has been segmented so that the floor holder receives its own line.

Example 13: Bmr010

601.519-604.014	c0	s	and if it's good enough we'll arrange windows machines to be available so ==

Example 14: Bmr010

601.519-602.707	c0	s	and if it's good enough we'll arrange windows machines to be available .
603.465-604.014	c0	fh	so ==

Regarding the tags <aa>, <ar>, <bk>, and <g>, the largest problem is determining whether or not an utterance boundary exists after speech labeled with the tag <aa>, <ar>, or <bk>, that is if speech from the same speaker immediately follows, or if a boundary exists before speech labeled with the tag <g>, that is if speech from the same speaker immediately precedes the portion labeled with the tag <g>. This problem only emerges if the speech surrounding the portions labeled with the tags previously specified is such that the prosody bears no indication of a boundary between utterances, the speaker speaks so quickly that a boundary cannot be discerned, or else no significant pause is found to mark a boundary. When the issue arises that a boundary cannot be marked between speech labeled with the previously mentioned tags and the surrounding speech, then it is permissible for an utterance to have multiple DAs. Section 3.3 details the format of labels for utterances which have multiple DAs.

Another issue regarding segmentation concerns otherwise complete utterances being segmented in such a way that yields abandoned utterances. For instance, a complete utterance may be quite lengthy and appear as though it ought to be segmented. However, segmenting the utterance may yield incomplete utterances that would be marked as abandoned. As the original intact utterance is complete and some of the segmented portions are marked as being abandoned, it is clear that segmenting the utterance in a way that yields abandoned utterances is incorrect.

As an addendum to the aforementioned system of segmentation, if uncertainty exists as to whether or not to segment an utterance, a general guideline is to segment the utterance regardless. Also, portions of speech that constitute one utterance but for some reason, perhaps mistakenly, are segmented as multiple utterances are merged to form one utterance.

SECTION 3: HOW TO LABEL

3.1 Basic Format of DAs and Labels

The basic format of a **DA** is as follows[2]:

<general tag> [^ specific tag]

The basic format of a **label** is as follows (depending upon the utterance, the portions enclosed in brackets may or may not be necessary):

<general tag> [[**^** <specific tag>] [**|** <general tag> [**^** <specific tag>]] [**.** <disruption form>]]

3.2 Label Construction

The general tag is a mandatory component of every label. Only one general tag is present in each DA. Specific tags and disruption forms (which indicate when a speaker has been interrupted, trails off, or else is indecipherable) are included within a label only when an utterance cannot be sufficiently characterized by a general tag and when further characterization is needed. Specific tags are appended to general tags when necessary and are not used alone. For the purpose of uniformity among annotators, when multiple specific tags are appended to a general tag, they are attached in alphabetical order[3].

In the following sets of tags, the first set contains general tags, the second set contains specific tags, and the third set contains disruption forms. Detailed descriptions of the tags in the three sets can be found in Section 5. Note that the tags found in Set 1 are

[2] Throughout this manual, when discussing format, the convention of enclosing portions in brackets denotes that, depending upon an utterance, those portions may or may not be necessary.

[3] As specific tags are attached in alphabetical order, the tag <2> is the last tag within the alphabetically ordered hierarchy, rather than the first.

only used as general tags, the tags found in Set 2 are only used as specific tags (in conjunction with a general tag), and tags in Set 3 are only used as disruption forms.

Set 1: General Tags

s	qy	qw	qr	qrr	qo	qh	b	fg	fh	h

Set 2: Specific Tags

aa	aap	am	ar	arp	ba	bc	bd	bh	bk	br
bs	bsc	bu	by	cc	co	cs	d	df	e	f
fa	fe	ft	fw	g	j	m	na	nd	ng	no
r	rt	t	tc	t1	t3	2				

Set 3: Disruption Forms

Disruptions

%-	%--

Indecipherable

x	%

Within a DA, when specific tags are necessary, they are attached to the general tag with a caret (^), thus rendering the following depiction of a DA:

< general tag >^< specific tag $_1$ >^< specific tag $_2$ >^< specific tag $_3$ > ...^< specific tag $_n$ >

Disruption forms are attached to and separated from the end of a DA with a period < **.** >, as seen in the following representation:

< general tag > [^ < specific tag $_1$ > ...^< specific tag $_n$ >] **.** < disruption form >

It must be noted that, in some cases, a disruption form is present within an utterance without sufficient information to assign a DA to that utterance. In such instances, a label comprised solely of a disruption form is necessary.

Additionally, if for some reason an utterance is not to be labeled with a DA, then that particular utterance receives a label consisting only of the tag <z>. For instance, if an utterance contains data that is not to be labeled on account of it containing digits, containing pre- or post-meeting chatter, pertaining to a "bleeped" portion in the corresponding audio file, or else is simply not relevant to the labeling task, a label comprised solely of the tag <z> is used. As the tag <z> is used to mark utterances which otherwise would be labeled with DAs but instead are intentionally not to be labeled, it is clear why the tag <z> is not included within the other groups of tags (i.e. general tags, specific tags, and disruption forms). The tag <z> does not provide any information regarding the characteristics and functions of utterances as the tags of the other groups do, and for this reason it is separated from those groups.

The following is a partial list of sample labels that are acceptable within the previously established conventions for label construction:

s	qy	qr	b	fg	%
s^bk	qy^d^f^g^rt	qr^rt	b.%	fh^rt	%-
s^nd	qy^bh	qrr.%--	b.x	h	%--
s^aa^rt.%--	qy^bu.%-	qh^rt.%	b^rt	z	x

Listed below is an incomplete list of sample labels that are not acceptable within the previously established conventions for label construction:

s^s	aa^bk	x.%--	%--.s^qy^d	s^z
s^s^aa	%.%--	%--.x	b.%-	z.%--

It is worthy of mention that other restrictions apply in constructing labels. Such restrictions include particular specific tags which may only appear with certain general tags, particular general tags which have a limited set of applicable specific tags, and sets of specific tags which are prohibited from appearing in the same DA. Restrictions applying to the usage of tags are discussed in the individual tag descriptions in Section 5.

3.3 Annotating Utterances Containing Multiple DAs

In cases where one DA does not suffice to represent an utterance, two DAs are used. Such a need arises in cases as those described in Section 2, usually with tags such as <fg>, <fh>, <h>, <aa>, <ar>, <bk>, and <g> which correspond to short utterances.

Often, an utterance requires multiple DAs when a floor grabber <fg> or floor holder <fh> is uttered at the beginning of a statement <s> or question, when a short answer of the nature <aa>, <ar>, or <bk> is following by a longer explanation, or when a statement is followed by a tag question <g>. In some cases, an utterance requires multiple DAs when a statement <s> is followed by a short answer of the nature <aa>, <ar>, or <bk>. In which case, the DAs can be separated in both the label and the portion of the transcript containing the utterance with a pipe bar < | >.

The pipe bar < | > is only used when sequential portions of an utterance that operate closely together require different characterizations. For instance, a pipe bar is not used for an agreement <aa> and a question that immediately follows it. In fact, an agreement followed by a question does not constitute an utterance but constitutes two separate utterances instead. Rather, an agreement immediately followed by an explanation of the agreement, a longer, narrative form of agreement, or a direct reference to what the agreement regards would require a pipe bar so long as the prosody and lack of significant pauses warrants such usage of a pipe bar.

The use of a pipe bar indicates that segmenting an utterance is not necessary, despite that the initial portion of an utterance, or last portion in the case of <g>, has a different DA than the rest of the utterance.

The pipe bar is indicated in the appropriate location within the label as well as within the transcription. Within the label, the pipe bar separates the DAs. Within the transcript, the pipe bar separates the portions of an utterance to which the different DAs apply. This is done in such a manner that the DA to the left of the pipe bar in the label pertains to the portion of the utterance to the left of the pipe bar in the transcript and the DA to the right of the pipe bar in the label pertains to the portion of the utterance to the right of the pipe bar in the transcript.

Example 1 demonstrates the correct usage of a pipe bar, whereas Example 2 and Example 3 depict the incorrect usage of a pipe bar.

Example 1: Bmr012

94.861-99.771	c4	fg|s^t	um - | everyone should have at least two forms possibly three in front of you depending on who you are .

Example 2: Bmr012

94.861-99.771	c4	s^t|fg	um - | everyone should have at least two forms possibly three in front of you depending on who you are .

Example 3: Bmr012

94.861-99.771	c4	fg|s^t	um - everyone | should have at least two forms possibly three in front of you depending on who you are .

3.4 Disruption Forms

Disruption forms are used to mark utterances that are indecipherable, abandoned, or interrupted. Only one disruption form may be used per utterance.

Disruption forms are included in a label in one of three formats, depending upon the nature of an utterance. When a DA is not detected, a disruption form alone may comprise an entire label. When used in conjunction with a DA, disruption forms are marked using either a period < **.** > or a pipe bar < | >.

If an utterance contains a disruption form and is too short to determine which DA applies to it, then only the disruption form is marked in the label. An utterance that is indecipherable may actually be quite lengthy, but because it cannot be deciphered, an appropriate DA cannot be assigned to it and only the disruption form is marked. Example 4 depicts a disrupted utterance which contains insufficient information to provide a DA:

Example 4: Bro014

1207.310-1207.880	c1	%-	but i- ==

Exceptions occasionally apply to short utterances deemed indecipherable. Utterances which appear to be backchannels, for instance, yet are indecipherable may be labeled with the appropriate DA along with a period and the applicable disruption form. Such treatment of indecipherable utterances is only employed when there is a high probability that the specific DA applies to the utterance based upon the surrounding context of the short utterance and the speaker's speech patterns. The following are two sample labels pertaining to short indecipherable utterances:

b.%	b.x

A period or a pipe bar is used in conjunction with a disruption form if a disruption form is indeed applicable to an utterance and if an utterance contains sufficient information to assign to it a DA. For instance, if an utterance, such as a statement, is interrupted or abandoned, the DA is marked and then followed by a period and the appropriate disruption form, as seen in Example 5:

Example 5: Bro014

495.681-499.134	c4	s.%--	some people are arguing that it would be better to have weights on ==

In the case of Example 5, the utterance contains sufficient information to determine that it is indeed a statement, despite being abandoned. If an utterance does not contain adequate information to decide which DA applies to it, then a DA is not marked.

Two types of instances exist in which an utterance containing a pipe bar requires a disruption form. In the first, an utterance requiring a pipe bar, such as what is discussed in Section 3.3, is either abandoned or incomplete. To the left of the pipe bar is a DA containing a tag such as <fg> or <aa> and to the right is a statement or explanation of some sort that is either incomplete or abandoned. Note that the disruption form only applies to the DA to the right of the pipe bar. Keeping in mind that the portion of the utterance to the right of the pipe bar contains sufficient information to assign to it a DA and is also abandoned or incomplete, its DA is followed by a period and the appropriate disruption form, as seen in Example 6:

Example 6: Bro014

1897.760-1904.500	c0	s^bk|s.%--	yeah | hopefully i think what we want to have is to put these features in s- - some kind of ==

In the second instance in which an utterance containing a pipe bar requires a disruption form, the portion of the utterance to the right of the pipe bar does not contain sufficient information to assign to it a DA. This portion may be abandoned, interrupted, or indecipherable. The DA designated to the portion of the utterance to the left of the pipe bar clearly begins upon the onset of the utterance and ends at the point where the pipe bar is placed. The DA pertaining to the initial portion of the utterance is marked, a pipe bar is placed after the DA in the label and at the point where that particular DA ends in the transcript, and a disruption form is marked after the pipe bar, as seen in Example 7 and Example 8:

Example 7: Bmr028

1187.370-1188.240	c1	fg|%-	yeah | he ==

Example 8: Bro014

403.710-405.428	c2	s^aa|%--	yeah | it's uh ==

The distinction between the use of the pipe bar and a period exists in how an utterance can be divided. An utterance divided by a pipe bar behaves in some ways as two separate utterances. The segment of the utterance to the left of the pipe bar will be annotated with a particular DA that is different from the DA used to annotate the right, that is if it is possible to assign a DA. The pipe bar exists as a clear boundary which marks where one DA ends and another begins in a single utterance. The portion to the right of the pipe bar behaves as a separate utterance in that it alone is the specific segment which is interrupted, abandoned, or indecipherable. The portion to the left is complete.

With regard to periods, and even labels consisting solely of disruption forms, no clear and comparable boundary as found in utterances requiring pipe bars exists. The exact region within an utterance where the disruption form occurs does not behave as a separate segment of the utterance that can be marked clearly with a mechanism such as a pipe bar. It is also unnecessary to use a pipe bar to mark where an interruption begins or where a speaker abandons his utterance, since the DA to the left of the pipe bar may also apply to the other side where the disruption form is marked.

Additionally, the reasoning behind why a disruption form is not used as a tag within a DA is that the tags used within a DA apply primarily to the function of an entire utterance. Disruption forms, however, usually apply only to the end of the utterance. For this reason, the use of periods with disruption forms is deemed necessary.

3.5 Quotes

Utterances that contain quoted material are to end with punctuation that reflects the DA of the utterance overall. If a quoted question is embedded within a statement, a period, rather than a question mark, is used at the end of the utterance in the transcript and no other punctuation is used.

A colon in the label signifies that there is quoted material in the transcription. The DA to the left of the colon characterizes the function of the entire utterance and the DA to the right of the colon characterizes only the quote. If the quoted material only consists of a

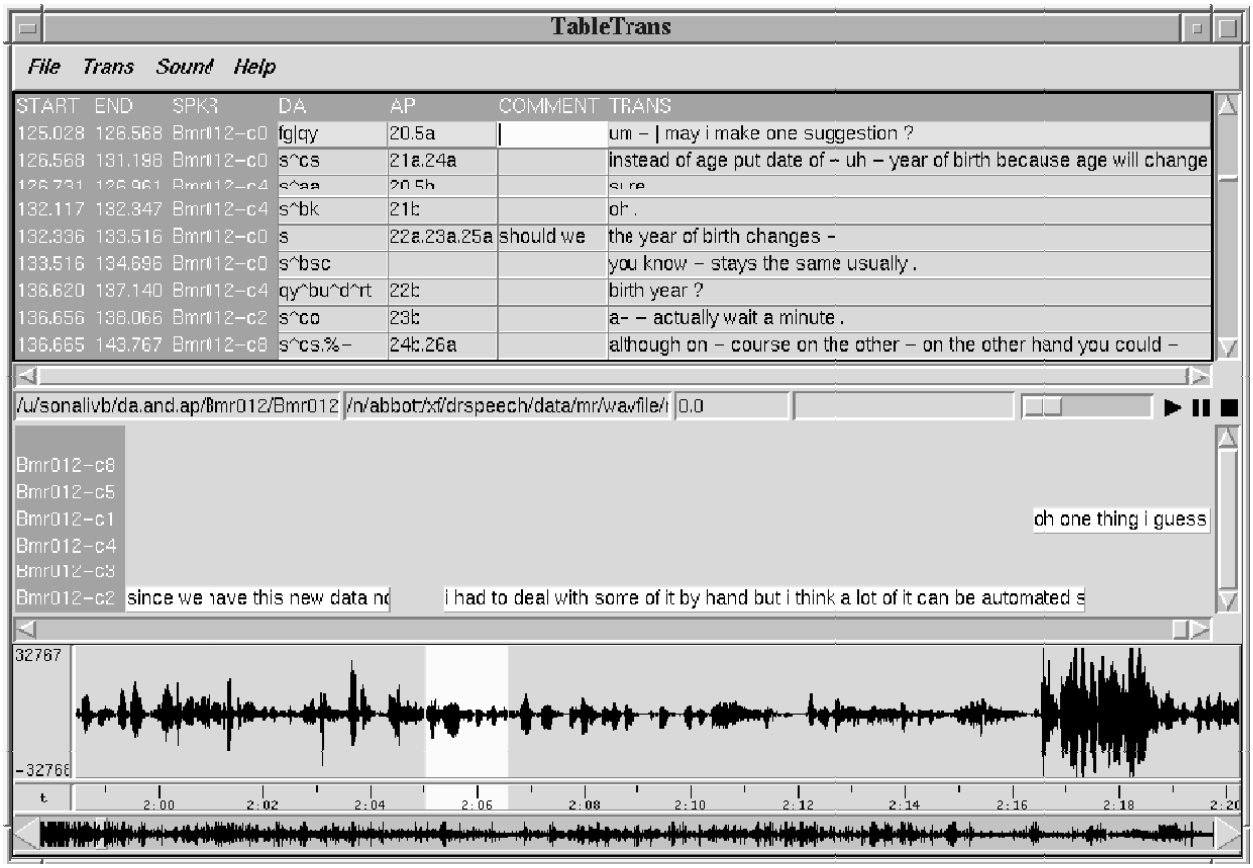
few words, such as a noun phrase, DA annotation of the quotation is unnecessary. Example 9 demonstrates the manner with which quotes are handled:

Example 9: Bmr026

941.984-944.924	c1	s^cs:qw	and just say an e- - just ask him that you know wha- - what should you do .
945.464-947.864	c1	s:qy	and in my answer back was are you sure you just want one .

3.6 Using TableTrans (Annotation Interface)

A. The Interface



There are three sections of TableTrans: the labeling and transcription section located at the top, the time-segmented transcription located in the middle, and the waveform located at the bottom.

In the labeling and transcription section, the first and second columns on the left provide the start and end times for each utterance and the third column denotes the speaker or channel number. DA and adjacency pair (AP) labels are entered in the fourth and fifth columns. The comment field is located in the sixth column and is primarily for an annotator's notes regarding an utterance. The last column on the right, under the "Trans" heading, provides the transcript of the utterances.

In order to label a meeting, the "Open Annotation File" command must be selected from the "File" menu. A sub-menu will appear providing three formats that can be used. "Table Format" is the format that is most widely used. A window will appear with a "Feature List" and a "Delimiter" to which clicking the "OK" button is necessary. Shortly after, the segment of the meeting to be annotated will appear.

Although the data within the fourth, fifth, sixth, and seventh columns may be altered within the interface, the Time-Segmented section, which is the first two columns and shows the annotator a series of utterances in chronological order, and the third column denoting the speaker cannot be modified.

B. TableTrans Commands

COMMAND	ACTION
Changing the Transcript	
Ctrl-s	Splits the current row at the location of the cursor in the TRANS field.
Ctrl-m	Merges the current row with the next row by the same speaker.
Moving within a Field	
Ctrl-f or left-arrow	Moves forward one character in a field.
Ctrl-b or right arrow	Moves backward one character in a field.
Ctrl-p or up-arrow	Moves up to previous row.
Ctrl-n or down-arrow	Moves down to next row.
Shift + left-arrow	Moves to previous field in the same row.
Shift + right-arrow	Moves to next field in the same row.
right-click	(In the Time-Segmented Transcription window) Opens up Comment Field Window
Ctrl-1	Plays a segment
Ctrl-a	Moves cursor to the beginning of a field
Ctrl-e	Moves cursor to the end of a field

C. Printing Commands

Annotators can print out their comments using the program "csvcomment." The command "csvcomment <csv_file>" is entered in the terminal window, where <csv_file> is the name of the ".csv" file to print.

D. Playing the Sound File

To open up the wave file of a meeting to be labeled, a link command can be made from the location where the sound file is saved in the annotator's home directory. After returning to the TableTrans interface, "Open Sound File" is selected from the "File" menu. The file can then be opened after browsing through the annotator's home directory.

SECTION 4: ADJACENCY PAIRS

4.1 Purpose and Definition

Labeling adjacency pairs (AP) in meetings provides a means to extract the information provided by the interaction between speakers. Adjacency pairs reflect the structure of conversation and are paired utterances such as question-answer, greeting-greeting, offer-acceptance, and apology-downplay. (Levinson 1983)

APs are defined as sequences of two utterances that are:

1. produced by different speakers
2. ordered with a first part (marked with “a”) and a second part (marked with “b”) (Levinson 1983)

An example of an AP is shown below:

Example 1: Bro016

113.976-116.502	c4	s^bu	but you were looking at mel cepstrum .
116.883-117.850	c5	s^aa	yes .

In Example 1, the utterances depict direct interaction between the two speakers.

4.2 Labeling Adjacency Pairs

Adjacency pairs consist of two parts, where each part is produced by a different speaker. The basic form of an AP is seen below:

<AP number><AP part>

This format allows APs to be enumerated as: 1a, 1b, 2a, 2b, and so on. A different number is assigned for each AP, yet every AP will contain an "a" part and a "b" part. A labeled AP is seen in Example 2:

Example 2: Bmr023

312.382-314.770	c2	qy^rt	30a	are you implying that it's currently disorganized ?
314.770-318.470	c3	s^na	30b	in my mind .

Although APs are to be marked sequentially in ascending order, it is possible that the numerical value of an AP jumps ahead of the numerical value of the previous AP by more than a value of one (e.g., an AP has a numerical value of 5 and the following AP has a numerical value of 7 instead of 6). However, such is only permitted so long as the sequential order of the APs is preserved and the numerical values are not repeated or used cyclically for entirely different APs.

4.3 Labeling Conventions

Specific labeling conventions have been established when marking APs in instances in which an utterance contains multiple AP parts, an AP part consists of multiple utterances, multiple speakers pertain to the same AP part, and an AP is overlooked.

A. Multiple AP Parts per Utterance

If an utterance functions as a "b" part of one AP and an "a" part of another AP, then both APs are marked with a period < . > separating the two APs, as seen below:

<AP number><AP part>**.**<AP number><AP part>

A portion of a conversation in which APs are labeled is seen in Example 3:

Example 3: Bro021

66.555-68.227	c2	s^rt	4a	well the first thing maybe is that the p- - eurospeech paper is uh accepted .
69.904-70.928	c2	fh		um ==
70.928-71.952	c2	fh		yeah .

72.059-74.710	c5	qw^rt	4b.5a	this is - what - what do you uh - what's in the paper there ?
74.702-81.090	c2	s^rt	5b.6a	so it's the paper that describe basically the um system that were proposed for the aurora .
80.320-82.794	c5	qy^bu^d^rt	6b.7a	the one that we s- - we submitted the last round ?
82.614-83.700	c2	s^aa	7b.8a	right yeah .
83.110-83.750	c5	s^bk	8b	uhhuh .

Refer to Section D for details regarding the treatment of utterances requiring three AP parts.

B. Continued AP Parts

A continued AP part is an AP part consisting of multiple utterances by the same speaker. When a continued AP part arises, a plus sign <+> is placed at the end of the AP. Example 5 depicts an instance where an AP part consists of multiple utterances:

Example 5: Bro016

1494.110-1499.560	c1	qy^rt	20a	do you have something simple in mind for - i mean vocal tract length normalization ?
1497.570-1501.320	c5	s^ar|s^nd	20b	uh no | i hadn't - i hadn't thought - it was - thought too much about it really .
1501.320-1503.200	c5	s^df^nd	20b+	it just - something that popped into my head just now .
1503.200-1505.070	c5	s.%--	20b++	and so i - i ==
1505.690-1509.900	c5	s^cs	20b+++	i mean you could maybe use the ideas - a similar idea to what they do in vocal tract length normalization .

Additionally, an utterance consisting of a tag question <g> is included within an AP part, assuming the utterance containing the statement <s> preceding it is a portion of the AP part. In which case, the utterance containing the tag question will receive the appropriate number of plus signs when labeled with an AP.

If an utterance contains multiple APs, where one or both is a continued AP part, a period < **.** > is inserted between the two APs to separate them (e.g., 5b++.6a+).

C. Multiple Speakers per AP Part

In some cases, an AP part consists of two or more speakers. This occurs most often with the "b" part and quite rarely with the "a" part. When such an occurrence arises, the corresponding AP number and AP part are marked. Then each speaker contributing to the same AP part receives a numerical value based upon the order in which the speakers make their utterances. So the first speaker to contribute to an AP part receives a value of 1, the second a 2, and so on. A hyphen <-> followed by a speaker's numerical value is then appended to the AP. The format of an AP consisting of multiple speakers is seen below:

<AP number><AP part> - <numerical value>

AP parts containing multiple speakers are seen in Example 5:

Example 5: Btr001

150.780-152.664	c5	s^bu	9a	parentheses meaning uncertainty .
151.730-152.365	c3	s^aa	9b-1	yeah .
152.467-153.164	c2	s^aa	9b-2	uhhuh .

If, for instance, the speaker designated as c2 in Example 5 continued speaking so that a continued AP part resulted, then his next utterance would be labeled as 9b-2+, the next 9b-2++, and so on as necessary. When continued AP parts occur within AP parts consisting of multiple speakers, each speaker retains his designated numerical value and plus signs <+> are appended after the numerical values as necessary.

Additionally, if an utterance contains multiple APs, where one or both is an AP part consisting of multiple speakers, a period < **.** > is inserted between the two APs to separate them (e.g., 5b-1.6a+, 1b-3+.2a).

D. Handling Overlooked APs

As stated in Section 4.2, APs are to be marked sequentially in ascending order. Occasionally, an AP is overlooked. If marking an overlooked AP with the next numerical value in sequence results in a non-sequential ordering of APs then an additional convention is implemented to handle the overlooked AP.

For instance, if a meeting is labeled with APs in sequence starting with a numerical value of 1 and ending with a value of 50 and an overlooked AP exists between an AP with a numerical value of 34 and an AP with a numerical value of 35, the overlooked AP is not to receive a numerical value of 51. Instead, the AP receives a numerical value of 34 followed by an underscore <_> and the appropriate AP part. The AP part is followed by a hyphen with a numerical value and plus signs when necessary. An overlooked AP located between two APs has the following format:

<AP number of previous AP>_<AP part>[- <numerical value>][$+_1$, $+_2$, …$+_n$]

If a number of overlooked APs exist in sequence, for instance if three APs exist between APs 34 and 35, then a slight modification of the above convention is necessary. The first overlooked AP receives an AP in the format detailed above. The second overlooked AP receives an AP in the same format but with two underscore <_> symbols instead of one. The third overlooked AP receives an AP in the same format but with three underscore symbols and so on, thus yielding the following format:

<AP number of previous AP>$_{\_1}$, $_{\_2}$, …$_{\_n}$ <AP part>[- <numerical value>][$+_1$, $+_2$, …$+_n$]

E. Labeled Meeting Sample

Example 6 depicts the labeling conventions discussed in Sections A through C. What is particularly unique about this example is that it contains an utterance requiring two “a” parts. Additionally, this utterance requires a total of three AP parts – two “a” parts and one “b” part – when utterances usually require at most two.

Example 6: Bmr003

1594.720-1595.830	c3	qy^d	47b.48a	you've already - you've already done some ?
1595.360-1596.610	c2	s.%-	48b-1	she - she's done one – she's one ==
1595.400-1595.950	c4	s^aa|s^na	48b-2	yes | i have .
1595.570-1597.070	c0	s^na	48b-3.49a.50a	she's - she's done about half a meeting .
1596.530-1597.570	c3	s^bk	49b-1	oh- - oh i see .
1597.130-1597.510	c2	s^bk	49b-2	right .
1597.570-1597.840	c3	s^bk	49b-1+	o_k .
1597.840-1598.100	c3	s^ba	49b-1++	good .
1597.760-1597.990	c2	s^bk	49b-2+	right .
1598.170-1598.360	c0	qy^d^g^rt	50a+	right ?

1598.580-1598.950	c2	s.%-		i'm go- ==
1598.580-1598.980	c0	qy^d^rt	50a++	about half ?
1599.150-1600.160	c4	s^no	50b.51a	s- - i'm not sure if it's that's much .

This utterance requires a “b” part as it contains the response to an earlier utterance, which constitutes the “a” part of the AP with a numerical value of 48. The “a” part of the AP with a numerical value of 49 only consists of one utterance and receives a number of responses. The utterance requires another “a” part for the AP with a numerical value of 50 as this utterance, along with the speaker’s following two utterances, comprise the “a” part for yet another AP.

F. Complex Form of an AP

The following is a complex form of an AP, taking into account the aforementioned conventions:

<AP number>[$\__1$, $\__2$, …$\__n$]<AP part>[-<numerical value>][$+_1$, $+_2$, …$+_n$][**.**<AP number> >[$\__1$, $\__2$, …$\__n$]<AP part> …]

4.4 Restrictions on Using Adjacency Pairs

Certain restrictions apply to which tags can or cannot be labeled with an AP.

APs denote direct interaction between speakers. Backchannels <b>, which serve simply to encourage the current speaker, are never marked with APs. Backchannels are not uttered directly to a speaker as a response and do not function in a way that elicits a response either. Rhetorical question backchannels <bh>, receive APs when uttered as acknowledgments and do not receive APs when uttered as backchannels.

Floor holders <fh> and floor grabbers <fg> are also never marked with APs, since they, like backchannels, are not said directly to anyone. Holds <h>, however, are marked. The definition of a hold entails that a speaker is given the floor and is expected to speak in response to something and "holds-off" prior to making an utterance. As the speaker is expected to speak and then utters a hold, which is usually followed by a response, the hold is considered part of the response.

Mimics <m> and collaborative completions <2> are always marked with APs, as they are always in direct reference to another speaker's utterance.

When indecipherable utterances appear, if the utterance can be characterized with a DA and it appears as though the utterance functions within an AP, then an AP is marked accordingly. Otherwise, no AP is marked.

In some cases, it is quite difficult to determine to which utterance a response refers. If such difficulty arises, then an AP is not marked. For instance, a scenario may arise where two or three speakers utter statements <s> simultaneously and another speaker utters an acknowledgment <bk>. As an acknowledgment by one speaker to another speaker is usually marked with an AP, if it cannot be determined whom a speaker is acknowledging, then an AP is not marked.

SECTION 5: TAG DESCRIPTIONS

5.1 Preliminaries

This section provides a detailed description of each tag and the rules governing the usage of each tag. The tags are categorized into thirteen groups according to syntactic, semantic, pragmatic, and functional similarities of the utterances they mark. Beneath a group heading will be a general description of the group along with explanations of the tags within the group. Most tag descriptions will contain examples[4] from data to further elucidate a tag's usage.

With regard to the examples provided within this section, it is of much use to listen to the corresponding audio portions, as some examples cannot be fully comprehended otherwise. In particular, utterances marked as floor grabbers <fg>, floor holders <fh>, holds <h>, backchannels <b>, acknowledgements <bk>, and accepts <aa> share a common vocabulary which renders examples of these tags in text insufficient in fully communicating how utterances marked as such are identified.

5.2 Group 1: Statements

This group contains only one tag, <s>, and serves as the default general tag.

▪ Statement <s>

The <s> tag is the most widely used tag in the MRDA tagset. Unless an utterance is completely indecipherable or else can be further described by a general tag as being a type of question, backchannel, floor grabber, floor holder, or hold, then its default status as a statement remains.

When necessary, specific tags are appended to the <s> tag to further characterize utterances. The use of the <s> tag is seen in Example 1 through Example 4:

4 In some examples, when displaying surrounding context, unnecessary lines, such as those which are irrelevant to characterizing a particular tag within the tag descriptions, may be edited out. The content of utterances within the examples remains unchanged.

Example 1: Bro004			
578.567-585.527	c3	s	if we exclude english um - there is not much difference with the data .
Example 2: Bed016			
70.600-71.470	c5	s^ba	it's a great story .
Example 3: Bro021			
3201.960-3204.850	c1	s^bu	so this changes the whole mapping for every utterance .
Example 4: Bro021			
3204.850-3205.490	c1	s^bk	okay .

5.3 Group 2: Questions

This group contains all general tags pertaining to questions. The tag description for elaborations <e> provides instructions regarding the treatment of questions followed by elaborations.

▪ Y/N Question <qy>

This tag marks utterances in the form of yes/no questions if and only if they have the pragmatic force along with the syntactic and prosodic indications of a yes/no question (i.e. subject-auxiliary inversion or question intonation). Essentially, an utterance is considered a yes-no question if it sounds as if it elicits a yes or no answer. This is not to say that all yes/no questions will receive yes or no answers. A question may be asked in a yes/no manner, but the response it receives may not be a simple yes or no. Regardless of the answer, the utterance is still considered a yes/no question.

Basic yes/no questions are seen in Example 5 through Example 8:

Example 5: Bro016			
58.863-61.782	c4	qy^rt	do you think that would be the case for next week also ?

Example 6: Bmr027			
2049.340-2051.730	c5	qy^rt	did i say that ?
Example 7: Bmr027			
1836.000-1838.580	c4	qy^bu^rt	didn't they want to do language modeling on you know recognition compatible transcripts ?
Example 8: Bmr012			
6.805-17.875	c1	qy^rt	is this channel one ?

The tag <qy> is also used as the general tag for tag questions <g> (e.g., "Yeah?", "Isn't it?", etc.) and rhetorical question backchannels <bh> (e.g., "Really?", "Isn't that interesting?", etc.). Many declarative questions <d> are also in the form of yes/no questions. Example 9 through Example 11 exhibit these characteristics:

Example 9: Bro016			
513.765-514.316	c4	qy^d^g^rt	right ?
Example 10: Bmr027			
2016.230-2017.440	c5	qy^bh	oh really ?
Example 11: Bmr027			
514.316-514.867	c4	qy^bu^d^rt	the insertion number is quite high ?

Additionally, a convention has been established in handling instances when a yes/no question is followed by an elaboration <e> which requires its own line. In such cases, the following elaboration could be considered a declarative yes/no question <qy^d>. Instead, the elaboration receives a DA of <s^e>, along with any other necessary specific tags. An instance of a yes/no question followed by an elaboration is seen in Example 12:

Example 12: Bro021			
316.709-319.202	c5	qy^rt	wasn't there some experiment you were going to try ?

319.202-325.216	c5	s^e.%--	where you did something differently for each um uh - i don't know whether it was each mel band or each uh um f f t bin or someth- ==

In some cases, it may be difficult to determine whether an utterance is a yes/no question or an "or" question <qr>. The tag description for <qr> details how distinguish between the two tags in certain scenarios.

▪ Wh-Question <qw>

Wh-questions are questions that require a specific answer. These usually contain "wh" words such as the following: what, which, where, when, who, why, or how. However, not all questions containing a "wh" word are considered wh-questions. The section on open-ended questions <qo> elucidates this point. Wh-questions are shown in Example 13 and Example 14:

Example 13: Bmr012

62.153-64.053	c3	qw^r^t3	why didn't you get the same results and the unadapted ?

Example 14: Bmr012

231.944-233.704	c2	qw^t3	i guess - what time do we have to leave ?

Declarative wh-questions often appear as wh-questions prior to wh-movement. An instance in which a declarative wh-question is used is seen in Example 15.

Example 15: Bed003

2889.130-2890.200	c1	qw	what's the technical term ?
2890.330-2890.750	c3	qw^d^rt	for which ?
2891.010-2892.820	c1	s^rt	for the uh - nodes that are observable .

In some cases, utterances that do not contain wh-words are labeled as wh-questions because they function as wh-questions. Such an instance is seen in Example 16:

Example 16: Bmr012

61.563-61.713	c0	qw^br^t3	hm ?

In Example 16, the utterance functions as a wh-question, in that "hm?" is akin to "what?" as a request for repetition. "Huh?", "excuse me?", and "pardon?" also appear as wh-questions in that they can also function in the same manner as what is exemplified in Example 16. Caution must be taken to distinguish whether such utterances are indeed wh-questions or if they are floor grabbers, floor holders, holds, backchannels, yes/no questions that are rhetorical question backchannels, or acknowledgments.

Declarative wh-questions that do not contain "wh" words are often confused with declarative forms of other questions because they appear the same syntactically. Despite this syntactic similarity, they differ functionally based upon the response that the question seeks. In determining whether an utterance is a declarative wh-question that does not contain a "wh" word, the surrounding context, in particular the response the question generates, is crucial to note. Most often, declarative wh-questions that do not contain "wh" words are requests for repetition, such as those seen in Example 17 through Example 19.

Example 17: Bmr031

947.610-948.925	c8	s	it's still yeah two or three d v ds .
948.925-950.240	c8	%-	but ==
949.569-951.874	c2	fg|s	yeah | not if you have to distribute the video also .
949.941-950.878	c5	qw^br^d	two or three ?
951.125-953.860	c8	s^df	if you use both sides and the two layer and all that .

Example 18: Bro003

3193.230-3198.820	c2	fh|s^cc	and um | for the broader class nets we're - we're going to increase that .
3198.820-3204.400	c2	s^df	because the um the digits nets only correspond to about twenty phonemes .
3205.460-3208.780	c2	fh	so .
3207.200-3207.840	c8	qw^br^d^rt	broader class ?
3208.780-3210.430	c2	h|s	um | the broader - broader training corpus nets .

Example 19: Bro003

3400.840-3402.950	c8	qw^br^d^rt	and - and you're saying about the

			spanish ?
3403.290-3404.350	c4	s	the spanish labels .
3405.000-3409.590	c4	s	that was in different format .

▪ Or Question <qr>

"Or" questions offer the listener at least two answers or options from which to choose. Section 2 and Section 3.3, which deal with segmentation and multiple DAs within an utterance, are quite helpful in determining if a question is actually an "or" question or if it is a yes/no question <qy> followed by an "or" clause after a yes/no question <qrr>. Select "or" questions can be seen in Example 20 through Example 23:

Example 20: Bmr001

305.466-307.826	c0	qr^rt	are we going to - i mean - is it going to be over there or is it going to be in there ?

Example 21: Bed003

1214.120-1215.140	c4	qr	are you assuming that or not ?

Example 22: Bmr001

339.042-342.612	c1	qr^rt	do we have like a cabinet on order or do we just need to do that ?

Example 23: Bmr007

165.987-167.447	cB	qr	is this the same as the e mail or different ?

In terms of the responses "or" questions receive, the obvious response is one in which a speaker selects one of the options posed within the "or" question. Sometimes the "or" question is interrupted and answered as if it is a yes/no question. In these cases, the question is marked as an "or" question if it seems as if the speaker would have continued the question in an "or" question format if he had not been interrupted. In other instances, the speaker asking the question might abandon his utterance, and the speaker answering the question may respond as if the question were a yes/no question without having interrupted the question at all.

If a speaker abandons a question that is seemingly an "or" question, it is actually a rather cumbersome task determining whether the question is indeed an "or" question or not. The point where the speaker abandons his question is of crucial importance. If the speaker abandons while posing at least a second option or after having posed at least two options, the question can be considered an "or" question. If the speaker abandons after saying the word "or" and has not issued a second option, the question could either be an abandoned "or" question or a yes/no question followed by an "or" clause, as mentioned above. If the speaker abandons at the word "or" abruptly, the utterance is most likely an "or" question. If the speaker trails off at the word "or" so that the word "or" is lengthened and sounds reminiscent of a floor holder <fh>, the "or" is segmented from the utterance or else separated by a pipe bar and is labeled as an abandoned "or" clause after a yes/no question <qrr.%--> and the remainder of the utterance is labeled as a yes/no question.

Example 24 through Example 31 depict instances of interrupted and abandoned "or" questions:

Example 24: Bed011			
2776.460-2779.490	c1	qr.%-	is that roughly the equivalent of - of what i've seen in english or is it ?==
Example 25: Bmr005			
2018.090-2023.710	c5	qr.%-	you know - did she miss some overlaps or did she ?==
Example 26: Bmr007			
369.570-372.515	cB	qr.%-	is this uh just raw counts or is it ==?
Example 27: Bmr013			
1987.000-1989.000	c2	qr.%--	well - oh wa- - in terms of the speakers or the conditions or the ?==
Example 28: Bmr013			
2064.000-2069.000	c1	qr^rt.%--	do the transcribers actually start wi- - with uh - transcribing new meetings or are they ?==
Example 29: Bmr014			
582.763-585.270	c8	qr.%--	has that started or is that ?==

Example 30: Bmr001

944.512-945.412	c8	qr^rt.%--	per channel or ?==

Example 31: Bmr009

1748.000-1751.000	c2	qr.%--	and north midland like like - uh illinois or ?==

If an utterance is suspected to be an "or" question but the speaker abandons or is interrupted before saying "or" and has not posed a second option, the utterance cannot be considered an "or" question since there is insufficient evidence to label it with the <qr> tag.

Furthermore, even with the presence of the word "or" along with a second option, it may be difficult to determine whether an utterance is an "or" question or a yes/no question, wh-question, or an open-ended question. If the question is actually presenting two specific options, the question is an "or" question. The question is not an "or" question if it presents one option and ends with a clause such as "or something." If a question ends with such a clause, the clause is not labeled separately with the tag <qrr>. Example 32 through Example 34 show instances when questions that are seemingly "or" questions are to be labeled as otherwise:

Example 32: Bmr005

3550.080-3551.680	c2	qy^d^rt^2	lapel mikes or something ?

Example 33: Bmr006

2057.610-2061.670	c0	qw	what if there was a door slam or something ?

Example 34: Bmr010

425.800-429.800	c6	qy	is there a - a transformation uh - like principal components transformation or something ?

▪ Or Clause After Y/N Question <qrr>

This tag marks when a speaker adds an "or" clause to a yes/no question. The previous description of "or" questions <qr> in conjunction with Section 2 and Section 3.3, which deal with segmentation and multiple DAs within an utterance, are also quite useful in determining whether a segment is an "or" clause and how to treat it.

As with the description of the tag <qr>, utterances marked with <qrr> must actually be posing some sort of option, rather than being a wh-question, for instance, preceded by the word "or."

Oftentimes, "or" clauses following yes/no questions are abandoned or else interrupted and the entire utterance consists of the word "or." In these cases, the label for such an utterance contains the <qrr> tag along with the appropriate disruption form.

Example 35 through Example 39 display in context instances where the tag <qrr> is used:

Example 35: Bed003

1867.670-1868.970	c1	qy^rt	do you have the true source files ?
1868.970-1870.270	c1	qrr	or just the class ?

Example 36: Bmr018

405.920-411.860	c1	qy^rt	the - i guess the question on my mind is do we wait for the transcribers to adjust the marks for the whole meeting before we give anything to i b m ?
411.860-413.440	c1	qrr	or do we go ahead and send them a sample ?

Example 37: Bmr001

2178.450-2179.950	c0	qy^d^rt	so - is it - it's going to disk ?
2179.950-2180.340	c0	qrr.%--	or is this ?==

Example 38: Bmr018

2722.490-2727.000	c1	qr	did they ever try going - going the other direction from simpler task to more complicated tasks ?
2727.000-2728.000	c1	qrr.%--	or ?==

Example 39: Bro004

1922.810-1928.020	c1	qy	so do you - are you - w- - did you have something going on - on the side with uh - or on - on this ?
1928.020-1928.130	c1	qrr.%-	or ?==

▪ Open-ended Question <qo>

An open-ended question places few syntactic or semantic constraints on the form of the answer it elicits. A question containing a "wh" word and consequently appearing to be a wh-question <qw> may actually be an open-ended question instead. Additionally, a question that is seemingly a yes/no question or an "or" question may actually be an open-ended question. As a wh-question, a yes/no question, and an "or" question require a specific answer, an open-ended question, as its name suggests, does not seek a specific answer at all. Rather, an open-ended question is asked in a broad sense.

Open-ended questions are seen in Example 40 through Example 48:

Example 40: Bmr007

112.365-116.868	c3	fh|qo^d^rt	um | and anything else ?

Example 41: Bmr007

117.088-118.018	c3	qo^d	nothing else ?

Example 42: Bmr007

92.862-98.798	c3	fh|qo^d^rt	um | and anything else anyone wants to talk about ?

Example 43: Bmr013

654.000-657.000	c3	qo^rt	d- e- - anybody do you have any - anybody have any opinion about that ?

Example 44: Bmr026

2307.190-2309.690	c5	qo	anybody have any intuitions or suggestions ?

Example 45: Bmr007			
1681.390-1683.180	c3	fg|qo	but - | what - what do you think about that ?
Example 46: Bmr014			
2691.750-2693.090	c4	qo^j	how about them energy crises ?
Example 47: Bmr007			
100.580-102.340	c0	qo^t	what about the um - your trip yesterday ?
Example 48: Bed006			
666.870-667.530	c2	qo^d	questions ?

▪ Rhetorical Question <qh>

The tag <qh> marks questions to which no answer is expected. Such questions are used by the speaker for rhetorical effect; they are essentially statements formulated as questions. Although rhetorical questions and rhetorical question backchannels <bh> are similar, <bh> lacks semantic content, functions mostly as a continuer, and is not used by a speaker who has the floor. Rhetorical questions are seen in Example 49 through Example 55:

Example 49: Bed011			
2204.540-2206.420	c2	qh^rt	i mean is this realistic ?
Example 50: Bmr005			
3802.380-3802.680	c4	qh^aa	why not ?
Example 51: Bmr005			
525.596-530.188	c4	qh^cs	so why don't you - you start with that ?
Example 52: Bmr009			
2089.900-2090.800	c3	qh	s- - i mean who cares ?

Example 53: Bmr009			
2512.610-2513.290	c1	qh^ba	isn't that wonderful ?
Example 54: Bmr009			
2778.960-2779.800	c0	qh^co	why don't you read the digits ?
Example 55: Bmr012			
1414.430-1415.430	c1	fh|qh	uh - | but who knows ?

5.4 Group 3: Floor Mechanisms

This group contains all general tags pertaining to mechanisms of grabbing or maintaining the floor. The only disruption forms that can be appended to tags within this section are the indecipherable tag <%> and the nonspeech tag <x>. Additionally, no specific tag may be appended to the tags denoted as floor mechanisms. Section 2 and Section 3.3 detail the issues regarding segmentation with floor mechanisms.

▪ Floor Grabber <fg>

Floor grabbers usually mark instances in which a speaker has not been speaking and wants to gain the floor so that he may commence speaking. They are often repeated by the speaker to gain attention and are used by speakers to interrupt the current speaker who has the floor. Most often, floor grabbers tend to occur at the beginning of a speaker's turn.

In some cases, none of the speakers will have the floor, resulting in multiple speakers vying for the floor and consequently using floor grabbers to attain it. During such occurrences, many speakers talk over one another without actually having the floor.

Floor grabbers are also used to mark instances in which a speaker who has the floor begins losing energy during his turn and then uses a floor grabber to either regain the attention of his audience or else because it seems as though he is relinquishing the floor, which he does not wish to do. Such mid-speech floor grabbers are usually followed by a change in topic.

Floor grabbers are generally louder than the surrounding speech. Although the energy of a floor grabber is relative to the energy of the surrounding speech, it is also relative to the energy of a speaker's normal speech.

Common floor grabbers include, but are not limited to, the following: "well," "and," "but," "so," "um," "uh," "I mean," "okay," and "yeah." It is worth mentioning that the identification of floor grabbers is not merely based purely on the vocabulary used, but rather on the speaker's actual attempt, whether successful or not, to gain the floor.

As previously mentioned, floor grabbers are not to be identified solely based upon the vocabulary used, as floor grabbers, floor holders <fh>, holds <h>, backchannels <b>, acknowledgements <bk>, and accepts <aa> share a very similar vocabulary. In order to properly distinguish whether an utterance is performing as a floor grabber, floor holder, hold, backchannel, acknowledgement, or accept, it is necessary to take into account the details provided within the individual tag descriptions and to listen to the audio portions corresponding to the examples within those tag descriptions. Utterances labeled with these tags tend to appear very similar in text yet emerge exceedingly different in sound.

As floor grabbers and backchannels are often confused on the basis of having a similar vocabulary, they are actually quite distinct in sound. The main distinctions between the two is that backchannels have a lower energy level in relation to the surrounding speech and are not used by someone who has or is attempting to gain the floor. Also, backchannels are considered "background" speech.

The floor grabbers seen in Example 56 through Example 60 are shown merely to illustrate how they appear in text. The surrounding context has been omitted for each example, as it provides little to no information regarding how to identify floor grabbers.

Example 56: Bed004

1017.990-1018.180	c4	fg	but uh ==

Example 57: Bed004

1052.310-1052.620	c2	fg	okay .

Example 58: Bed004

2264.780-2265.060	c2	fg	yeah but ==

Example 59: Bmr012

1814.65-1817.01	c2	fg|s.%-	well | or also for you know - if people are not ==

Example 60: Bmr012

1822.12-1824.17	c4	fg|qy^df	well i mean - | is the - is the handheld really any better ?

▪ Floor Holder <fh>

A floor holder occurs mid-speech by a speaker who has the floor. A floor holder is usually an utterance such as "uh" or "so" and is used as a means to pause and continue holding the floor. In some cases, a speaker will utter a floor holder at the end of his turn as a means to relinquish the floor.

The duration of a floor holder is usually longer than that of the other words spoken by a speaker. Also, the energy of a floor holder is often similar to that of the surrounding speech by the same speaker. Common floor holders include, but are not limited to, the following: "so," "and," "or," "um," "uh," "let's see," "well," "and what else," "anyway," "I mean," "okay," and "yeah."

In terms of placement, floor holders do not occur at the beginning of a speaker's turn, but rather occur throughout the middle and at the end[5] of a speaker's turn. Although floor holders do not occur at the beginning of a speaker's turn or speech, they may occur at the beginning of a speaker's utterance. If a speaker begins his turn with a floor grabber followed by a floor holder, it is permissible to label the suspected floor holder as such.

Section 2 discusses the treatment of floor holders in succession.

Floor holders are often found mid-utterance. In such cases, if an utterance is complete and splitting it to mark the floor holder would yield an incomplete utterance, the utterance remains intact and the floor holder is not marked.

In some cases, an utterance will end with a typical floor-holding word such as "um" or "uh" and, despite the presence of a common floor-holding word, a floor holder is not actually present, since the floor-holding word lacks the duration or "pause" property common to most floor holders. If such occurs, the utterance, while containing the floor-holding word, is simply marked as incomplete and the floor-holding word is not marked as an actual floor holder.

As previously mentioned, floor holders are not to be identified solely based upon the vocabulary used, as floor holders, floor grabbers <fg>, holds <h>, backchannels <b>, acknowledgements <bk>, and accepts <aa> share a very similar vocabulary. In order to properly distinguish whether an utterance is performing as a floor holder, floor grabber, hold, backchannel, acknowledgement, or accept, it is necessary to take into account the details provided within the individual tag descriptions and to listen to the audio portions corresponding to the examples within those tag descriptions. Utterances labeled with these tags tend to appear very similar in text yet emerge exceedingly different in sound.

5 As mentioned in Section 2, floor holders are not permitted to occur at the end of utterances. The treatment of floor holders within the transcript is discussed in Section 2 and Section 3.3.

Example 61 through Example 65 present floor holders in context:

Example 61: Bed003

2524.030-2526.510	c1	s	so it's a - it's a rather huge huge thing .
2526.510-2531.970	c1	fh|s.%--	but um - um - | we can sort of ==

Example 62: Bed003

2579.930-2581.760	c4	s	like all the different sort of general schemas that they might be following .
2581.760-2583.600	c4	fh	okay .

Example 63: Bed004

1336.010-1339.280	c2	s	i think we got plenty of stuff to talk about .
1340.180-1344.840	c2	fh|s	and then um - | just see how a discussion goes .

Example 64: Bed004

1596.700-1598.000	c2	s^arp	no i understand that .
1598.000-1599.540	c2	fh	but i- - but um ==

Example 65: Bed004

1672.310-1673.880	c2	fg	okay so so ==
1673.880-1675.440	c2	fh	uh ==

▪ Hold <h>

The <h> tag is used when a speaker who is given the floor and is expected to speak "holds off" prior to making an utterance. The <h> tag is predominantly used when a speaker is responding to a question that he in particular was asked, and that speaker pauses or "holds off" prior to answering the question.

Common holds include, but are not limited to, the following: "so," "um," "uh," "let's see," "well," "I mean," "okay," and "yeah."

Holds are very similar to floor holders <fh> in the way that they sound, however holds occur at the beginning of a speaker's turn, as opposed to floor holders which occur in the middle or at the end of a speaker's turn.

Although the primary distinction between holds and floor holders is location, holds are not collapsed with floor holders as they provide explicit information regarding a speaker's turn. Utterances marked as holds explicitly indicate that a speaker is given the floor, whereas utterances marked as holds indicate that a speaker merely has the floor.

If a speaker's initial utterance is marked as a hold and his following utterances appear to be either holds or floor holders, those following utterances are marked as holds. In other words, if a speaker's initial utterance is a hold and his following utterances are seemingly floor holders, those utterances appearing as floor holders are marked as holds until an utterance is encountered that is to be marked with a question tag or with the statement tag. After such a question or statement is encountered, any following segment within that same speaker's speech that appears to be a floor holder is marked as a floor holder and not as a hold.

As previously mentioned, holds are not to be identified solely based upon the vocabulary used, as holds, floor grabbers <fg>, floor holders <fh>, backchannels <b>, acknowledgements <bk>, and accepts <aa> share a very similar vocabulary. In order to properly distinguish whether an utterance is performing as a hold, floor grabber, floor holder, backchannel, acknowledgement, or accept, it is necessary to take into account the details provided within the individual tag descriptions and to listen to the audio portions corresponding to the examples within those tag descriptions.

Example 66[6] through Example 68 present instances of holds in context:

Example 66: Bro021

817.043-821.220	c1	qw	i mean what was the rest of the system ?
820.060-821.922	c2	h	um ==
823.605-827.084	c2	s	yeah it was - it was uh the same system .
828.960-829.683	c2	fh	uhhuh .
830.079-831.107	c2	s^r	it was the same system .
838.050-839.197	c2	fh	huh ==

Example 67: Bro021

3238.590-3243.580	c1	qy^d^rt	so you estimated uh f- - completely forgetting what you had before ?

6 In Example 66, the word "uhhuh" is used as a floor holder <fh>. Although the word "uhhuh" is not commonly used as a floor holder, this instance exemplifies the need to listen to corresponding audio portions in order to correctly assess the function of an utterance and not to label utterances according to the vocabulary used alone.

3244.200-3248.840	c4	h	um ==
3248.840-3251.170	c4	s^ar|s^nd	no no no | it's not completely noise .

Example 68: Bro018

1542.550-1546.120	c5	qy^rt	does there some kind of a distance metric that they use ?
1546.120-1549.520	c5	qw	or how do they for cla- - what do they do for classification ?
1550.050-1550.740	c0	h	um ==
1550.740-1551.150	c0	h	right .
1551.150-1559.900	c0	s	so the - the simple idea behind a support vector machine is um - you have - you have this feature space .

5.5 Group 4: Backchannels and Acknowledgments

This group contains the general tag for backchannels <b> and the specific tags for acknowledgments <bk>, assessments/appreciations <ba>, and rhetorical question backchannels <bh>. The commonality among the tags of this group is that they are most often used to mark utterances that are often responses, in the form of acknowledgments or backchannels, to a speaker who has the floor as that speaker is talking. Such responses generally do not elicit feedback. Also, utterances marked with these tags generally do not serve the purpose of halting the speaker who has the floor.

It may seem as though the tags <bk> and <ba> could be grouped with the tags in Group 5, since they are responses of a sort, they are instead placed in Group 4 due to the nature of the utterances they mark. The tags in Group 5 are limited to being orthogonally categorized as positive, negative, or uncertain. Utterances marked with <bk> are perceived as being neutral, whereas utterances marked with <ba> can be either positive or negative. Thus the tag <ba> is not included within Group 5 as its dynamic nature would prevent the preservation of the orthogonal categorization scheme within Group 5. Additionally, utterances marked with the tag <ba> generally tend to have more in common with utterances marked with the tag <bk> than with the tags in Group 5. These similarities are discussed in the tag description for <ba>.

▪ Backchannel <b>

Utterances which function as backchannels are not made by the speaker who has the floor. Instead, backchannels are utterances made in the background that simply indicate that a listener is following along or at least is yielding the illusion that he is paying attention. When uttering backchannels, a speaker is not speaking directly to anyone in particular or even to anyone at all.

Common backchannels include the following: "uhhuh," "okay," "right," "oh," "yes," "yeah," "oh yeah," "uh yeah," "huh," "sure," and "hm."

The nature of backchannels does not usually permit utterances such as "uh," "um," and "well" as being perceived as backchannels, since these utterances do not indicate that a speaker is following along, but rather that a speaker has something to say or else is attempting to say something.

As previously mentioned, backchannels are not to be identified solely based upon the vocabulary used, as backchannels, floor grabbers <fg>, floor holders <fh>, holds <h>, acknowledgements <bk>, and accepts <aa> share a very similar vocabulary. In order to properly distinguish whether an utterance is performing as a floor grabber, floor holder, hold, backchannel, acknowledgement, or accept, it is necessary to take into account the details provided within the individual tag descriptions and to listen to the audio portions corresponding to the examples within those tag descriptions. Utterances labeled with these tags tend to appear very similar in text yet emerge exceedingly different in sound.

Furthermore, backchannels are most often confused with acknowledgments and accepts than with floor grabbers, floor holders, and holds. One method in distinguishing if the <b>, <bk> or <aa> tag is appropriate lies in the point at which the utterance occurs with regard to the speaker who has the floor's utterance. Acknowledgments generally appear after another speaker has completed a phrase or an utterance, as they are acknowledging the semantic significance of what is said. Accepts usually occur at the end of another speaker's utterances, as they are agreeing with what is said. Backchannels, although they can occur in the same locations as acknowledgments and accepts, can also be found in the middle of another speaker's phrase. Such mid-phrasal placement is a strong indicator that an utterance is a backchannel, rather than an acknowledgment or an accept, as the speaker uttering the backchannel lacks adequate semantic information from the other speaker's utterance to acknowledge it or agree to it. Additionally, backchannels are usually uttered with a significantly lower energy level than the surrounding speech, while acknowledgments tend not to be quite so low as backchannels and accepts are generally at the same level or else higher.

Additionally, the only specific tag that may be appended to a backchannel is the rising tone tag <rt>.

Backchannels in context are seen in Example 69 through Example 71:

Example 69: Bro018

1821.160-1829.060	c2	s	but i think that uh - this was a couple years ago .
1821.510-1821.820	c5	b	huh .

Example 70: Bro018

2005.020-2012.090	c5	qy^rt	do you get out a - uh - a vector of these ones and zeros and then try to find the closest matching phoneme to that vector ?
2006.210-2006.410	c0	b	uhhuh .

Example 71: Bro007

837.018-838.648	c1	s^df	well also just to know the numbers .
837.345-837.565	c3	b	yeah .
838.648-838.828	c1	b	right .

▪ Acknowledgment <bk>

The <bk> tag is used to express a speaker's acknowledgment of a previous speaker's utterance or of a semantically significant portion of a previous speaker's utterance. Acknowledgments are neither positive nor negative, as they only serve to acknowledge, not to agree or disagree. In some cases, a speaker will acknowledge his own utterance or a semantically significant portion of his own utterance.

Common acknowledgments, in addition to mimicked portions, include, but are not limited to, the following: "I see," "okay," "oh," "oh okay," "yeah," "yes," "uhhuh," "huh," "ah," "all right," and "got it." If an utterance is suspected to be an acknowledgment solely based upon the vocabulary used, yet does not sound as though it is an acknowledgment, then it should not be marked as one.

As opposed to backchannels, acknowledgments encode a level of direct communication between speakers. A speaker who acknowledges a previous speaker's utterance is actually speaking directly to that previous speaker, yet is usually not seeking a response from the previous speaker. As stated in the tag description for backchannels, the tags <bk>, <b>, and <aa> are often confused with one another. The tag description for backchannels elucidates how to distinguish among the three tags.

Acknowledgements also tend to be confused with floor grabbers <fg>, floor holders <fh>, and holds <h> due to their similar vocabularies. In order to properly distinguish the function of an utterance, it is necessary to take into account the details provided

within the individual tag descriptions and to listen to the audio portions corresponding to the examples within those tag descriptions. Utterances labeled with the <bk>, <fg>, <fh>, and <h> tags, as well as with the <b> and <aa> tags, tend to appear very similar in text yet emerge exceedingly different in sound.

Restrictions apply to the usage of the <bk> tag with other specific tags. The <bk> tag is only used when the primary function of an utterance is to acknowledge a portion of another speaker's speech. The use of other tags to mark an utterance, such as those in Group 5, indicates that an utterance serves a different primary purpose, such as agreeing or disagreeing. So, when a tag from Group 5 is used to mark an utterance, the <bk> tag may not be used in conjunction with that tag.

The <bk> tag also may not be used with <ba>, as the <ba> tag encodes the acknowledging nature of <bk> within its definition and thus renders the <bk> tag redundant when the two are used in conjunction. The use of the <ba> tag also indicates that an utterance is either positive or negative, whereas an utterance marked with the <bk> tag is neutral. The <bk> tag may not be used with <bh>, as <bh> is a type of backchannel or acknowledgment, depending upon its usage, and may encode the acknowledging nature of <bk> thus rendering the use of the <bk> tag redundant when used in conjunction.

The specific tags with which <bk> is permitted to be used in conjunction are <m>, <r>, <rt>, <fe>, <t1> and <t3>. When used in conjunction with the <bk> tag, a tag from this list merely indicates a feature of the acknowledgment. In the case of the tag <fe>, when used in conjunction with the tag <bk>, it indicates that an exclamatory acknowledgment was uttered. When used with another functional tag, such as <aa> or <cs>, the tag <fe> indicates that an exclamatory agreement or an exclamatory suggestion has been made.

Acknowledgments in context are seen in Example 72 through Example 76:

Example 72: Bmr012

58.784-60.504	c3	qw^t3	so why didn't you get the same results and the unadapted ?
62.153-64.053	c3	qw^r^t3	why didn't you get the same results as the unadapted ?
64.235-68.995	c0	s^t3	oh because when it estimates the transformer pro- - produces like single matrix or something .
67.730-69.010	c3	s^bk^t3	o- - oh i see .

Example 73: Bed003

151.920-155.150	c1	s	it opens the assistant that tells you that

			the font type is too small .
155.780-156.120	c2	s^bk	ah .

Example 74: Bed003

158.220-159.100	c2	s^nd	i'd prefer not to .
159.140-159.500	c1	s^bk	okay .

Example 75: Bed003

166.460-169.010	c2	s^rt	because i'm going to switch to the javabayes program .
167.820-168.400	c1	s^bk	oh okay .

Example 76: Bed003

1615.540-1617.810	c2	s^rt	so we can rel- open it up again .
1616.130-1616.410	c3	s^bk	okay .

▪ Assessment/Appreciation <ba>

Assessments/appreciations are acknowledgments directed at another speaker's utterances and function to express slightly more emotional involvement than what is seen in the utterances marked with the <bk> tag. The <ba> tag is similar to the <bk> tag in that it acknowledges another speaker's utterance, however it lacks the neutral nature of the <bk> tag. Utterances marked with <ba> can be either positive or negative. When negative, utterances marked with the <ba> tag are often criticisms.

Utterances which function as acknowledgments in the senses discussed under the tag descriptions for <bk>, <bh>, and <ba> may only be marked with one of these tags to express the acknowledging nature of an utterance, not a combination of these tags.

As with the <bk> tag, the <ba> tag encodes a level of direct communication between speakers. When appreciating or assessing the contents of a previous speaker's utterance, a speaker is actually speaking directly to the previous speaker, yet usually is not seeking a response from the previous speaker.

Although most utterances marked with the <ba> tag tend to be quite short, some utterances tend to be somewhat lengthy. This is due to the very nature of the <ba> tag. In briefly expressing appreciation or assessing a situation, which is usually the case, a speaker's utterance may be something to the likes of "that's great," "that's terrible," "good enough," "wow," or "excellent." Brief utterances such as these are often uttered as exclamations, thus requiring the <fe> tag.

Longer appreciations tend to be akin to utterances such as "so I think that's a really great way to approach it." Longer assessments tend to appear as criticisms, which take many forms. Comments and opinions on an aspect a speaker has noticed within the contents of another speaker's speech are often marked as assessments/appreciations also.

In some cases, utterances which are assessments/appreciations are also affirmative answers <na>, dispreferred answers <nd>, or negative answers <ng>. In these cases, an utterance that is assessing or appreciating is also communicating that it is agreeing or disagreeing. An utterance such as "I think that would be worth doing" would function as an assessment/appreciation in that it embeds the speaker's own opinion. Assuming the utterance is actually agreeing to another speaker's previous utterance, the utterance also functions as an affirmative answer in that it accepts and agrees to what the previous speaker said. An utterance such as "that's wonderful" is an assessment/appreciation, yet is not an agreement since it only expresses an assessment.

In determining whether an utterance is indeed an assessment/appreciation, it is necessary to ensure that the assessment/appreciation is actually uttered in reference to another speaker's utterance.

A variety of assessments/appreciations are seen in Example 77 through Example 89:

Example 77: Bed006

172.462-173.242	c3	s^ba	it's very exciting .

Example 78: Bed006

257.526-257.916	c3	s^ba	that's good .

Example 79: Bed006

266.653-267.043	c2	s^ba	wonderful .

Example 80: Bed006

347.295-347.615	cA	s^ba	it's fine .

Example 77: Bed006

172.462-173.242	c3	s^ba	it's very exciting .

Example 78: Bed006

257.526-257.916	c3	s^ba	that's good .

Example 79: Bed006

266.653-267.043	c2	s^ba	wonderful .

Example 80: Bed006

347.295-347.615	cA	s^ba	it's fine .

Example 81: Bmr021

261.000-262.000	c4	s^ba^fe	wow !

Example 82: Bed006

1333.750-1337.640	c2	s^ba	but it's - so this time we - we are at an advantage .

Example 83: Bed008

1873.870-1876.850	c2	fg|s^ba	uh - | anyway this is crude .

Example 84: Bed008

2035.000-2036.000	c2	s^ba	but this is a good discussion .

Example 85: Bed008

3878.640-3880.450	c4	s^ba	so this is slightly uh - more complicated .

Example 86: Bed008

4997.490-5002.340	c0	s^ba	that's uh - that's a whole lot of constructions .

Example 87: Bed017

1462.890-1467.820	c2	s^ba	so it's probably not that easy to simply have a symbolic uh computational model .

Example 88: Bmr002

1992.220-1996.800	c2	s^ba	and i was very impressed by how well you could hear separate speakers .

Example 89: Bmr021

747.750-749.530	c0	fg|s^ba^cs	well | it seems like just shortening them is a good short term solution .

▪ Rhetorical Question Backchannel <bh>

Rhetorical question backchannels lack semantic content and are syntactically similar to rhetorical questions, however they function as backchannels and acknowledgments. Rhetorical question backchannels can be uttered as backchannels, which is often the case, in that they can be made in the background and simply indicate that a listener is following along or at least is yielding the illusion that he is paying attention. In these cases, the use of a rhetorical question backchannel indicates that a speaker is not speaking directly to anyone in particular or even to anyone at all. When uttered as an acknowledgment, the rhetorical question backchannel expresses a speaker's acknowledgment of a previous speaker's utterance or of a semantically significant portion of a previous speaker's utterance. As acknowledgments, rhetorical question backchannels encode a level of direct communication between speakers. A speaker who acknowledges a previous speaker's utterance is actually speaking directly to that previous speaker, yet is usually not seeking a response from the previous speaker. However, when acknowledgments are uttered as rhetorical question backchannels, they often receive answers such as "yeah." Additionally, when a rhetorical question backchannel functions as an acknowledgment, it is unnecessary to mark the <bk> tag.

As stated in the tag descriptions for <bk> and <ba>, the default tag for acknowledgments is the <bk> tag. If further descriptions apply to an acknowledgment and a <ba> or <bh> tag is deemed necessary, than only one of these tags is used. The <bk> tag cannot be used in conjunction with the <ba> or <bh> tags.

Common rhetorical question backchannels include, but are not limited to, the following: "oh really?", "yeah?", "isn't that interesting?", and "you think so?".

Rhetorical question backchannels always receive the Y/N question general tag <qy>.

Example 90 through Example 99 present instances of rhetorical question backchannels:

Example 90: Bed003

2136.810-2137.060	c1	qy^bh	yeah ?

Example 91: Bed003

2319.660-2319.910	c2	qy^bh	really ?

Example 92: Bed003

3493.590-3494.000	c3	qy^bh	oh really ?

Example 93: Bmr005

1358.460-1358.690	c3	qy^bh^rt	yeah ?

Example 94: Bmr012

671.580-672.090	c4	qy^bh^d^rt	oh it did ?

Example 95: Bmr014

522.800-523.120	c8	qy^bh^m^rt	no ?

Example 96: Bmr014

2357.840-2358.290	c8	qy^bh	oh they won't ?

Example 97: Bmr021

193.000-194.000	c5	qy^bh	isn't that something ?

Example 98: Bmr021

859.540-860.670	c5	qy^bh	is that right ?

Example 99: Bro021

170.110-170.542	c5	qy^bh	huh ?

5.6 Group 5: Responses

Group 5 is orthogonally divided into three subgroups: positive utterances, negative utterances, and uncertain utterances. The tags in Group 5 are often used to characterize responses to questions and suggestions.

POSITIVE

▪ Accept <aa>

The <aa> tag is used for utterances which exhibit agreement to or acceptance of a previous speaker's question, proposal, or statement. Utterances marked with the <aa> tag are quite short, as their lengthy counterparts are marked with the <na> tag.

Common utterances marked with the <aa> tag include, but are not limited to, the following: "yeah," "yes," "okay," "sure," "uhhuh," "right," "I agree," "exactly," "definitely," and "that's true."

Additionally, the word "no" can be marked with the <aa> tag if it is used to agree to a syntactically negative statement or question, as seen in Example 104.

Utterances marked with the <aa> tag may be confused with backchannels and acknowledgments. Generally, utterances marked with the <aa> tag have much more energy and are more assertive than backchannels and acknowledgments. The tag descriptions for backchannels and acknowledgments further elucidate the distinctions among the three tags.

Accepts are not to be identified solely based upon the vocabulary used, as accepts, floor grabbers <fg>, floor holders <fh>, holds <h>, backchannels <b>, and acknowledgements <bk> share a very similar vocabulary. In order to properly distinguish whether an utterance is performing as an accept, floor grabber, floor holder, hold, backchannel, or acknowledgement, it is necessary to take into account the details provided within the individual tag descriptions and listen to the audio portions corresponding to the examples within those tag descriptions. Utterances labeled with these tags tend to appear very similar in text yet emerge exceedingly different in sound.

Accepts in context are seen in Example 100 through Example 104:

Example 100: Bro017

2264.620-2271.560	c3	s.x	if you want to decrease the importance of a c- - parameter you have to

			increase it's variance .
2267.450-2267.830	c1	s^aa	yes .
2269.590-2269.840	c1	s^aa.x	right .
2269.690-2269.980	c4	s.x	multiply .
2270.470-2270.690	c1	s^aa	yes .
2271.610-2272.050	c1	s^aa	exactly .

Example 101: Bro022

1575.820-1579.190	c0	s^df	because when you train up the aurora system you're uh - you're also training on all the data .
1579.190-1582.560	c0	s.%--	i mean it's ==
1580.350-1580.920	c2	s^aa	that's right .
1580.920-1581.490	c2	s^aa	yeah .

Example 102: Bro022

1475.950-1477.970	c4	s	and it was about six point six percent .
1477.390-1477.780	c2	s^bk	oh .
1477.790-1478.630	c1	s^aa	right right right right .
1478.630-1479.470	c1	s^bk	okay .

Example 103: Bro026

2416.730-2418.050	c2	s	because that's what you're going to be using .
2418.050-2418.210	c2	qy^d^g^rt	right ?
2418.250-2418.740	c3	s^aa	yeah .
2418.740-2419.220	c3	s^aa^r	yeah .

Example 104: Bro026

854.850-858.060	c2	s^nd	although you - you know you haven't tested it actually on the german and danish .
858.060-858.360	c2	qy^d^g^rt	have you ?
858.850-859.520	c0	s^aa	no we didn't .

▪ Partial Accept <aap>

The <aap> tag marks when a speaker explicitly accepts part of a previous speaker's utterance. Partial accepts are often conditional responses that accept or agree to another speaker's utterance.

Partial accepts are often confused with partial rejections <arp>. The distinction is that an utterance marked with the <aap> tag focuses on agreeing with or accepting part of a previous speaker's utterance. An utterance marked with the <arp> tag focuses on disagreeing with or rejecting part of a previous speaker's utterance.

Partial accepts in context are seen in Example 105 through Example 108:

Example 105: Bed003

922.295-924.105	c1	s^bu^rt	well the - the - sort of the landmark is - is sort of the object .
924.105-925.915	c1	qy^d^g	right ?
925.915-927.595	c1	qy^d^g^rt	the argument in a sense ?
927.230-928.260	c4	s^aap	usually .

Example 106: Bmr024

1147.330-1156.120	c3	fh|qy^bu^d	um so | it's wizard in the sen- - usual sense that the person who is asking the questions doesn't know that it's uh a machi- - not a machine ?
1155.600-1156.190	c5	s^aap	at the beginning .

Example 107: Bmr006

944.455-949.460	c3	s	but i think that - i'm raising that because i think it's relevant exactly for this idea up there that if you think about well gee we have this really complicated setup to do well maybe you don't .
950.300-961.150	c3	s^cs	maybe if - if - if really all you want is to have a - a - a recording that's good enough to get a - uh a transcription from later you just need to grab a tape recorder and go up and make a recording .
950.660-951.260	c1	s^aap	for some of it .

Example 108: Bro007

1605.290-1612.800	c2	s^cs	and - and perhaps i was thinking also a fourth one with just - just a single k l t .
1612.800-1616.550	c2	s^df	because we did not really test that .
1616.550-1620.300	c2	s^cs	removing all these k l t's and putting one single k l t at the end .
1622.760-1626.240	c1	s^na	yeah i mean that would be pretty low maintenance to try it .
1626.970-1628.480	c1	fh|s^aap	uh - | if you can fit it in .

▪ Affirmative Answer <na>

The <na> tag marks an utterances that act as narrative affirmative responses to questions, proposals, and statements. The <na> tag is much like the <aa> tag in that they both exhibit agreement to or acceptance of a previous speaker's question, proposal, or statement. The difference between the two tags is that, as the <aa> tag is used for shorter utterances, the <na> tag is used for lengthy utterances.

In order to determine whether an utterance requires the <na> tag, the surrounding context is generally required. Without surrounding context, an utterance requiring the <na> tag may be considered merely as a statement <s> without any additional specific tags representing agreement or acceptance.

Instances of the <na> tag in context are seen in Example 109 through Example 111:

Example 109: Bed011

1528.600-1530.280	c2	s	nobody's interested in that except for the speech people .
1529.120-1529.290	c3	s^aa	right .
1529.290-1530.300	c3	s^na	no we don't care about that at all .

Example 110: Bmr001

374.134-377.954	c8	s	a cabinet is probably going to cost a hundred dollars two hundred dollars something like that .
378.105-381.715	c0	s^na	yeah i mean - you know - we - we can spend under a thousand dollars or something without - without worrying about it .

Example 111: Bmr007

1656.590-1664.310	cA	s	if - if the goal were to just look at overlap you would - you could serve yourself - save yourself a lot of time but not even transcri- transcribe the words .
1666.090-1668.990	c1	s	well i was thinking you should be able to do this from the acoustics on the close talking mikes .
1668.990-1671.900	c1	qy^d^g	right ?
1671.140-1674.800	cB	s^na	well that's - the - that was my - my status report .

NEGATIVE

▪ Reject <ar>

The <ar> tag marks negative words such as "no" and other semantic equivalents that offer negative responses to questions, proposals, and statements. The <ar> tag marks brief negative responses to questions, proposals, and statements in the same manner that the <aa> tag marks brief affirmative answers.

Common utterances marked with the <ar> tag include, but are not limited to, the following: "no," "nope," "no way," "nah," "not really," and "I don't think so."

When syntactically negative questions or statements arise, responses in the form of "yes," "yeah," or the like can function as rejections. As discussed in the tag description for <aa>, negative responses such as "no" can function as agreements in these cases.

Rejections in context are seen in Example 112 through Example 116:

Example 112: Bed003

259.160-264.920	c4	qy.%-	but are you saying that in this particular domain it happens the - that landmarkiness cor- - is correlated with ?==
263.409-264.019	c3	s^ar	no .

Example 113: Bed003

545.980-548.160	c4	qy	and are those mutually exclusive sets ?
547.610-547.990	c3	s^ar	not at all .

Example 114: Bed003

1758.350-1760.280	c2	qy^rt	i didn't n- - is there an ampersand in dos ?
1761.030-1761.370	c3	s^ar	nope .

Example 115: Bed003

3022.070-3023.720	c2	qy^rt	do you want to trade ?
3023.360-3024.610	c1	h|s^ar	um - | no .

Example 116: Bed011

2776.460-2779.490	c1	qr.%-	is that roughly the equivalent of - of what i've seen in english or is it ?==
2779.390-2780.180	c2	s^ar	no not at all .

▪ Partial Reject <arp>

The <arp> tag marks when a speaker explicitly rejects part of a previous speaker's utterance. Partial rejections are often responses posing exceptions when rejecting another speaker's utterance.

Partial rejections are often confused with partial accepts <aap>. As stated in the tag description for <aap>, the distinction between the two is that an utterance marked with the <aap> tag focuses on agreeing with or accepting part of a previous speaker's utterance. An utterance marked with the <arp> tag focuses on disagreeing with or rejecting part of a previous speaker's utterance. An utterance marked with the <aap> tag is formulated in a positive manner, whereas an utterance marked with the <arp> tag is formulated in a negative manner.

Partial rejections in context are seen in Example 117 through Example 119[7]:

7 The tag <sj> is seen in Example 19. This tag was formerly part of the MRDA tagset eliminated in the revision of the tagset. Appendix 4 details tags which are no longer a part of the MRDA tagset.

Example 117: Bed003

1352.970-1355.790	c2	qy^bu^rt	also - you know - didn't we have a size as one ?
1357.120-1357.350	c3	qw^br	what ?
1357.330-1358.250	c2	s^r^rt	the size of the landmark .
1359.860-1361.550	c3	s^arp	um - not when we were doing this .

Example 118: Bed003

1131.440-1132.880	c2	s	it would actually slow that down tremendously .
1136.540-1137.290	c3	s^arp	not that much though .

Example 119: Bmr018

505.460-507.485	c4	s	but you're listening to the mixed signal and you're tightening the boundaries .
507.485-509.510	c4	s^bsc	correcting the boundaries .
509.510-512.510	c4	s	you shouldn't have to tighten them too much because thilo's program does that .
511.313-512.073	c0	sj.x	should be pretty good .
512.550-515.710	c3	s^arp	except for it doesn't do well on short things remember .

▪ Dispreferred Answer <nd>

The <nd> tag marks statements which act explicit narrative forms of negative answers to previous speakers' questions, proposals, and statements in the same manner in which the <na> tag acts as an agreement with or acceptance of a previous speaker's utterance. As with the <na> tag, the <nd> tag marks lengthier utterances than those marked with the <ar> tag which exhibit rejection.

Surrounding context is generally required to determine whether an utterance requires the <nd> tag. Without surrounding context, an utterance requiring the <nd> tag may be considered merely as a statement <s> without any additional specific tags representing rejection.

Dispreferred answers are often confused with negative answers <ng>. The main distinction between the two tags is that the <nd> tag marks utterances that offer explicit rejections and the <ng> tag marks utterances that offer implicit rejections through the use of hedging.

Dispreferred answers in context are seen in Example 120 through Example 124:

Example 120: Bmr001

948.121-951.731	c8	s^bu^rt	we figured out that it was t- - twelve gig- - twelve gigabytes an hour .
949.056-949.806	c1	s^nd	it was more than that .

Example 121: Bed003

156.910-157.510	c1	qy^rt	do you want to try ?
158.220-159.100	c2	s^nd	i'd prefer not to .

Example 122: Bed003

1163.060-1166.150	c4	s	so i thought that was directly given by the context switch .
1163.130-1166.160	c3	s^nd	that's a different thing .

Example 123: Bmr005

781.990-783.000	c4	s	probably de- - probably depends on what the prepared writing was .
785.281-786.821	c1	s^bk|s^nd	yeah | i don't think i would make that leap .

Example 124: Bmr024

1987.890-1989.760	c1	s^bs	he's saying get a whole different drive .
1989.680-1990.810	c5	s^nd	but there's no reason to do that .

▪ Negative Answer <ng>

As opposed to a dispreferred answer <nd> which explicitly offers a negative response to a previous speaker's question, proposal, or statement, a negative answer <ng> implicitly offers a negative response with the use of hedging.

The negative answer tag <ng> is often confused with the maybe tag <am> and the no knowledge tag <no>. The maybe tag <am> marks utterances in which a speaker asserts that his response is probable, yet not definite, and the no knowledge tag <no> marks utterances in which a speaker does not know an answer. A negative answer <ng> essentially offers an indirect negative response. In uttering an indirect negative response, a speaker may employ responses similar to those marked with the maybe tag

<am> and no knowledge tag <no> to hedge around uttering a direct refusal or negative response.

Oftentimes, negative answers <ng> appear as alternative suggestions to a previous speaker's question, proposal, or statement.

Negative answers <ng> in context are seen in Example 125 through Example 133[8]:

Example 125: Bed004

350.465-352.450	c4	qy^rt	y- - you guys have plans for sunday ?
352.900-353.470	c4	s.%--	we're - we're not ==
353.470-360.645	c4	s	it's probably going to be this sunday but um w- - we're sort of working with the weather here .
360.645-367.820	c4	s^df	because we also want to combine it with some barbecue activity where we just fire it up and what - whoever brings whatever you know can throw it on there .
368.787-371.447	c4	s	so only the tiramisu is free nothing else .
373.980-377.050	c1	s^ng	well i'm going back to visit my parents this weekend .

Example 126: Bmr005

4094.420-4099.430	c2	qw	what if we give people you know - we cater a lunch in exchange for them having their meeting here or something ?
4099.640-4103.350	c1	s^ng	well you know - i - i do think eating while you're doing a meeting is going to be increasing the noise .

Example 127: Bmr007

14.467-15.967	cB	qy^rt	and uh shall i go ahead and do some digits ?
16.724-17.504	c3	h|s^ng	uh | we were going to do that at the end .

8 Regarding the use of the tag <sj> in Example 133, refer to footnote 7.

Example 128: Bmr007

1750.790-1755.290	cA	s	we have - have in the past and i think continue - will continue to have a fair number of uh phone conference calls .
1756.380-1771.950	cA	fh|s^cs	and uh | and as a - to um as another c- c- comparison condition we could um see what - what what happens in terms of overlap when you don't have visual contact .
1774.140-1777.190	cB	s^ng	it just seems like that's a very different thing than what we're doing .

Example 129: Bmr007

1773.730-1774.870	c1	qy^rt	can we actually record ?
1775.870-1778.340	c3	fh|s^ng	uh | well we'll have to set up for it .

Example 130: Bmr014

2637.240-2645.800	cB	s	i mean so it's like i- - in a way it's - it's nice to have the responsibility still on them to listen to the tape and - and hear the transcript .
2645.800-2646.660	cB	s.%--	to have that be the ==
2647.970-2652.800	c8	s^ng	i mean most people will not want to take the time to do that though .

Example 131: Bmr024

1237.760-1240.380	c9	s^cs	maybe we can have him vary the microphones too .
1241.190-1243.470	c5	fg|s	so - so - so | for their usage they don't need anything .
1243.880-1246.890	c4	s^ng	but - but i'm not sure about the legal aspect of - of that .

Example 132: Bmr024

2385.660-2389.950	cB	s.%--	it might be that one more iteration would - would help but it's sort of ==
2390.330-2390.650	cB	fh	you know .
2390.440-2392.350	c3	s^ng	or maybe - or maybe you're doing one too many .

Example 133: Bmr024

818.269-825.296	c5	s	sure there - there might be a place where it's beep seven beep eight beep eight beep .
826.056-829.156	c5	s	but you know they - they're - they're going to macros for inserting the beep marks .
830.078-831.768	c5	sj	and so i - i don't think it'll be a problem .
831.768-832.708	c5	s^cs	we'll have to see .
832.708-833.648	c5	sj^r	but i don't think it's going to be a problem .
834.643-834.903	c3	s^bk	okay .
835.101-836.021	c3	fg|s^ng	well | i - i - i don't know .
836.021-848.194	c3	s^cs	i - i think that that's - if they are in fact going to transcribe these things uh certainly any process that we'd have to correct them or whatever is - needs to be much less elaborate for digits than for other stuff .

UNCERTAIN

- ## Maybe <am>

The maybe tag <am> marks utterances in which a speaker's utterance conveys probability or possibility by using the word "maybe" or other words denoting possibility and probability. An utterance marked with the <am> tag is one which the speaker asserts that his utterance is probable or possible, yet not definite.

The <am> tag is often confused with suggestions <cs> which have the form of "maybe we should..."

Maybes <am> in context are seen in Example 134 through Example 138:

Example 134: Bed003

1228.410-1231.250	c1	qw^rt	we- - what set the - they set the context to unknown ?
1232.500-1233.580	c3	s	right now we haven't observed it .

1233.580-1236.710	c3	s^am	so i guess it's sort of averaging over all those three possibilities .

Example 135: Bed003

2969.930-2971.610	c3	qy^rt	is srini going to be at the meeting tomorrow ?
2971.610-2971.870	c3	qy^rt	do you know ?
2972.580-2972.910	c4	s^am	maybe .

Example 136: Bed003

3206.200-3214.190	c1	s.%--	but you know - if we take a subject that is completely unfamiliar with the task or any of the set up we get a more realistic ==
3212.060-3213.000	c3	s^am	i guess that would be reasonable .

Example 137: Bmr009

1752.000-1754.000	c0	qw	so - so what accent are we speaking ?
1756.500-1761.000	c3	s^am	probably western yeah .

Example 138: Bmr018

1890.390-1893.760	c0	s^df	because you have to uh - maneuver around on the - on both windows then .
1895.010-1895.960	c4	qr^d	to add or to delete ?
1896.110-1896.480	c0	s	to delete .
1898.510-1898.860	c4	s^bk^rt	okay .
1898.970-1900.440	c3	fg|%-	anyways | so i - i guess ==
1900.380-1904.150	c4	s^am	that - maybe that's an interface issue that might be addressable .

▪ No Knowledge <no>

The no knowledge tag <no> marks utterances in which a speaker expresses a lack of knowledge regarding some subject.

The most common expressions found within utterances marked with the no knowledge tag are "I don't know" and "I'm not sure." However, in some cases, utterances consisting of "I don't know" are actually floor holders <fh> and are not to be marked with the no knowledge tag.

Utterances marked with the no knowledge tag may be confused with utterances marked with the negative answer tag <ng>. The tag description for the <ng> tag elucidates this issue.

Instances of utterances labeled with the no knowledge tag, where some are shown in context, are seen in Example 139 through Example 146:

Example 139: Bed003

142.790-146.410	c1	s	but if you really want to find out what it's about you have to click on the little light bulb .
147.130-148.810	c2	s^no	although i've - i've never - i don't know what the light bulb is for .

Example 140: Bed003

1281.990-1284.650	c3	s^no	but uh - i don't know y- what the right thing is to do for that .

Example 141: Bed004

1417.360-1418.320	c2	s^no	yeah i don't understand it .

Example 142: Bmr001

68.756-70.816	c0	fg|s^no	um - | i have no idea which one i'm - i'm on .

Example 143: Bmr001

354.108-359.588	c1	qy	do we have any money at all that we can go out and spend on things like cabinets or a hard drive or things like that ?
359.791-360.451	c0	h|s^no	oh - i mean - | i don't know .

Example 144: Bmr001

366.306-368.646	c0	h|qw^rt	uh | how much are we talking about here ?
371.211-374.134	c8	h|s^no	um - | i don't know .

Example 145: Bmr001

1365.460-1366.620	c0	qy	didn't we already get that ?

1365.650-1366.140	c8	s^no.%	oh god knows .

Example 146: Bed003

2112.730-2113.480	c0	qw	who was it trained on ?
2113.770-2114.510	cB	h|s^no	uh | i have no idea .
2114.740-2115.330	cB	s^no	i don't remember .

5.7 Group 6: Action Motivators

This group contains specific tags pertaining to future action. Whether the future action occurs immediately or after a long period of time is not relevant.

The tags in Group 6 either indicate that a command or a suggestion has been made regarding some action to be taken at some point in the future or else indicate that a speaker has committed himself to executing some action at some point in the future.

▪ Command <co>

The <co> tag marks commands. In terms of syntax, a command may arise in the form of a question (e.g., "Do you want to go ahead?") or as a statement (e.g., "Give me the microphone.").

Commands are often confused with suggestions <cs>. The distinction between the two entails considering what sort of response such an utterance could receive as well as the role of the speaker within the meeting. In terms of responses, commands are uttered as orders, where a failure to comply (e.g., a "no" answer), in an extreme sense, is perceived as a sign of indignation toward the speaker uttering the command. With regard to a suggestion, rejecting a suggestion is not considered as impolite as rejecting a command. If an utterance yields the illusion that it may be a command or a suggestion, considering whether the utterance could receive a response that is a rejection and whether that rejection is considered impolite is a helpful method to determine if the utterance is a command or a suggestion. If a rejection is considered impolite, the utterance is considered a command, otherwise it is considered a suggestion.

In terms of the role of a speaker within a meeting, generally suggestions made by the speaker running a meeting are perceived as commands. If the speaker running the meeting says to another speaker, "let's try that one," such an utterance is considered a command. Whereas, if the same utterance is made by another speaker who is not running the meeting, then the utterance is considered a suggestion instead. However,

this is not to say that all suggestions made by the speaker running a meeting are to be considered as commands. In distinguishing between commands and suggestions made by a speaker running a meeting, it is helpful to consider the method regarding whether a rejection is impolite as discussed in the previous paragraph.

Commands are seen in Example 147 through Example 162. Note that commands that appear to be suggestions within these examples are actually commands made by the speaker running the meeting.

Example 147: Bed003			
160.020-160.440	c1	s^co	continue .
Example 148: Bed003			
177.840-178.190	c4	s^co	proceed .
Example 149: Bed003			
581.856-582.226	c3	s^co	wait .
Example 150: Bed003			
1440.550-1441.820	c1	s^co	let's get this uh - b- - clearer .
Example 151: Bed003			
1467.230-1473.090	c2	s^co	explain to me why it's necessary to distinguish between whether something has a door and is not public .
Example 152: Bed003			
1670.450-1675.190	c1	s^co	close it and - and load up the old state so it doesn't screw - screw that up .
Example 152: Bed003			
1761.440-1762.790	c3	s^co	just s- - l- - start up a new d o s .
Example 153: Bmr001			
127.000-127.450	c1	s^co	fill it out .

Example 154: Bmr001

131.458-131.988	c8	s^co	just write it down .

Example 155: Bmr001

2016.020-2017.270	c0	s^co	well - let's do some more while we got them here .

Example 156: Bmr005

4248.000-4250.020	c8	fh|s^co	so | we should think about trying to wrap up here .

Example 157: Bmr007

3080.090-3082.130	c3	qw^co	so why don't you explain it quickly ?

Example 158: Bro026

236.320-247.993	c2	s^co^t	but i guess maybe the thing - since you weren't - yo- - you guys weren't at that - that meeting might be just - just to um - sort of recap uh - the - the conclusions of the meeting .

Example 159: Bro026

311.870-317.825	c2	fh|s^co^t	uh - | maybe describe roughly what - what we are keeping constant for now .

Example 160: Bro026

2068.470-2071.780	c2	s^co	yeah so maybe just c c hari and say that you've just been asked to handle the large vocabulary part here .

Example 161: Bro021

2611.590-2618.090	c1	s^bk|s^co	okay | so now once you get that - that one then you - then you do a first- - or second order or something taylor series expansion of this .

Example 162: Bro026

614.735-617.130	c2	s^co^t	and then uh - maybe you should just continue telling what - what else is in the - the form we have .

▪ Suggestion <cs>

The suggestion tag marks proposals, offers, advice, and, most obviously, suggestions.

Suggestions are often found in constructions such as "maybe we should..." Suggestions containing the word "maybe" are not to be confused with the maybe tag <am>. Additionally, if the phrase "excuse me" precedes something for which a speaker is negotiating permission (Jurafsky 35), then it is marked as a suggestion rather than an apology <fa>.

Suggestions are also often confused with commands <co>. The tag description for <co> clarifies how such might occur.

Suggestions are seen in Example 163 through Example 173:

Example 163: Bro018

948.67-950.165	c5	fg|s^cs	yeah | i was just going to say maybe it has something to do with hardware .

Example 164: Bro021

28.107-28.938	c5	qy^cs^rt	should we take turns ?

Example 165: Bro021

28.938-29.768	c5	qy^cs^d^rt	you want me to run it today ?

Example 166: Bro021

33.052-36.270	c5	s^cs	let's see maybe we should just get a list of items .

Example 167: Bro021

414.758-419.812	c1	s^cs	i- - i really would like to suggest

			looking um a little bit at the kinds of errors .
Example 168: Bro021			
1967.920-1969.610	c2	s^cs	maybe you have to standardize this thing also .
Example 169: Bro021			
1987.380-2000.980	c1	qw^cs	um given that we're going to have for this test at least of - uh boundaries what if initially we start off by using known sections of nonspeech for the estimation ?
Example 170: Bro021			
2054.740-2058.370	c4	s^cs	if you want you c- - i can say something about the method .
Example 171: Bro021			
2340.390-2341.720	c1	s^cs	maybe we can take it off line .
Example 172: Bro021			
2564.920-2566.410	c1	s^cs	i think these things are a lot clearer when you can use fonts - different fonts there .
Example 173: Bro021			
711.142-715.021	c1	s^cs	and maybe you'd want to have something that was a little more adaptive .

▪ Commitment <cc>

The commitment tag <cc> is used to mark utterances in which a speaker explicitly commits himself to some future course of action. Commitments are not to be confused with suggestions in which a speaker suggests that he, the speaker himself, execute some action. With commitments, a speaker mentions what he will do in the future, not what he might do.

Commitments are seen in Example 174 through Example 181:

Example 174: Bmr018			
278.930-281.910	c0	s^cc	i'll - i'll - i'll um - get - make that available .
Example 175: Bmr018			
526.910-527.560	c4	s^cc^j	i'll work on that .
Example 176: Bmr024			
1972.600-1974.890	c5	s^cc	my intention is to do a script that'll do everything .
Example 177: Bmr026			
196.510-198.560	c5	s^cc	i'll send it out to the list telling people to look at it .
Example 178: Bmr026			
202.562-203.282	c0	s^cc	i'll try to get to that .
Example 179: Bmr026			
211.838-212.668	c0	s^cc	i'm just going to do it .
Example 180: Bmr026			
218.868-227.628	c0	s^cc	i'm going to send out to the participants uh - with links to web pages which contain the transcripts and allow them to suggest edits .
Example 181: Bmr026			
271.030-271.440	c5	s^cc	i'll wait .

5.8 Group 7: Checks

This group contains specific tags pertaining to understanding or being understood.

▪ "Follow Me" <f>

The <f> tag marks utterances made by a speaker who wants to verify that what he is saying is being understood. Utterances marked with the <f> tag explicitly communicate or else implicitly communicate the questions "do you follow me?" or "do you understand?" In implicitly communicating those questions, a speaker's utterance may be a tag question <g>, such as "right?" or "okay?", where a sense of "do you understand?" is being conveyed.

Tag questions marked with the "follow me" <f> tag often occur in instances in which a speaker is attempting to be instructional or else is offering an explanation. After an instruction or explanation, a speaker may utter a tag question <g> that is also a "follow me" in order to gauge whether what he is saying is understood.

Instances of the "follow me" tag, some of which are shown with their surrounding context, are seen in Example 182 through Example 187:

Example 182: Bed008

589.304-590.304	c5	qy^d^f^rt	this is understandable ?

Example 183: Bmr006

23970.340-3971.190	c1	qy^f^rt	do you know what i'm saying ?

Example 184: Bmr007

2821.400-2823.070	c3	qy^d^f^rt	you know what i mean ?

Example 185: Bmr008

670.000-676.000	c4	qy^d^f	well - i guess i was thinking maybe you know how you were taking information off of the digits and putting it onto that ?

Example 186: Bro021

1267.930-1268.770	c0	s.%--	i - i - i was thinking ==
1268.770-1272.600	c0	s^bk|s	okay | so just set to - set to some really

			low number the - the nonvoiced um phones .
1272.600-1274.520	c0	qy^d^f^g^rt	right ?
1274.520-1276.440	c0	s	and then renormalize .

Example 187: Bro016

264.902-267.287	c4	s	i mean y- - don't want to do this over a hundred different things that they've tried .
267.287-268.822	c4	s	but you know for some version that you say is a good one .
268.822-270.356	c4	qy^d^f^g	you know ?
273.619-279.864	c4	qw	how - how much uh does it improve if you actually adjust that ?
284.961-288.832	c4	s	but it is interesting .

▪ Repetition Request

An utterance marked as a repetition request indicates that a speaker wishes for another speaker to repeat all or part of his previous utterance. Repetition requests are usually used when a speaker could not decipher another speaker's previous utterance and wishes to hear that portion again.

Common repetition requests include, but are not limited to, the following: "what?", "sorry?", "huh?", "pardon?", "excuse me?", and "say that again." The tag description for wh-questions <qw> proves to be quite useful in determining the general tag for some repetition requests.

Instances of repetition requests, some of which are shown with their surrounding context, are seen in Example 188 through Example 195:

Example 188: Bed003

1291.740-1300.550	c1	fh|qw^rt	um | how long would it take to - to add another node on the observatory and um - play around with it ?
1301.430-1302.290	c3	qw^br^rt	another node on what ?

Example 189: Bed003

1352.970-1355.790	c2	qy^bu^rt	also - you know - didn't we have a size as one ?

1357.120-1357.350	c3	qw^br	what ?

Example 190: Bed003

3146.860-3148.940	c3	qw	so who would be the subject of this trial run ?
3149.670-3149.910	c1	qw^br	pardon me ?

Example 191: Bmr018

2495.240-2495.770	c0	qw^br	what did you say ?

Example 192: Bro015

365.840-366.470	c3	qw^br	what was that again ?

Example 193: Bmr008

3114.260-3116.010	c8	qw	what about doing it with just the single channels ?
3117.010-3117.270	c2	qw^br^rt	sorry ?

Example 194: Bmr005

2687.890-2688.970	c2	qw^rt	how many meetings is that ?
2689.200-2689.640	c8	qw^br^rt	what's that ?

Example 195: Bmr030

243.000-244.000	c1	qw	how much memory does he have ?
244.000-245.000	c0	qy^br^d^rt	i'm sorry ?

▪ Understanding Check <bu>

The understanding check tag marks when a speaker checks to see if he understands what a previous speaker said or else to see if he understands some sort of information.

With understanding checks, a speaker usually states what he is trying verify as correct and follows that with a tag question <g>. Only the utterance, or portion of the utterance if a pipe bar is used, containing the information to be verified is marked with the <bu> tag. Tag questions <g> are not marked with the <bu> tag as they do not contain the information that is to be verified.

Understanding checks are often confused with repetition requests <br> and summaries <bs>. With a repetition request, a speaker is seeking to hear what another speaker said again, whereas, with an understanding check, a speaker is seeking to verify if what he is saying is indeed correct. With a summary, a speaker summarizes something that was previously said and is not seeking any sort of verification of correctness.

Understanding checks in context are seen in Example 196 through Example 199:

Example 196: Bed003

1907.630-1909.300	c2	s	there's a bayes net spec for - in x m l .
1909.400-1910.680	c3	qy^bu^rt	he's - like this guy has ?
1910.780-1911.550	c3	qy^bu^d^g^rt	the javabayes guy ?

Example 197: Bed011

1988.840-1994.600	c2	s	i e uh - it's either uh - for sightseeing for meeting people for running errands or doing business .
2006.120-2010.250	c1	qy^bu^d	so business is supposed to uh - be sort of - it - like professional type stuff ?
2010.250-2012.320	c1	qy^d^g	right ?

Example 198: Bed011

1504.790-1525.140	c2	s	the reading task is a lot shorter .
1511.580-1516.010	c3	s.%--	and other than that yeah i guess we'll just have to uh - listen ==
1516.010-1520.440	c3	s^bu	although i guess it's only ten minutes each .
1520.440-1520.670	c3	qy^d^g^rt	right ?

Example 199: Bmr012

231.944-233.704	c2	qw^t3	i guess - what time do we have to leave ?
234.144-234.774	c2	qy^bu^d^rt^t3	three thirty ?

5.9 Group 8: Restated Information

This group, as the name states, contains specific tags pertaining to information that has been restated. The group is further divided into two subgroups: repetition and correction.

REPETITION

▪ Repeat <r>

The repeat tag <r> is used when a speaker repeats himself. This often occurs in response to repetition requests <br> or else to place emphasis on a certain point.

In repeating himself, a speaker repeats all or part of one of his previous utterances. However, in order for an utterance to be considered a repeat, it must be a repeat of an utterance made at most a few seconds prior to the repeat. Also, the guidelines regarding segmentation, as discussed in Section 2, are to be taken into consideration so that utterances in which a speaker begins speaking and then starts over using the same words are within the same utterance are not segmented and the pipe bar is not employed so that the repeated portions are labeled as repeats.

It is not required that a speaker repeat himself verbatim in order for a utterance to be marked with the repeat tag <r>. If a speaker repeats himself and the repeated utterance differs by a small number of words yet approximates the original utterance, the <r> tag may be used. However, the <r> tag is not to be used if a speaker alters an utterance so much so that no obvious structural likeness can be seen. For instance, if a speaker says, "my pen has run out of ink" and then says "my pen's run out," the second statement can be considered a repeat of the first. However, if the speaker's second utterance was instead "there's no ink in my pen," that utterance would not be considered a repeat of the first.

Additionally, in repeating himself, a speaker's utterance marked as a repeat may contain more speech in addition to what was repeated. For instance, if a speaker says, "I have to leave at one," and then follows that utterance with "I have to go at one and make some phone calls," the latter utterance is still considered a repeat despite the additional information.

Repeats <r> are not to be confused with mimics <m>. As previously stated, a repeat occurs when a speaker repeats his own utterance. A mimic occurs when a speaker repeats another speaker's utterance. Repeats are also not to be confused with summaries <bs> where a speaker summarizes his own utterances as many structural differences occur between the summary and the information being summarized.

Repeats in context are seen in Example 200 through Example 202:

Example 200: Bro017

1821.640-1822.990	c1	s	and hev- - everything is fixed .
1822.990-1823.950	c1	s^r	everything is fixed .

Example 201: Bro017

1827.470-1828.860	c1	s	for both - you would have to do .
1829.110-1829.720	c5	s^bu^m	you would do it on both .
1829.560-1829.720	c1	s^aa	yeah .
1829.720-1830.390	c5	s.%-	so you'd actually ==
1829.830-1830.870	c1	s^r	you have to do bo- - both .

Example 202: Bro025

870.243-872.737	c1	qy^bu^d^rt	and there didn't seem to be any uh penalty for that ?
873.030-873.386	c2	qy^br^rt	pardon ?
873.390-876.620	c1	qy^bu^d^r^rt	there didn't seem to be any penalty for making it causal ?

▪ Mimic <m>

The mimic tag marks when a speaker mimics another speaker's utterance, or portion of another speaker's utterance.

As with repeats <r>, mimics do not have to be repeated verbatim in order to be considered mimics. This condition is discussed in the tag description for repeats <r>.

Also, if a speaker's utterance is marked as a mimic, it may contain more speech in addition to what is mimicked. For instance, if one speaker says, "there's a problem with the phone system," and then another speaker follows that utterance with "there's a problem with the phone system concerning what aspect?," the latter utterance would still be considered a mimic despite the additional speech.

Mimics are often forms of acknowledgments <bk> and, when such is the case, are labeled in conjunction with the <bk> tag. The most common scenario when a mimic is a form of acknowledgment occurs as a speaker who has the floor is talking and another speaker acknowledges the speaker who has the floor by mimicking part of what he says.

In other cases, a speaker will mimic another speaker and phrase the mimic in the form of a declarative question as a request for more information about what they mimicked. For instance, if a speaker's utterance is "I went to the restaurant" and another speaker's utterance in response is "the restaurant?", the response is a mimic of the first utterance and acts as a request for more information about the restaurant.

Mimics <m> are not to be confused with repeats <r>. As previously stated, A mimic occurs when a speaker repeats another speaker's utterance. A repeat occurs when a speaker repeats his own utterance.

Also, mimics are not to be confused with summaries <bs> where a speaker summarizes another speaker's utterances as many structural differences occur between the summary and the information being summarized.

Mimics in context are seen in Example 203 through Example 211:

Example 203: Bed003

1875.040-1875.550	c3	s^co^rt	go up one .
1875.700-1876.410	c2	s^bk^m	up one .

Example 204: Bed004

1567.700-1568.320	c4	qw	what's tourbook ?
1569.180-1570.630	c1	s^m.%--	tourbook ==

Example 205: Bmr001

1700.790-1704.110	c8	s	so - so they - they're going to - they're going to have to make speaker assignments or something like this .
1704.030-1705.880	c1	s^bk^m	they're going to have to make speaker assignments .

Example 206: Bmr001

878.126-878.426	c8	s^bc	nine .
878.352-878.672	c1	s^bk^m	nine .

Example 207: Bmr001

1043.710-1044.080	c8	s	it's a pain .
1044.500-1044.810	c1	s^bk^m	it's a pain .

Example 208: Bmr005

1492.390-1495.610	c3	s	i - i - i - i consider - i consider acoustic events uh - the silent too .
1497.240-1497.860	c1	s^m	silent .

Example 209: Bmr005

2785.520-2786.340	c8	s^na	it's what we're aiming for .
2786.060-2786.970	c2	s^bk^m	that we're aiming for .

Example 210: Bmr009

1963.930-1966.420	c3	s	well you have a like techno speak accent i think .
1965.700-1967.180	c0	qy^bu^d^m^rt	a techno speak accent ?

Example 211: Bmr012

123.504-124.024	c3	s^cs	california .
124.251-124.871	c4	s^bk^m	california .

▪ Summary <bs>

The <bs> tag marks when a speaker summarizes a previous utterance or discussion, regardless of whose speech he is summarizing.

Summaries are not to be confused with understanding checks <bs>. Understanding checks restate information for validation while summaries do not require validation. Furthermore, a DA may not contain both the <bs> and <bu> tags.
Summaries are also not to be confused with repeats <r> and mimics <m>. The tag descriptions for repeats and mimics detail how such might occur.

Summaries in context are seen in Example 212 and Example 213:

Example 212: Bro011

75.120-82.956	c3	fh|s^rt	well - uh | first we discussed about some of the points that i was addressing in the mail i sent last week .
87.253-90.293	c3	s^rt	about the um - well - the downsampling problem .

91.763-94.322	c3	s	uh - and about the fit- - uh the length of the filters .
98.530-100.610	c1	qw^rt	so what's the - w- - what was the downsampling problem again ?
98.609-98.929	c3	%-	so we had ==
100.610-101.180	c1	s	i forget .
100.813-105.273	c3	s	so the fact that there - there is no uh - low pass filtering before the downsampling .
107.394-113.682	c3	s	there is because there is l d a filtering but that's perhaps not uh - the best .
114.640-117.470	c1	s|s^aa	depends what it's frequency characteristic is | yeah .
117.680-119.610	c1	s^cs	so you could do a - you could do a stricter one .
118.240-118.580	c4	qy^rt^t3	is the system on ?
120.255-120.545	c1	s^am	maybe .
122.143-125.083	c3	s.%--	so we discussed about this about the um ==
125.550-126.740	c1	qy^rt	was there any conclusion about that ?
128.482-129.032	c3	h|s^co^na^rt	uh - | try it .
130.300-130.640	c1	s^bk	i see .
135.230-140.890	c1	s^bs	so again this is th- - this is the downsampling uh - of the uh - the feature vector stream .

Example 213: Bro017

539.307-543.396	c1	s	so i mean uh - uh - add moderate amount of noise to all data .
544.447-549.417	c1	s	so that makes uh - th- - any additive noise less addi- - less a- - a- - effective .
549.417-549.737	c1	qy^d^g^rt	right ?
549.550-549.870	c5	s^aa	right .
549.957-552.487	c1	s.%--	because you already uh - had the noise uh - in a ==
552.487-555.017	c1	s	and it was working at the time .
555.017-557.032	c1	s.%--	it was kind of like one of these things you know but ==
559.870-566.410	c1	s	so well you know just take a - take a spectrum and - and - and add of the constant c to every - every value .
560.570-561.820	c5	s.%-	well you're - you're basically y- ==
567.550-569.560	c5	s^bs	so you're making all your training data more uniform .

CORRECTION

▪ Correct Misspeaking <bc>

The <bc> tag is used when a speaker corrects another speaker's utterance. Corrections are based upon whether the word choice of a speaker is corrected or the pronunciation of a word is corrected.

Instances in which the correct misspeaking tag <bc> are used are shown in context in Example 214 through Example 217:

Example 214: Bro012

1221.540-1225.420	c5	s^ar|s^rt	oh no | i've ninety four .
1218.660-1219.640	c1	s^bc	ninety three point six four .

Example 215: Bed012

2122.730-2124.280	c2	s^j^2	killing machines !
2125.890-2126.880	c1	s^bc	reasoning machines .

Example 216: Bmr011

3098.000-3100.000	c6	s	native speaking native speaking english .
3100.000-3102.000	c7	s^bc	i bet he meant native speaking american .

Example 217: Bmr011

1308.000-1309.000	c1	s^rt	and there we're already using fourteen .
1309.000-1311.000	c7	s^bc	and we actually only have fifteen .

▪ Self-Correct Misspeaking <bsc>

The <bsc> tag marks when a speaker corrects his own error, with regard to either pronunciation or word choice.

Segmentation is an issue regarding the <bsc> tag. As with repeats, a speaker may begin an utterance and correct himself within the same utterance. In such cases, the utterance is not segmented and the pipe bar is not employed to mark the <bsc> tag. Section 2 details the guidelines surrounding how and why utterances are segmented.

Instances in which the self-correct misspeaking tag <bsc> are used are shown in context in Example 218 through Example 223:

Example 218: Bed003

567.066-574.026	c3	s^bk|s	okay | so - yeah so note the four nodes down there the - sort of the things that are not directly extracted .
574.316-575.176	c3	s^bsc	actually the five things .

Example 219: Bed003

1013.070-1013.210	c3	s^aa	yeah .
1013.260-1013.420	c3	s^ar^bsc	no .

Example 220: Bmr009

301.025-303.500	c2	fh|s	um and uh | they don't look very separate .
303.750-305.600	c2	fh|s^bsc	uh | separated .

Example 221: Bmr013

1632.080-1632.920	c8	s^rt.%--	well we did the hand ==
1632.920-1633.760	c8	s^bsc	the one by hand .

Example 222: Bmr024

653.072-659.242	c5	h|s.%--	uh so | we have a whole bunch of digits that we've read and we have the forms and so on um but only a small number of that ha- ==
659.384-660.524	c5	s^bsc	well not a small number .

Example 223: Bmr018

507.485-508.498	c4	s^e	and you're tightening the boundaries .
508.498-509.51	c4	s^bsc	correcting the boundaries .

5.10 Group 9: Supportive Functions

This group contains tags that apply to utterances in which a speaker supports his own argument by defending himself, offering an explanation, or else offering additional details and utterances in which a speaker attempts to support another speaker by finishing the other speaker's utterance.

▪ Defending/Explanation <df>

The <df> tag marks cases in which a speaker defends his own point or offers an explanation. Often, the word "because" signals an explanation.

The <df> tag is often confused with the elaboration tag <e>. The two tags differ in that, as the <df> tag marks utterances in which a speaker defends a point or offers an explanation, the <e> tag marks utterances in which a speaker offers further details.

Example 224 through Example 229 present instances of the <df> tag in context:

Example 224: Bmr005

949.459-951.044	c4	s^ar	no no it isn't sensitive at all .
951.044-951.837	c4	s^df	i was just - i was jus- - i was overreacting just because we've been talking about it .

Example 225: Bmr005

1012.960-1019.350	c4	s^arp	but i - i mean - i think also to some extent its just educating the human subjects people in a way .
1019.350-1022.540	c4	s^df	because there's if uh - you know - there's court transcripts there's - there's transcripts of radio shows .

Example 226: Bmr007

14.467-15.967	cB	qy^rt	and uh shall i go ahead and do some digits ?
16.724-17.504	c3	h|s^ng	uh | we were going to do that at the end .
17.504-18.284	c3	qy^d^rt	remember ?
18.700-19.840	cB	s^bk|s	okay | whatever you want .

20.396-23.856	c3	s^co^df	just - just to be consistent from here on in at least that - that we'll do it at the end .
Example 227: Bmr009			
459.997-463.620	c2	s	but i had maybe made it too complicated by suggesting early on that you look at scatter plots .
463.620-467.244	c2	s^df	because that's looking at a distribution in two dimensions .
Example 228: Bro008			
1356.660-1357.940	c4	s^na	yeah because a lot of time that's true .
1357.940-1366.720	c4	s^df	there were a lot of times when we would try something and it didn't work right away even though we had an intuition that there should be something there .
Example 229: Bro015			
449.830-450.490	c0	s^nd	this week i haven't .
450.490-453.980	c0	s^df^ng	i've been - my whole time's been taken up with uh meeting recorder stuff .

▪ Elaboration <e>

The elaboration tag marks when a current speaker elaborates on a previous utterance of his by adding further details as opposed to simply continuing to speak on the same topic. When a speaker describes something using an example, the example is regarded as an elaboration.

The elaboration tag is often confused with the defending/explanation tag <df> which marks utterances in which a speaker defends a point or offers an explanation. As the defending/explanation tag revolves around reasons, the elaboration tag revolves around details.

A convention has been established in handling instances when a question is followed by an elaboration <e> which requires its own line. In such cases, the following elaboration could be considered a declarative form of the question. Instead, the elaboration receives a DA of <s^e>, along with any other necessary specific tags. The reasoning behind labeling an elaboration following a question as a statement <s> rather than a

question is that, if the elaboration were to be considered a question, then the elaboration itself would be asking something. For instance, if a speaker were to ask, "have you gone to that restaurant I suggested?", and then followed that question with an elaboration such as "the one on Sixth Street," labeling the elaboration as a type of question would indicate that the elaboration, "the one on Sixth Street," was actually eliciting some sort of answer. Instead, the question, "have you gone to that restaurant I suggested?", seeks an answer and the elaboration, "the one on Sixth Street," merely adds a detail to the question without actually asking something.

Elaborations are shown in context in Example 230 through Example 237:

Example 230: Bed011

1516.010-1520.440	c3	s^bu	although i guess it's only ten minutes each .
1520.440-1520.670	c3	qy^d^g^rt	right ?
1521.030-1521.480	c3	s^e	roughly .

Example 231: Bro004

1179.080-1185.130	c1	qw	well what was - is that i- - what was it that you had done last week when you showed - do you remember ?
1185.310-1188.230	c1	s^e^rt	wh- - when you showed me the - your table last week .

Example 232: Bmr024

1424.290-1427.230	c5	fg|s^df	well but - but | i put it under the same directory tree .
1427.230-1429.620	c5	fh|s^e	you know | it's in user doctor speech data m r .

Example 233: Bro004

2028.080-2038.300	c3	s^cs	so uh - we were thinking about is perhaps um - one way to solve this problem is increase the number of outputs of the neural networks .
2040.010-2044.450	c3	s^e.%--	doing something like um - um - phonemes within context and ==

Example 234: Bro004

2170.080-2175.840	c3	s	and basically the net- - network is trained almost to give binary decisions .

2177.730-2181.920	c3	s^e	and uh - binary decisions about phonemes .

Example 235: Bro004

2261.170-2264.060	c3	s	so you - you have more information in your features .
2264.060-2272.160	c3	s^e	so um - you have more information in the uh - posterior spectrum .

Example 236: Bro011

546.896-555.660	c1	fh|s^co^t^tc	so um - | i suggest actually now we - we - we sort of move on and - and hear what's - what's - what's happening in - in other areas .
555.660-562.490	c1	s^e^t	like what's - what's happening with your investigations about echos and so on .

Example 237: Bro011

1471.250-1476.140	c1	fh|s	and uh - | because in the ideal case we would be going for posterior probabilities .
1476.140-1481.030	c1	s^e	if we had uh - enough data to really get posterior probabilities .
1481.430-1486.460	c1	s^e	and if the - if we also had enough data so that it was representative of the test data .
1486.460-1491.500	c1	s^e	then we would in fact be doing the right thing to train everything as hard as we can .

▪ Collaborative Completion <2>

The collaborative completion tag <2> tag marks utterances in which a speaker attempts to complete a portion of another speaker's utterance. Whether the speaker whose utterance is completed by another speaker agrees with the content of the completion is inconsequential. If a speaker does agree with the completion, then the agreement is marked with the appropriate tag.

In some cases, a speaker attempts to complete another speaker's utterance and, in doing so, interrupts and stops the speaker whose utterance he is trying to complete. The interrupted speaker then resumes speaking, usually having either accepted or rejected the collaborative completion. If the collaborative completion is accepted, the tags <aa>, <na>, and <aap> are used to characterize the acceptance. Acceptance of a collaborative completion usually arises in the form of a "yes" word, as those labeled with the <aa> tag, or else by mimicking the completion, and such is marked with the <na> tag. If the collaborative completion is rejected, the tags <ar>, <nd>, <ng>, and <arp> are used to characterize the rejection. Rejection of a collaborative completion usually arises in the form of a "no" word, as those labeled with the <ar> tag, or else by a speaker completing his utterance in a manner which differs from the collaborative completion, and such is marked with either the <nd> or <ng> tag.

Collaborative completions in context are seen in Example 238 through Example 245:

Example 238: Bed003

463.416-469.753	c2	s.%-	because we were thinking uh - if they were in a hurry there'd be less likely to - like - or th- ==
469.220-469.780	c3	s^2	want to do vista .

Example 239: Bed003

593.810-599.330	c3	s	that kind of thing is all uh - sort of - you know - probabilistically depends on the other things .
598.030-599.260	c4	qy^bu^d^rt^2	inferred from the other ones ?

Example 240: Bmr007

1652.350-1654.960	cB	s	well but from the acoustic point of view it's all good .
1655.120-1655.620	c4	s^aa^2	is the same .

Example 241: Bmr009

1937.990-1941.720	c3	s.%--	i think originally it was north - northwest but ==
1941.420-1941.930	c0	s^2	northwest .

Example 242: Bmr012

435.384-437.674	c2	s.%-	but there's a significant amount of ==
436.608-437.368	c5	qy^d^rt^2	non zero ?

Example 243: Bmr012

1825.930-1828.470	c2	s	but i d- - i know the lapel is really suboptimal .
1827.450-1827.910	c4	qy^rt^2	is awful ?

Example 244: Bro004

1462.620-1472.340	c3	s^e	the uh - the um - networks are trained with noise from aurora - t i digits .
1471.470-1471.880	c4	s^2	aurora two .

Example 245: Bmr008

177.000-180.000	c1	qw	how fine a resolution do you need on that for this ?
181.000-182.000	c2	s^2	is the question .

5.11 Group 10: Politeness Mechanisms

This group contains tags that apply to utterances in which speakers exhibit courteousness.

▪ Downplayer <bd>

The downplayer tag <bd> marks cases in which a speaker downplays or de-emphasizes another utterance. The utterance that is downplayed may be uttered by the same speaker or a different speaker.

Apologies, compliments, and other courteous utterances are often downplayed. In other cases, a speaker makes a strong assertion and then downplays it.

Downplayers vary in form. Some may be long utterances and others may be quite short. The following is a list of common short downplayers: "that's okay," "that's all right," "it's okay," "I'm kidding," "it's just a thought," and "never mind."

Downplayers in context are presented in Example 246 through Example 252:

Example 246: Bmr012

960.050-960.790	c8	s^ba	congratulations .
961.254-964.724	c2	s^bd	well it was i mean - i really didn't do this myself .

Example 247: Bmr005

954.368-958.498	c1	s^t	i - i came up with something from the human subjects people that i wanted to mention .
958.498-959.743	c1	s^bd	i mean it fits into the area of the mundane .

Example 248: Bed006

1953.730-1954.170	c1	s^fa	sorry .
1955.080-1955.380	cA	s^bd	it's okay .

Example 249: Bro018

501.447-503.797	c2	s	but suppose you don't really know what the right thing is .
504.377-508.497	c2	s	and that's what these sort of dumb machine learning methods are good at .
510.950-511.540	c2	s^bd	it's just a thought .

Example 250: Bmr011

2778.000-2779.000	c0	s.%--	and then the other thing is ==
2780.000-2781.500	c0	s^bd	i don't know if this is at all useful .

Example 251: Bmr029

1232.580-1238.270	c2	s.%--	the - the other difference that we'd have to take care of is that ==
1238.270-1242.430	c2	fh|s	uh - | yeah we - we don't have a mike that uh is particular to a person .
1242.430-1244.510	c2	s	and so we'll have to do some clustering .
1244.510-1249.770	c2	s	and that'll be another another uh issue too .
1252.160-1253.810	c2	s^bd	but it - it - i could be wrong .

Example 252: Bro010

631.950-633.005	c2	s	so you would think as long as it's under half a second or something .
633.005-633.533	c2	s^bd	uh i'm not an expert on that .

▪ Sympathy <by>

The <by> tag marks utterances in which a speaker exhibits sympathy. Oftentimes, the phrase "I'm sorry" is used sympathetically. However, that very phrase also has the potential to be marked as a repetition request <br> or as an apology <fa>, depending upon its function.

Instances of the <by> tag in context are displayed in Example 253 through Example 255:

Example 253: Bed003

3033.120-3034.070	c1	s^rt	so i had to reboot .
3033.440-3034.140	c4	s^by^fe^rt	oh no .

Example 254: Bmr027

1972.740-1977.040	c0	s	and then you can see here g p s was misinterpreted .
1977.450-1978.850	c0	s^by.%--	it's just totally understanda- ==

Example 255: Bmr027

2186.760-2189.800	c3	s.%--	without thinking about it when i offered up my hard drive last week ==
2189.260-2190.040	c5	s^by^fe	oh no !

▪ Apology <fa>

An utterance is marked as an apology <fa> when a speaker apologizes for something he did (e.g., after coughing, sneezing, interrupting another speaker, etc.).

The phrase "I'm sorry," depending upon its usage, may be interpreted as a repetition request <br> or as sympathy <by>.

Additionally, the phrase "excuse me" can be used as an apology <fa> or else can be found within a suggestion <cs>. The phrase is found within a suggestion when it precedes something for which a speaker is negotiating permission (Jurafsky 35).

Apologies <fa>, some of which are in context, are shown in Example 256 through Example 261:

Example 256: Bmr001

876.821-877.541	c1	s	so we could have eight .
876.899-877.029	c8	s^aa	yeah .
878.126-878.426	c8	s^bc	nine .
878.352-878.672	c1	s^bk^m	nine .
878.672-879.432	c1	s^fa|s^r	excuse me | nine .

Example 257: Bmr005

832.753-837.990	c5	s^fa	sorry to interrupt .

Example 258: Bmr009

1563.000-1566.500	c0	s.%--	because the date is when you actually read the digits and the time and ==
1566.500-1568.250	c0	s^fa	excuse me .
1568.250-1570.000	c0	s^bsc	the time is when you actually read the digits but i'm filling out the date beforehand .

Example 259: Bmr018

217.760-219.630	c1	s^fa	he's - i - i'm sorry i should have forwarded that along .

Example 260: Bmr026

1202.170-1203.530	c3	s^fa	oh i'm sorry i misunderstood .

Example 261: Bmr006

1202.100-1205.320	c9	s^fa	sorry i- have to - sorry i have to leave .

▪ Thanks <ft>

The <ft> tag marks utterances in which a speaker thanks another speaker.

Instances of the <ft> tag, one of which with surrounding context, are shown in Example 262[9] through Example 264:

Example 262: Bed003

216.310-217.340	c4	sj^ba	nice coinage .
219.833-220.463	c2	s^ft	thank you .

Example 263: Bmr007

3266.710-3267.720	c8	s^ft	thanks .
3267.810-3268.270	c8	s^ft	appreciate that .

Example 264: Bmr024

2928.220-2929.450	c3	s^ft	thank you for the box .

▪ Welcome <fw>

The <fw> tag marks utterances which function as responses to utterances marked with the thanks tag <ft>. Phrases such as "you're welcome" and "my pleasure" are marked with the welcome tag <fw>.

No instances of the <fw> tag exist within the Meeting Recorder data.

5.12 Group 11: Further Descriptions

This group contains various tags that do not fit into any of the pre-established groups. The tags within this group characterize meeting agendas, changes in topic, exclamatory material, humorous matter, self talk, third party talk, as well as syntactic and prosodic features of utterances.

9 Regarding the use of the tag <sj> in Example 262, refer to footnote 7.

▪ Exclamation <fe>

The <fe> tag marks utterances in which a speaker expresses excitement, surprise, or enthusiasm. Utterances marked with the <fe> tag, excluding quotes, are punctuated with an exclamation mark < ! > within the transcript.

Utterances marked with the <fe> tag can range from consisting of one word to a lengthy string of words. The most salient factor in determining if an utterance is an exclamation is the level of energy. Exclamations usually have a much higher energy than that of the surrounding utterances.

Instances of the <fe> tag are seen in Example 265 through Example 279:

Example 265: Bed003			
47.760-47.920	c3	s^fe	wow !
Example 266: Bed003			
119.945-120.205	c2	s^fe	aha !
Example 267: Bed003			
1626.000-1626.240	c4	s^fe	whew !
Example 268: Bed003			
1676.950-1677.070	c2	s^fe	oops !
Example 269: Bed003			
1761.080-1761.190	c4	s^fe	god !
Example 270: Bed003			
1794.550-1794.750	c2	s^fe	oh !
Example 271: Bed003			
2004.230-2004.480	c3	s^fe	ha !
Example 272: Bed004			
3200.900-3201.260	c2	s^fe	oh yeah !

Example 273: Bmr009			
2394.570-2396.130	c0	s^fe	oh no !
Example 274: Bed003			
133.711-134.431	c4	s^fe^j	i can read !
Example 275: Bmr005			
1956.430-1962.910	c4	s^fe^m	twelve minutes !
Example 276: Bmr008			
3293.420-3294.600	c3	s^fe^t3	oh it's seventy five per cent !
Example 277: Bed006			
2876.320-2877.010	cA	s^fe^j	damn this project !
Example 278: Bro012			
3213.110-3215.050	c0	s^fe^rt	then do some more spectral subtraction !
Example 279: Bmr015			
525.983-527.896	c0	s^ba^fe	so that's amazing you showed up at this meeting !

▪ About-Task <t>

The about-task tag marks utterances that are in reference to meeting agendas or else address the direction of meeting conversations with regard to meeting agendas.

The about-task tag is not to be confused with the topic change tag <tc>. The topic change tag marks utterances which either end or begin a topic regardless of a meeting agenda. The about-task tag marks utterances which regard previously established items to be discussed or managed within a meeting. However, this is not to say that an utterance can only be marked by either the about-task tag or the topic change tag. Rather, both tags may be used to label an utterance so long as an utterance is changing a topic in reference to a meeting agenda. For instance, if a speaker is talking about a topic that is not part of the meeting agenda and then he or another speaker changes the topic and mentions the agenda, then the utterance in which the change in

topic and reference to the agenda occurred would be marked with the tags <t> and <tc>.

Additionally, a restriction applies to the usage of the about-task tag. The about-task tag is used to mark utterances which mention agendas and agenda items. In essence, the about-task tag marks utterances which revolve around what tasks are to be completed within the course of a meeting. So what is marked with the about-task tag is what is to be accomplished within a meeting, but when an agenda item is in the process of being "accomplished," it is not marked by the about-task tag. For instance, if a speaker mentions that an agenda item is to discuss a certain subject and then other speakers begin to discuss that subject, then the utterance mentioning that the agenda item to discuss a subject is marked with the about-task tag. However, the actual discussion about the subject is not marked with the about-task tag.

Example 280 through Example 289 display instances in which the about-task tag is used:

Example 280: Bmr005			
381.017-383.717	c4	s^t	um - so i - i do have a - an agenda suggestion .
Example 281: Bmr006			
1224.410-1229.080	c3	fh|s^t^tc	and | then um i guess another topic would be where are we in the whole disk resources question .
Example 282: Bmr006			
4464.590-4466.090	c3	s^co^t^tc	let's do digits .
Example 283: Bmr007			
1938.400-1941.590	c3	s^t^tc	speaking of taking control you said you had some research to talk about .
Example 284: Bmr008			
15.000-18.000	c1	s^co^rt^t	let's discuss agenda items .
Example 285: Bmr010			
239.005-242.305	c6	qh^t^tc	so yeah why don't we do the speech nonspeech discussion ?

Example 286: Bmr012

209.361-211.781	c4	qy^cs^rt^t	okay so should we do agenda items ?

Example 287: Bmr012

219.415-223.365	c4	s^t	uh - well i have - i want to talk about new microphones and wireless stuff .

Example 288: Bmr014

51.589-52.929	c8	qo^t	any agenda items today ?
53.672-61.382	c4	s^t	i want to talk a little bit about getting - how we're going to to get people to edit bleeps parts of the meeting that they don't want to include .

Example 289: Bro022

35.044-41.771	c0	qy^cs^rt^t^tc	so should we just do the same kind of deal where we go around and do uh status report kind of things ?

▪ Topic Change <tc>

The <tc> tag marks utterances which either begin or end a topic. As the <tc> tag marks when a topic changes, once the topic has indeed changed and a new topic is in the course of discussion, the discussion of the new topic is not marked with the <tc> tag.

Oftentimes, a speaker will utter a floor grabber <fg> and then introduce a new topic. As the floor grabber appears as though it is used as a mechanism to gain the floor and introduce a new topic, and in effect signals a change in topic, it is not marked with the <tc> tag. Rather, only utterances which convey a change in topic are marked with the <tc> tag. In which case, a speaker must specify in his utterance that he wishes to end a topic or else he must state that he wishes to begin a new topic either by initiating and specifying a new topic or else by merely stating that he wishes to talk about something else.

The <tc> tag may be used in conjunction with the about-task tag <t>. The tag description for the about-task tag details the rules governing such usage.

Topic changes, some of which with surrounding context, are shown in Example 290 through Example 296:

Example 290: Bro015

713.450-713.910	c3	fg	let's see .
715.580-725.090	c3	fh|s^cs^t^tc	um | why don't - why don't we uh - if there aren't any other major things why don't we do the digits and then - then uh - turn the mikes off .

Example 291: Bro007

1770.390-1776.060	c1	s^co^t^tc	k uh - if nobody has anything else maybe we should go around do - do our digits - do our digits duty .

Example 292: Bmr008

2697.000-2698.000	c3	s^t^tc	okay enough on forms .

Example 293: Bro004

3756.280-3766.420	c1	s^co^t^tc	so with that maybe we should uh - go to our digit recitation task .

Example 294: Bro013

1899.320-1899.750	c0	fg	okay .
1902.920-1905.180	c0	fh|s^tc	um | i think we're sort of done .

Example 295: Bro013

691.240-691.550	c0	fg	okay .
691.680-692.500	c0	s^tc	that was that topic .
692.500-693.140	c0	qw^t^tc	what else we got ?

Example 296: Bro015

96.560-99.450	c3	s	anyway hynek will be here next week and maybe he'll know more about it .
105.440-105.990	c2	fg	oh yeah .
106.680-111.530	c2	s^tc	well the news more specifically t- - for aurora .
111.530-112.450	c2	fh	um ==
113.880-121.622	c2	s	so i guess there was again a conference call but uh they are not decide on everything yet .

▪ Joke <j>

The <j> tag marks utterances of humorous or sarcastic nature. If a speaker is attempting to be humorous, then the utterances containing humorous material are marked with the <j> tag, regardless of how those utterances received by other speakers.

Utterances marked with the <j> tag are often context dependent, in that jokes are often made with regard to the current topic at hand. A majority of jokes require the surrounding context in order to be perceived as jokes, as when jokes are seen without surrounding context, they usually tend not to appear as being humorous or sarcastic.

Example 297 through Example 301 display jokes with surrounding context:

Example 297: Bro021

1877.030-1878.270	c5	qw^rt	what - what is v t s again ?
1878.070-1881.140	c4	s	uh vectorial taylor series .
1880.420-1881.070	c5	s^bk	oh yes .
1881.070-1881.710	c5	s^aa	right right .
1882.530-1885.350	c5	s	i think i ask you that every single meeting .
1885.350-1886.750	c5	qy^g	don't i ?
1884.860-1885.590	c4	qw^br	what ?
1886.750-1888.160	c5	s	i ask you that question every meeting .
1887.310-1888.120	c4	s^aa	yeah .
1888.080-1890.790	c1	s^j	so that'd be good from - for analysis .
1890.790-1892.140	c1	s^df^j	it's good to have some uh cases of the same utterance at different - different times .
1891.680-1893.200	c5	s^bk	yeah .
1893.200-1894.720	c5	qw^j	what is v t s ?

Example 298: Bro017

2173.380-2175.970	c1	s^cs.%--	but what you can do - i'm confident we ca- ==
2175.970-2178.550	c1	s	well i'm reasonably confident and i putting it on the record .
2178.550-2178.730	c1	qy^d^f^rt	right ?
2178.730-2183.790	c1	s^j	i mean y- - people will listen to it for - for centuries now .

Example 299: Bro016

1386.190-1388.280	c5	qy	do you have speaker information ?

1388.930-1393.370	c4	s^j	social security number .
1389.800-1392.410	c5	s^ba	that would be good .
1391.980-1395.370	c1	s	like we have male female .
1392.410-1394.130	c5	s^j	bank pin .

Example 300: Bro014

8.347-9.712	c1	fg	okay .
9.712-11.077	c1	qy^j^rt	did you solve speech recognition last week ?

Example 301: Bro014

40.831-41.701	c2	qy^rt	is he going to come here ?
42.154-44.306	c1	h	uh ==
44.306-45.382	c1	s^j^na	well we'll drag him here .
45.382-46.458	c1	s^j	i know where he is .

▪ Self Talk <t1>

The <t1> tag is used when a speaker talks to himself. Often, utterances marked as self talk are quieter and softer than the surrounding speech.

A case in which the self talk tag is used occurs when a speaker is writing something down and consequently repeats what he writes to himself. In other instances, a speaker may be attempting to make some sort of a calculation or solve a problem and talks to himself in the process of figuring out the answer.

Although it has been mentioned that certain types of utterances, such as backchannels <b> and floor holders <fh>, are not forms of direct communication between speakers, these utterances are not considered self talk either.

Example 302 through Example 305 display instances of the self talk tag, most of which are shown with surrounding context.

Example 302: Bmr007

787.674-792.891	c8	s.%--	in that case um my c- the coding that i was using - since we haven't uh incorporated adam's uh coding of overlap yets the coding of ==
792.891-798.109	c8	s^t1	yeah yets is not a word .

Example 303: Bro018

2987.260-2989.580	c2	s.%--	i - i - i th- - i think he ==
2991.360-2992.210	c2	qo^t1	what am i saying here ?

Example 304: Bro014

50.154-51.928	c4	s^t1	doo doo doo .
53.633-54.207	c4	s^t1	doo doo .

Example 305: Bro021

2230.830-2235.540	c1	fh|s.%--	uh - | so that's log of x plus log of one plus uh ==
2236.170-2236.760	c1	fh	well .
2237.360-2238.270	c1	qy^rt^t1	is that right ?
2238.270-2239.180	c1	s^e^t1.%--	log of ==
2238.710-2240.560	c3	s^t1	one plus n by x .

▪ Third Party Talk <t3>

The third party talk tag marks utterances of side conversations. Side conversations are conversations which are not directed toward the main conversation and may only consist of a handful of utterances or may be quite lengthy.

Instances of third party talk are shown in Example 306 through Example 309 with surrounding context.

Example 306: Bmr007

1389.340-1394.230	cA	s	so so - actually um that's in part because the nodding - if you have visual contact the nodding has the same function .
1394.230-1399.120	cA	s	but on the phone in switchboard you - you - that wouldn't work .
1398.900-1399.680	cB	s^na	yeah you don't have it .
1399.120-1401.260	cA	s	so so you need to use the backchannel .
1401.140-1405.880	cB	s^t3.%--	your mike is ==
1403.000-1410.570	c0	qy^r^rt	so in the two person conversations when there's backchannel is there a

			great deal of overlap in the speech ?
1405.880-1410.630	cB	s^co^t3	that is an earphone so if you just put it so it's on your ear .
1410.570-1411.000	c0	qrr.%--	or ?==
1411.000-1417.160	c0	s	because my impression is sometimes it happens when there's a pause .
1411.170-1411.450	c1	s^aa	yes .
1411.250-1411.660	cB	s^t3	there you go .
1412.160-1412.380	c1	b	yeah .
1412.630-1412.940	cB	s^ft^t3	thank you .

Example 307: Bro004

1109.570-1111.640	c2	qy^d^rt^t3	these numbers are uh - ratio to baseline ?
1110.650-1111.840	c1	qw.%-	so i mean - wha- - what's the ?==
1111.840-1121.980	c1	qy^bu^d	this - this chart - this table that we're looking at is um - sho- - is all testing for t i digits ?
1123.260-1126.910	c3	s^rt	so you have uh - basically two uh - parts .
1123.610-1123.880	c9	s^t3	bigger is worse .
1123.880-1125.290	c9	s^t3	this is error rate i think .
1125.570-1125.690	c9	s^ar^t3.%	no no .
1125.640-1126.040	c2	s^t3	ratio .
1126.910-1130.580	c3	s^rt	the upper part is for t i digits .
1130.580-1134.240	c3	s^rt	and it's divided in three rows of four - four rows each .
1128.380-1128.640	c9	s^aa^t3	yeah yeah yeah .

Example 308: Bro003

2159.050-2161.170	c0	qy^rt	is that - was that distributed with aurora ?
2161.170-2162.230	c0	qrr.%--	or ?==
2161.490-2161.730	c8	s.%	italian .
2161.960-2163.020	c2	qr^bu^d^rt^t3	one l or two l's ?

Example 309: Bed012

998.980-1001.180	c1	s^rt	and we get a certain - we have a situation vector and a user vector and everything is fine .
1001.540-1004.130	c1	%-	an- - an- - and - and our - and our ==
1002.750-1005.980	c2	qy^rt^t3	did you just sti- - did you just stick the

			m- - the - the - the microphone actually in the tea ?
1005.790-1008.320	c0	s^ar^t3	no .
1008.500-1009.530	c1	fh	and um ==
1009.480-1010.290	c0	s^ng^t3	i'm not drinking tea .
1010.290-1011.100	c0	qw^t3	what are you talking about ?
1011.770-1012.260	c2	s^bk^t3	oh yeah .
1012.260-1012.750	c2	s^fa^t3	sorry .
1013.580-1017.780	c1	s^co^rt	let's just assume our bayes net just has three decision nodes for the time being .

▪ Declarative Question <d>

The declarative question tag marks questions which have the syntactic appearance of a statement. In declarative questions, the subject precedes the verb and subject-auxiliary inversion and wh-movement do not occur. It is not uncommon for a rising tone <rt> to be found on a declarative question, however a rising tone does not always function as an indicator that a question is being asked.

Additionally, tag questions <g> are often declarative questions. This is only the case when subject-auxiliary inversion does not occur (e.g., "you do?" rather than "do you?") or if the question consists of only one word (e.g., "right?") or does not contain a verb (e.g., "the tenth of July?"). However, if a question consists of one word and that word is a "wh" word, such as those mentioned in the tag description for wh-questions <wh>, then neither the tags <d> or <g> are used.

Declarative questions are seen in Example 310 through Example 324:

Example 310: Bro021

979.242-980.846	c1	qy^d^g^rt	right ?

Example 311: Bro013

2020.370-2020.610	c0	qy^d^f^g	you know ?

Example 312: Bro021

2493.820-2495.190	c4	qy^d^g^rt	no ?

Example 313: Bmr007

92.862-98.798	c3	fh|qo^d^rt	um | and anything else anyone wants to talk about ?

Example 314: Bmr007			
112.365-116.868	c3	fh|qo^d^rt	um | and anything else ?
Example 315: Bmr007			
117.088-118.018	c3	qo^d	nothing else ?
Example 316: Bmr007			
171.144-171.704	c0	qy^d^rt^2	same idea ?
Example 317: Bmr007			
628.021-630.973	c3	qy^bu^d	oh so the bottom three did have s- stuff going on ?
Example 318: Bmr007			
653.124-653.594	c3	qy^d	you don't know ?
Example 319: Bmr021			
342.000-343.000	c4	qy^bu^d^rt	a wired one ?
Example 320: Bed006			
2804.550-2807.290	c4	qy^bu^d^rt	or you'd like - so you're saying you could practically turn this structure inside out ?
Example 321: Bmr024			
929.052-930.972	c4	qy^d	the references for - for those segments ?
Example 322: Bmr024			
1075.910-1081.850	c3	fg|qy^d^t^tc	um | another one that we had on adam's agenda that definitely involved you was s- - something about smartkom ?
Example 323: Bro017			
2117.620-2122.540	c5	qy^d^rt	so that effectively the c one never really

			contributes to the score ?

Example 324: Bro017

2487.900-2489.260	c5	qy^d^rt	see how many cycles we used ?

▪ Tag Question <g>

A tag question follows a statement and is a short question seeking confirmation of that statement. Tag questions receive a general tag of <qy> and are often used in conjunction with the "follow me" tag and the declarative question tag <d>. The tag description for declarative questions <d> discusses the instances in which it may be used in conjunction with the tag <g>. Utterances preceding tag questions are labeled as statements <s> rather than declarative yes/no questions <qy^d>.

Tag questions are often found following statements marked with the understanding check tag <bu>.

Common utterances marked with the <g> tag include, but are not limited to, the following: "right?", "yes?", "yeah?", "no?", "okay?", "isn't it?", "correct?", "won't it?", "doesn't it?", and "you know?".

Tag questions in context are seen in Example 325 through Example 334:

Example 325: Bed011

2073.940-2074.690	c1	s^bu	exchange money is an errand .
2074.690-2075.440	c1	qy^d^g	right ?

Example 326: Bed003

407.887-409.477	c2	s	so then our next idea was to add a middle layer .
409.477-409.777	c2	qy^d^f^g	right ?

Example 327: Bed003

1391.100-1398.880	c1	s	in the sense that you know - if it's tom - the house of tom cruise you know - it's enterable but you may not enter it .
1399.230-1399.520	c1	qy^d^f^g^rt	you know ?

Example 328: Bed003

2298.190-2301.170	c1	s:s	and then the persons says um - yeah i want to see it .
2302.210-2302.320	c1	qy^d^g	yeah ?

Example 329: Bed004

3059.570-3065.040	c2	s	there - the - the land- - the construction implies the there's a con- - this thing is being viewed as a container .
3065.920-3066.250	c2	qy^d^f^g	okay ?

Example 330: Bmr001

95.697-98.097	c8	s	and this - this one is right at the end of the table .
98.477-98.757	c8	qy^d^f^g	okay ?

Example 331: Bmr005

1473.790-1474.370	c8	s^m	that's a lot of overlap .
1474.370-1474.940	c8	qy^d^g^rt	yeah ?

Example 332: Bmr001

1237.390-1238.960	c1	fg|s^bu	yeah | so we don't store any of our audio formats compressed in any way .
1238.960-1240.530	c1	qy^d^g	do we ?

Example 333: Bmr005

1257.220-1260.490	c8	fg|s^bu	well | you weren't talking about just overlaps .
1260.490-1260.740	c8	qy^d^g^rt	were you ?

Example 334: Bmr005

1763.010-1764.720	c2	fh|s	i mean - | the normalization you do is over the whole conversation .
1764.720-1766.490	c2	qy^g^rt	isn't it ?

▪ Rising Tone <rt>

The rising tone tag is used to mark utterances in which a speaker's tone rises at the end of his utterance. Rising tones at the end of utterances occur in both questions and statements. Although intonation does not constitute a dialog act, the use of the <rt> tag provides useful information for automatic speech recognition.

5.13 Group 12: Disruption Forms

As stated in Section 3.4, disruption forms are used to mark utterances that are indecipherable, abandoned, or interrupted. Only one disruption form may be used per utterance. Guidelines and restrictions surrounding the format and use of disruption forms that are not mentioned in the tag descriptions for the indecipherable, interrupted, abandoned, and nonspeech tags are found in Section 3.4.

Examples are not provided within the tag descriptions for the indecipherable, interrupted, and nonspeech tags, as they require the corresponding audio portion in order to convey why it is that an utterance is indecipherable, interrupted, abandoned, or is considered nonspeech.

Additionally, Section 2 discusses segmentation and proves to be of much assistance in using disruption forms.

▪ Indecipherable <%>

The indecipherable tag marks indecipherable speech such as mumbled or muffled words or utterances that are too difficult to hear on account of the microphone picking up sounds from breathing.

The indecipherable tag <%> is not to be confused with the nonspeech tag <x>. The nonspeech tag <x> is used for sound segments which are silent or otherwise contain non-vocal sounds such as doors slamming and phones ringing. The nonspeech tag <x> does not apply to sounds such as breathing and sighs, as these are vocal sounds. However, sounds such as coughing and sneezing may be considered vocal sounds but are instead categorized with the nonspeech variety.

▪ Interrupted <%->

The interrupted tag marks incomplete utterances in which a speaker stops talking on account of being interrupted by another speaker. This tag is not to be confused with the abandoned tag <%--> which is used to mark instances in which a speaker intentionally abandons an utterance.

As the most salient examples of the interrupted tag involve speakers giving up the floor immediately, the interrupted tag is even used in cases in which a speaker has the floor and is interrupted but does not immediately relinquish the floor. The reasoning behind using the interrupted tag rather than the abandoned tag <%--> in such instances is because the speaker gives up the floor on account of being interrupted.

▪ Abandoned <%-->

The abandoned tag marks utterances which are abandoned by a speaker. Abandoned utterances occur when a speaker trails off or else chooses to either reformulate an utterance or change the topic by abandoning his current utterance and beginning a new one.

The issues mentioned in Section 2 regarding segmentation are of crucial importance when using the abandoned tag. For instance, if a speaker begins an utterance and restarts it in a different manner, and the prosody and pauses are such that the original utterance and the restarted version constitute a single utterance, the entire utterance remains intact and is labeled in a way that reflects its completeness. The utterance is not split at the point between the beginning and the restarted portion, and the beginning portion is not marked as being abandoned. In Example 335, an utterance is shown that is restarted and remains intact, rather than being split at the region where it is restarted:

Example 335: Bro021

1730.970-1733.270	c3	s	and it - it - it gave like - i just got the signal out .

Abandoned utterances are seen with surrounding context in Example 336 through Example 339:

Example 336: Bro021

186.057-194.998	c2	s	well uh there is one thing that we can observe is that the mean are more different for - for c zero and c one than for the other coefficients .

195.634-196.920	c2	fh	and ==
198.663-199.323	c2	fh	yeah .
200.819-203.469	c2	s.%--	and - yeah it - the c one is ==
203.469-215.256	c2	s	there are strange - strange thing happening with c one is that when you have different kind of noises the mean for the - the silence portion is - can be different .

Example 337: Bro021

261.708-276.050	c2	fh|s^rt	um | a third thing is um that instead of t- - having a fixed time constant i try to have a time constant that's smaller at the beginning of the utterances .
276.050-279.990	c2	s^e	to adapt more quickly to the r- - something that's closer to the right mean .
280.273-282.108	c2	fh	t- - t- - um ==
283.723-286.491	c2	s^bk	yeah .
286.491-287.875	c2	s	and then this time constant increases .
287.875-289.259	c2	s.%--	and i have a threshold that ==
289.855-298.584	c2	s	well if it's higher than a certain threshold i keep it to this threshold to still uh adapt um the mean when - if the utterance is uh long enough to - to continue to adapt after like one second .

Example 338: Bro026

1235.390-1237.000	c3	qy^rt	would - would that set on the handset ?
1237.000-1237.420	c3	qrr.%--	or ?==

Example 339: Bro025

118.800-127.061	c1	s^na	yeah i mean it's - it's actually uh very similar .
127.061-128.844	c1	s.%--	i mean if you look at databases ==
129.611-130.740	c1	fh	uh ==
132.232-141.440	c1	s	the uh one that has the smallest - smaller overall number is actually better on the finnish and spanish .
142.317-147.387	c1	fh|s	uh | but it is uh worse on the uh aurora .
145.334-146.817	c4	s^2.%--	it's worse on ==
147.387-151.000	c1	s^bsc	i mean on the uh t i- - t i digits .

- **Nonspeech <x>**

The nonspeech tag marks any utterance that is unintelligible on account of non-vocal noises such as doors slamming, phones ringing, and problems with a recording. The nonspeech tag also marks coughing and sneezing sounds, as well as utterances filled with silence.

The nonspeech tag is not to be confused with the indecipherable tag <%> which marks utterances that are unintelligible on account of muffled speech, mumbling, breathing sounds, and sighing.

5.14 Group 13: Nonlabeled

Group 13 solely contains the nonlabeled tag <z>. As stated in Section 3.2, the tag <z> does not provide any information regarding the characteristics and functions of utterances as the tags of the other groups do, and for this reason it is separated from those groups.

- **Nonlabeled <z>**

The nonlabeled tag marks utterances that are not to be labeled with a DA. Types of utterances that are not to be labeled are those containing to pre- or post-meeting chatter, those pertaining to "bleeped" portions in the corresponding audio file, and those pertaining to the reading of digits. The tag <z> marks utterances which otherwise would be labeled with DAs but instead are intentionally not to be labeled.

An additional, but rare, instance in which the tag <z> is used arises when one speaker wears multiple microphones, thus causing his utterances to be recorded on multiple channels. In such a case, the speaker's utterance on his original microphone (i.e. the microphone he has been using throughout the meeting) receives the appropriate DA. Subsequent channels with the same utterance are labeled with the tag <z> and receive a note of "DUPLICATED-MICROPHONE" in the comment field.

As a side note, the convention of marking pre- and post-meeting chatter with the <z> tag was a fairly recent development. In which case, a number of utterances which are now marked with the <z> tag were originally marked with DAs consisting of the tags found in Groups 1 through 12 along with adjacency pairs. As these original DAs have been replaced with the <z> tag, the APs, however, have been preserved per chance they are of use for future research. As the information derived from APs is optimized with the use of corresponding DAs, APs corresponding to utterances marked with the

<z> tag can only provide optimal information upon being relabeled with DAs consisting of the tags found in Groups 1 through 12.

APPENDIX 1: LABELED MEETING SAMPLE

A labeled five-minute portion of Bro021 is shown below. Included are start and end times, channel numbers, DAs, adjacency pairs, and the corresponding portions of the transcript.

1828.250-1832.820	c3	s		i like plugged some groupings for computing this eigen- - uh uh uh s- - values and eigenvectors .
1832.820-1839.250	c3	s		so just - i just some small block of things which i needed to put together for the subspace approach .
1839.250-1845.680	c3	s		and i'm in the process of like building up that stuff .
1846.670-1849.080	c3	fh		and um ==
1850.400-1852.790	c3	fh		uh - yeah .
1854.120-1856.580	c3	s		i guess - yep i guess that's it .
1856.580-1859.040	c3	s		and uh th- - th- - that's where i am right now .
1859.620-1860.630	c3	fh		so .
1861.560-1863.000	c5	qo^tc	1a	oh how about you carmen ?
1862.830-1865.740	c4	s	1b	huh i'm working with v t s .
1866.330-1869.160	c4	fh|s		um | i do several experiment with the spanish database first .
1869.150-1873.400	c4	s^e	2a	only with v t s and nothing more .
276.050-279.990	c2	s^e		to adapt more quickly to the r- - something that's closer to the right mean .
1875.520-1876.580	c4	s^e		no l d a .
1873.400-1875.520	c4	s^e		not v a d .
1876.580-1877.640	c4	s^e		nothing more .
1877.030-1878.270	c5	qw^rt	2b.3a	what - what is v t s again ?
1878.070-1881.140	c4	s	3b.4a	uh vectorial taylor series .
1878.320-1879.090	c3	%-		new ==
1880.420-1881.070	c5	s^bk	4b	oh yes .
1881.070-1881.710	c5	s^aa	4b+	right right .
1881.350-1883.060	c4	s		to remove the noise too .
1882.530-1885.350	c5	s	5a	i think i ask you that every single meeting .
1885.350-1886.750	c5	qy^g	5a+	don't i ?

1884.860-1885.590	c4	qw^br	5b.6a	what ?
1886.750-1888.160	c5	s	6b.7a	i ask you that question every meeting .
1887.310-1888.120	c4	s^aa	7b-1	yeah .
1888.120-1888.930	c4	%-		if - well ==
1888.080-1890.790	c1	s^j	7b-2.8a	so that'd be good from - for analysis .
1890.790-1892.140	c1	s^df^j	7b-2+.8a+	it's good to have some uh cases of the same utterance at different - different times .
1892.140-1893.490	c1	fh		yeah .
1891.680-1893.200	c5	s^bk	8b	yeah .
1893.200-1894.720	c5	qw^j	8b+.9a	what is v t s ?
1895.100-1896.260	c4	s^m	9b	v t s .
1896.260-1897.410	c4	s.%--		i'm sor- ==
1897.410-1898.980	c4	s.%--		well um the question is that ==
1898.980-1900.540	c4	fh		well .
1900.540-1903.300	c4	s		remove some noise but not too much .
1903.700-1909.290	c4	fh|s		and | when we put the m- - m- - the them - v a d the result is better .
1909.290-1915.030	c4	s		and we put everything the result is better .
1915.030-1920.770	c4	s	10a	but it's not better than the result that we have without v t s .
1921.110-1921.780	c4	s^ar		no no .
1923.210-1924.060	c1	s^bk	10b	i see .
1924.060-1930.290	c1	s.%--	11a	so that given that you're using the v a d also the effect of the v t s is not so far ==
1929.630-1930.270	c4	s^na	11b	is not .
1930.780-1934.640	c1	qw^rt	12a	do you - how much of that do you think is due to just the particular implementation and how much you're adjusting it ?
1934.640-1938.490	c1	qw.%--	12a+	or how much do you think is intrinsic to ?==
1936.770-1937.830	c4	s^no	12b	pfft i don't know .
1937.830-1938.880	c4	s^df.%--	12b+	because ==
1938.880-1940.500	c4	fh		hhh ==
1939.210-1941.350	c2	qy	13a	are you still using only the ten first frame for noise estimation ?
1941.350-1943.490	c2	qrr.%--		or ?==
1944.260-1953.610	c4	h|s^rt	13b	uh | i do the experiment using

				only the f- - onl- - uh to use on- - only one fair estimation of the noise .
1944.890-1946.040	c2	qrr.%--		or i- ?==
1948.290-1948.820	c2	b		yeah .
1949.670-1950.580	c2	b		huh .
1953.610-1961.850	c4	s	13b+	and also i did some experiment uh doing um a lying estimation of the noise .
1962.430-1965.860	c4	s.%--		and well it's a little bit better but not ==
1966.550-1967.100	c4	x		n- ==
1967.920-1969.610	c2	s^cs		maybe you have to standardize this thing also .
1970.450-1974.600	c2	s^df.%--		because all the thing that you are testing use a different ==
1969.610-1970.450	c2	s^e		noise estimation .
1975.430-1975.930	c4	b		huh .
1975.490-1976.000	c3	b		huh .
1975.780-1978.860	c2	s^df		they all need some - some noise - noise spectra .
1978.860-1981.940	c2	s^df		but they use - every - all use a different one .
1976.720-1979.030	c4	s^ar|s		no | i do that two - t- - did two time .
1982.310-1983.860	c1	s		i have an idea .
1983.860-1985.620	c1	s.%--		if - if uh uh ==
1985.620-1986.500	c1	s^aa		y- - you're right .
1986.500-1987.380	c1	s		i mean each of these require this .
1987.380-2000.980	c1	qw^cs		um given that we're going to have for this test at least of - uh boundaries what if initially we start off by using known sections of nonspeech for the estimation ?
1999.540-2000.350	c4	b		uhhuh .
1999.630-2000.020	c2	b		uhhuh .
2003.140-2003.740	c1	qy^d^g^rt		right ?
2003.740-2005.860	c1	fh		s- - so e- - um ==
2003.760-2004.160	c2	b		yeah .
2004.160-2004.570	c2	b		uhhuh .
2005.860-2010.710	c1	s^df		first place i mean even if ultimately we wouldn't be given the boundaries uh this would be a good initial experiment to

				separate out the effects of things .
2010.710-2015.930	c1	qw		i mean how much is the poor you know relatively uh unhelpful result that you're getting in this or this or this ?
2015.930-2021.370	c1	qy		is due to some inherent limitation to the method for these tasks ?
2021.370-2031.420	c1	qw		and how much of it is just due to the fact that you're not accurately finding enough regions that - that are really n- - noise ?
2028.600-2029.070	c3	b		huh .
2030.230-2030.880	c4	b		uhhuh .
2030.780-2031.490	c2	b		uhhuh .
2032.080-2033.070	c1	fh		um ==
2033.070-2037.980	c1	s^df	14a	so maybe if you tested it using that you'd have more reliable stretches of nonspeech to do the estimation from .
2037.980-2042.900	c1	s	14a+	and see if that helps .
2042.880-2045.120	c4	s^bk	14b	yeah .
2045.120-2046.250	c4	s^tc		another thing is the them - the codebook .
2046.250-2047.370	c4	s^bsc		the initial codebook .
2047.370-2049.380	c4	s.%--		that maybe ==
2049.380-2050.380	c4	s		well it's too clean .
2050.380-2051.380	c4	fh		and ==
2051.240-2051.980	c1	b		uhhuh .
2051.380-2052.560	c4	s^df.%--		because it's a ==
2052.560-2053.150	c4	fh		i don't know .
2053.150-2053.740	c4	s.%--		the methods ==
2054.740-2058.370	c4	s^cs	15a	if you want you c- - i can say something about the method .
2058.420-2059.090	c1	s^aa	15b	uhhuh .
2059.380-2060.780	c4	s.%--		yeah in the ==
2065.040-2070.080	c4	s^df		because it's a little bit different of the other method .
2071.310-2072.790	c4	s.%--		well we have ==
2073.710-2088.990	c4	s		if this - if this is the noise signal uh in the log domain we have something like this .
2102.010-2103.390	c4	s		now we have something like this .
2103.390-2107.640	c4	s.%--		and the idea of these methods is

			to n- - given a um ==
2107.640-2111.900	c4	qw	how do you say ?
2108.620-2110.040	c1	b	huh huh .
2111.900-2115.240	c4	s	i will read because it's better for my english .
2116.130-2117.780	c4	%--	i- - i- - given ==
2117.780-2120.610	c4	s	is the estimate of the p d f of the noise signal .
2120.610-2131.340	c4	s	when we have a - um a statistic of the clean speech and an statistic of the noisy speech .

APPENDIX 2: UNUSED/MERGED SWBD-DAMSL TAGS

As indicated in Section 1.2, certain SWBD-DAMSL tags are not found in the MRDA tagset. Of these tags, some have been merged with other tags and others are not included in the MRDA tagset entirely. Below is a list of these tags. Each SWBD-DAMSL tag listed below is followed by a brief description indicating whether it has been merged or why it is not included in the MRDA tagset.

▪ About-communication <c>

Utterances such as "pardon me?" and "I can't hear you" that are marked with <c> in the SWBD-DAMSL tagset are considered Repetition Requests <br> in the MRDA tagset. The <br> tag is more specific in characterizing these utterances. Also, the <c> tag marks utterances such as "I heard a laugh in the background" and "I think a train went by" (Jurafsky et al. 1997). Such utterances generally do not tend to occur in the MRDA meetings. Rather than generally address communication with the <c> tag, the <br> tag is implemented for specificity.

▪ Statement-non-opinion <sd> and Statement-opinion <sv>

The <sd> and <sv> tags were quite difficult to use with the MRDA data, as their use resulted in a lack of agreement among annotators. They were eventually eliminated from the MRDA tagset and replaced with the <s> tag, which marks statements in general, without having to distinguish between "non-opinion" and "opinion." (For overt opinions, the <ba> tag is used).

▪ Open-option <oo>

This tag is no longer included in the MRDA tagset due to its redundancy with suggestions <cs>. Refer to Appendix 4 for more information.

▪ Conventional-opening <fp>

This tag is not included in MRDA tagset due to lack of use. Utterances that would be marked with this tag usually occur in pre-meeting chatter, which is marked with the <z> tag.

- ## Conventional-closing <fc>

This tag is not included in MRDA tagset due to lack of use. Utterances that would be marked with this tag usually occur in post-meeting chatter, which is marked with the <z> tag.

- ## Explicit-performative <fx>

This tag is no longer included in the MRDA tagset due to its lack of use. Refer to Appendix 4 for more information.

- ## Other-forward-function <fo>

This tag is not included in MRDA tagset due to lack of use.

- ## Yes Answers <ny>

This tag has been merged with the SWBD-DAMSL tag <aa> to form the MRDA tag <aa>.

- ## No Answers <nn>

This tag has been merged with the SWBD-DAMSL tag <ar> to form the MRDA tag <ar>.

- ## Quoted Material <q>

Due to the various DA tags quoted material within the MRDA data had the potential to receive, the use of the SWBD-DAMSL tag <q> was replaced with a convention that actually used DAs to characterize the quoted material. In doing so, more information regarding the character and function of quoted material is gained than through using a tag such as <q> to merely indicate that quoted material is present. Section 3.5 details the treatment of quoted material.

- ## Hedge <h>

This tag is not included in the MRDA tagset due to lack of use and ambiguity as to what sort of utterance would be labeled as a hedge as opposed to another label.

- **Continued from Previous Line <+>**

This tag is not included in the MRDA tagset because utterances continued from a previous line by the same speaker are given a new DA to depict the function of the continuation.

APPENDIX 3: UNIQUE MRDA TAGS

Due to the nature of the MRDA data, the SWBD-DAMSL tagset proved to be inefficient in accurately characterizing all facets of the MRDA data. Consequently, tags were created to account for areas where the SWBD-DAMSL tagset was insufficient. Below is a list of the tags that were created specifically for the MRDA data. Each tag listed below is followed by a brief description indicating why it entered the MRDA tagset.

▪ Interrupted <%->

Throughout the meetings, incomplete utterances arose on account of speakers abandoning their utterances or being interrupted. To characterize why an incomplete utterance arose, the interrupted tag was added (as the abandoned tag <%--> was already present).

▪ Topic Change <tc>

Within the MRDA data, many instances arose in which speakers attempted to change the topic. No other mechanism was present to mark such occurrences, so the <tc> tag entered the MRDA tagset to mark changes in topic.

▪ Floor Holder <fh>

The SWBD-DAMSL tagset contained the tag <h> (hold), which was also incorporated into the MRDA tagset. Utterances similar to those marked with <h> appeared mid-speech within the MRDA data. The <fh> tag was implemented to distinguish between a hold, which marks utterances in which a speaker "holds off" prior to answering a question or prior to speaking when he is expected to speak, and these mid-speech "holds.

▪ Floor Grabber <fg>

This tag entered the tagset as there were significant similarities among the means by which speakers “gained” the floor and also due to the lack of a tack to mark such instances. Speakers’ utterances often contained specific lexical items and higher

energy during these attempts to “gain” the floor. The <fg> tag entered the MRDA tagset as a means to mark such utterances.

- **Repeat <r>**

This tag entered the MRDA tagset in order to mark possible subtle changes in the manner in which a speaker repeats an utterance, whether for purposes of emphasis or in response to a repetition request.

- **Self-Correct Misspeaking <bsc>**

This tag was added to differentiate cases in which the primary speaker alone corrected his speech rather than being corrected by another speaker, which is indicated by the <bc> tag.

- **Understanding Check <bu>**

This tag entered the MRDA tagset as there seemed to be a large number of distinct cases in which a speaker wanted to check if his information was correct.

- **Defending/Explanation <df>**

This tag was added as speakers tended to defend their suggestions either immediately prior to making a suggestion or immediately after. Its usage was later expanded to include when speakers generally defended their points or offered explanations.

- **“Follow Me” <f>**

This tag was added as speakers tended to occasionally seek verification from their listeners that their utterances were understood or agreed upon.

- **Joke <j>**

This tag was added to mark utterances of humorous content and jokes, as there was previously no other means to mark such utterances.

▪ Rising Tone <rt>

Although this tag is not an actual dialog act, it was implemented to mark whether an utterance ended with a rising tone for the purpose of providing information for automatic speech recognition.

▪ Nonlabeled <z>

Certain utterances arose in the data that were intentionally not to be labeled. The <z> tag entered the MRDA tagset specifically for this purpose.

APPENDIX 4: FINAL MRDA TAGSET REVISIONS

As work on dialog act labeling progressed, the original tagset used underwent many changes and eventually evolved to the form that is presented within this guide. As most changes to the tagset occurred early on, in its final stages, the tagset underwent a scant number of changes prior to being finalized. During its final stages, a number of meetings were labeled and consequently do not reflect a few of the minute changes present within the current tagset. Those changes include the elimination of the <sj>, <fx>, and <oo> tags. Instances in which the <sj> tag was used are preserved within the data, however instances in which the <fx> and <oo> tags were used are not preserved and the data has subsequently been updated to reflect the current tagset.

▪ Subjective Statement <sj>

Originally, a distinction existed where the statement tag <s> marked objective and factual statements and the <sj> tag marked opinions and other subjective statements. The <sj> tag eventually merged with the <s> tag, as there was a lack of agreement among annotators regarding the use of the <sj> tag. The twenty-six meetings listed below currently contain the <sj> tag:

Bed003	**Bmr008**	**Bro004**
Bed004	**Bmr009**	**Bro005**
Bed009	**Bmr010**	**Bro007**
Bed010	**Bmr012**	**Bro008**
Bed011	**Bmr013**	**Bro012**
Bmr001	**Bmr014**	**Bro017**
Bmr005	**Bmr018**	**Bro018**
Bmr006	**Bmr024**	**Bro026**
Bmr007	**Bmr026**	

▪ Explicit Performative <fx>

This tag marked utterances in which a speaker made a declaration or performed some sort of act, such as the act of "firing" in saying "you're fired" and the act of "recommending" in saying "I recommend you try the other one." This tag was removed from the tagset completely due to its lack of use.

Although no examples exist in the data of the welcome tag <fw>, the welcome tag is complementary to the thanks tag <ft> and persists as a result of this relationship. The explicit performative tag lacks a complementary relationship of this sort.

▪ Open Option <oo>

This tag marked utterances in which a speaker posed multiple options. It was removed from the tagset completely due to its redundancy with suggestions <cs>.

BIBLIOGRAPHY

Jurafsky, Dan, Shriberg, Liz, and Biasca, Debra. 1997. “Switchboard SWBD-DAMSL Shallow-Discourse-Function Annotation Coders Manual, Draft 13.” Technical Report 97-02, University of Colorado, Boulder, Institute of Cognitive Science.

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge: Cambridge University Press.

INDEX OF TAGS