Lexical, Prosodic, and Syntactic Cues for Dialog Acts

Daniel Jurafsky*, Elizabeth Shriberg†, Barbara Fox*, and Traci Curl*
University of Colorado*
SRI International†

Abstract

The structure of a discourse is reflected in many aspects of its linguistic realization, including its lexical, prosodic, syntactic, and semantic nature. Multiparty dialog contains a particular kind of discourse structure, the **dialog act** (DA). Like other types of structure, the dialog act sequence of a conversation is also reflected in its lexical, prosodic, and syntactic realization. This paper presents a preliminary investigation into the realization of a particular class of dialog acts which play an essential structuring role in dialog, the **backchannels** or **acknowledgements tokens**. We discuss the lexical, prosodic, and syntactic realization of these and subsumed or related dialog acts like **continuers, assessments, yes-answers, agreements**, and **incipient-speakership**. We show that lexical knowledge plays a role in distinguishing these dialog acts, despite the widespread ambiguity of words such as *yeah*, and that prosodic knowledge plays a role in DA identification for certain DA types, while lexical cues may be sufficient for the remainder. Finally, our investigation of the syntax of assessments suggests that at least some dialog acts have a very constrained syntactic realization, a per-dialog act 'microsyntax'.

1 Introduction

The structure of a discourse is reflected in many aspects of its linguistic realization. These include 'cue phrases', words like *now* and *well* which can indicate discourse structure, as well as other lexical, prosodic, or syntactic 'discourse markers'. Multiparty dialog contains a particular kind of discourse structure, the **dialog act**, related to the speech acts of Searle (1969), the conversational moves of Carletta et al. (1997), and the adjacency pair-parts of Schegloff (1968) Sacks et al. (1974) (see also e.g. Allen and Core (1997; Nagata and Morimoto (1994)). Like other types of structure, the dialog act sequence of a conversation is also reflected in its lexical, prosodic, and syntactic realization. This paper presents a preliminary investigation into the realization of a particular class of dialog acts which play an essential structuring role in dialog, the **backchannels** or **acknowledgements tokens**. We discuss the importance of words like *yeah* as cue-phrases for dialog structure, the role of prosodic knowledge, and the constrained syntactic realization of certain dialog acts.

This is part of a larger project on automatically detecting discourse structure for speech recognition and understanding tasks, originally part of the 1997 Summer Workshop on Innovative Techniques in LVCSR at Johns Hopkins. See Jurafsky et al. (1997a) for a summary of the project and its relation to previous attempts to build stochastic models of dialog structure (e.g. Reithinger et al. (1996),Suhm and Waibel (1994),Taylor et al. (1998) and many others), Shriberg et al. (1998) for more details on the automatic use of prosodic features, Stolcke et al. (1998) for details on the machine learning architecture of the project, and Jurafsky et al. (1997a) on the applications to automatic speech recognition.

In this paper we focus on the realization of five particular dialog acts which are subsumed by or related to **backchannel** acts, utterances which give discourse-structuring feedback to the speaker. Four (**continuers, assessments, incipient speakership**, and to some extent **agreements**), are subtypes of backchannels. These four and the fifth type (**yes-answers**) overlap strongly in their lexical realization; many or all of them are realized with words like *yeah*, *okay*, *uh-huh*, or *mm-hmm*. Distinguishing true markers of agreements or factual answers from mere continuers is essential in understanding a dialog or modeling its structure. Knowing whether a speaker is trying to take the floor (**incipient speakership**) or merely passively following along (**continuers**) is essential for predictive models of speakers and dialog.

Tag	Example	Count	%
Statement	*Me, I'm in the legal department.*	72,824	36%
Continuer	*Uh-huh.*	37,096	19%
Opinion	*I think it's great*	25,197	13%
Agree/Accept	*That's exactly it.*	10,820	5%
Abandoned/Turn-Exit	*So, -/*	10,569	5%
Appreciation	*I can imagine.*	4,633	2%
Yes-No-Question	*Do you have to have any special training*	4,624	2%
Non-verbal	*<Laughter>,<Throat_clearing>*	3,548	2%
Yes answers	*Yes.*	2,934	1%
Conventional-closing	*Well, it's been nice talking to you.*	2,486	1%
Uninterpretable	*But, uh, yeah*	2,158	1%
Wh-Question	*Well, how old are you?*	1,911	1%
No answers	*No.*	1,340	1%
Response Ack	*Oh, okay.*	1,277	1%
Hedge	*I don't know if I'm making any sense*	1,182	1%
Declarative Question	*So you can afford to get a house?*	1,174	1%
Other	*Well give me a break, you know.*	1,074	1%
Backchannel-Question	*Is that right?*	1,019	1%

Table 1: *18 most frequent tags (of 42)*

2 The Tag Set and Manual Tagging

The SWBD-DAMSL dialog act tagset (Jurafsky et al., 1997b) was adapted from the DAMSL tag-set (Core and Allen, 1997), and consists of approximately 60 labels in orthogonal dimensions (so labels from different dimensions could be combined). Seven CU-Boulder linguistic graduate students labeled 1155 conversations from the Switchboard (SWBD) database (Godfrey et al., 1992) of human-to-human telephone conversations with these tags, resulting in 220 unique tags for the 205,000 SWBD utterances.

The SWBD conversations had already been hand-segmented into utterances by the Linguistic Data Consortium (Meteer and others, 1995; an utterance roughly corresponds to a sentence). Each utterance received exactly one of these 220 tags. For practical reasons, the first labeling pass was done only from text transcriptions without listening to the speech.

The average conversation consisted of 144 turns, 271 utterances, and took 28 minutes to label. The labeling agreement was 84% (κ = .80; (Carletta, 1996)). The resulting 220 tags included many which were extremely rare, making statistical analysis impossible. We thus clustered the 220 tags into 42 final tags. The 18 most frequent of these 42 tags are shown in Table 1. In the rest of this section we give longer examples of the 4 types which play a role in the rest of the paper.

A **continuer** is a short utterance which plays discourse-structuring roles like indicating that the other speaker should go on talking (Jefferson, 1984; Schegloff, 1982; Yngve, 1970). Because continuers are the most common kind of backchannel, our group and others have used the term 'backchannel' as a shorthand for 'continuer-backchannels'. For clarity in this paper we will use the term **continuer**, in order to avoid any ambiguity with the larger class of utterances which give discourse-structuring feedback to the speaker. Table 2 shows examples of continuers in the context of a Switchboard conversation.

Jefferson (1984) (see also Jefferson (1993)) noted that continuers vary along the dimension of incipient speakership; continuers which acknowledge that the other speaker still has the floor reflect 'passive recipiency', and those which indicate an intention to take the floor reflect 'preparedness to shift from recipiency to speakership'. She noted that tokens of passive recipiency are often realized as *mm-hmm*, while tokens of incipient speakership are often realized as *yeah*, or sometimes as *yes*. The example in Table 2 is one of Passive Recipiency. Table 3 shows an example of a continuer that marks **incipient speakership**. In our original coding, these were not labeled differently (tokens of passive recipiency and incipient speakership were both marked as 'backchannels'). Afterwards, we took all continuers which the speaker followed by further talk and coded them as **incipient speakership**. [1].

[1]This simple coding unfortunately misses more complex cases of incipiency, such as the speaker's next turns beginning

Table 2: *Examples: Continuers*

Spkr	Dialog Act	Utterance
B	**Statement**	*but, uh, we're to the point now where our financial income is enough that we can consider putting some away –*
A	**Continuer**	Uh-huh. /
B	**Statement**	*– for college, /*
B	**Statement**	*so we are going to be starting a regular payroll deduction –*
A	**Continuer**	Um. /
B	**Statement**	*— in the fall /*
B	**Statement**	*and then the money that I will be making this summer we'll be putting away for the college fund.*
A	**Appreciation**	Um. Sounds good.

Table 3: *Examples: Incipient Speakership.*

Spkr	Dialog Act	Utterance
B	**Wh-Question**	*Now, how long does it take for your contribution to vest?*
A	**Statement**	God, I don't know /
A	**Statement**	<laughter> It's probably a long time <laughter>.
A	**Statement**	I'm sure it's not till
A	**Statement**	like twenty-five years, thirty years.
B	**Incipient**	*Yeah, /*
B	**Statement**	*the place I work at's, health insurance is kind of expensive./*

The **yes-answer** DA (Table 4) is a subtype of the **answer** category, which includes any sort of answers to questions. **yes-answer** includes *yes*, *yeah*, *yep*, *uh-huh*, and such other variations on *yes*, when they are acting as an answer to a **Yes-No-Question**.

The various **agreements** (**accept**, **reject**, **partial accept** etc.) all mark the degree to which speaker accepts some previous proposal, plan, opinion, or statement. Because SWBD consists of free conversation and not task-oriented dialog, the majority of our tokens were **agree/accepts**, which for convenience we will refer to as **agreements**. These are used to indicate the speaker's agreement with a statement or opinion expressed by another speaker, or the acceptance of a proposal. Table 5 shows an example.

3 Lexical Cues to Dialog Act Identity

Perhaps the most studied cue for discourse structure are lexical cues, also called 'cue phrases', which are defined as follows by Hirschberg and Litman (1993): "Cue phrases are linguistic expressions such as NOW and WELL that function as explicit indicators of the structure of a discourse". This section examines the role of lexical cues in distinguishing four common DAs with considerable overlap in lexical realizations. These are **continuers**, **agreements**, **yes-answers**, and **incipient-speakership**.

What makes these four types so difficult to distinguish is that they all can be realized by common words like *uh-huh*, *yeah*, *right*, *yes*, *okay*.

But while some tokens (like *yeah*) are highly ambiguous, others, (like *uh-huh* or *okay*) are somewhat less ambiguous, occurring with different likelihoods in different DAs. This suggests a generalization of the 'cue word' hypothesis: while some utterances may be ambiguous, in general the lexical form of a DA places strong constraints on which DA the utterance can realize. Indeed, we and our colleagues as well as many other researchers working on automatic DA recognition, have found that the words and phrases in a DA were the strongest cue to its identity.

Examining the individual realization of our four DAs, we see that although the word *yeah* is highly ambiguous, in general the distribution of possible

a telling (Drummond and Hopper, 1993b)

Table 4: *Examples: yes-answer.*

Spkr	Dialog Act	Utterance
A	**Declarative-Question**	*So you can afford to get a house?*
B	**Yes-Answer**	Yeah, /
B	**Statement-Elaboration**	we'd like to do that some day. /

Table 5: *Example: Agreement*

Spkr	Dialog Act	Utterance
A	**Opinion**	*So, I, I think, if anything, it would have to be /*
A	**Opinion**	*a very close to unanimous decision. /*
B	**Agreement**	Yeah, /
B	**Agreement**	I'd agree with that. /

realizations is quite different across DAs. Table 6 shows the most common realizations.

As Table 6 shows, the Switchboard data supports Jefferson's (1984) hypothesis that *uh-huh* tends to be used for passive recipiency, while *yeah* tends to be used for incipient speakership. (Note that the transcriptions do not distinguish *mm-hm* from *uh-huh*; we refer to both of these as *uh-huh*). In fact *uh-huh* is twice as likely as *yeah* to be used as a continuer, while *yeah* is three times as likely as *uh-huh* to be used to take the floor.

Our results differ somewhat from earlier statistical investigation of incipient speakership. In their analysis of 750 acknowledge tokens from telephone conversations, Drummond and Hopper (1993a) found that *yeah* was used to initiate a turn about half the time, while *uh huh* and *mm-hm* were only used to take the floor 4% – 5% of the time. Note that in Table 6, *uh-huh* is used to take the floor 1402 times. The corpus contains a total of 15,818 tokens of *uh-huh*, of which 13,106 (11,704+1402) are used as backchannels. Thus 11% of the backchannel tokens of *uh-huh* (or alternatively 9% of the total tokens of *uh-huh*) are used to take the floor, about twice as many as in Drummond and Hopper's study. This difference could be caused by differences between SWBD and their corpora, and bears further investigation.

Drummond and Hopper (1993b) were not able to separately code **yes-answers** and **agreements**, which suggests that their study might be extended in this way. Since we did code these separately, we also checked to see what percentage of just the backchannel uses of *yeah* marked incipient speakership. We found that 41% of the backchannel uses of *yeah* were used to take the floor (4773/(4773+6961)) similar to their finding of 46%.

While *yeah* is the most common token for **continuer**, **agreement**, and **yes-answer**, the rest of the distribution is quite different. *Uh-huh* is much less common as an **yes-answer** than tokens of *yeah* or *yes* – in fact 86% of the **yes-answer** tokens contained the words *yes*, *yeah*, or *yep*, while only 14% contained *uh-huh*.

Note also that *uh-huh* is also not a good cue for agreements, only occurring 4% of the time. Tokens like *exactly* and *that's right*, on the other hand, uniquely specify agreements (among these four types). The word *no*, while not unique (it also marks incipient speakership), is a generally good discriminative cue for agreement (it is very commonly used to agree with negative statements).

We are currently investigating speaker-dependencies in the realization of these four DAs. Anecdotally we have noticed that some speakers used characteristic intonation on a particular lexical item to differentiate between its use as a **continuer** and an **agreement**, while others seemed to use one lexical item exclusively for backchannels and others for agreements.

4 Prosodic Cues to Dialog Act Identity

While lexical information is a strong cue to DA identity, prosody also clearly plays an important role. For example Hirschberg and Litman (1993) found that intonational phrasing and pitch accent play a role in disambiguating cue phrases, and hence in helping determine discourse structure.

Agreements			Continuer			Incipient Speaker			Yes-Answer		
yeah	3304	36%	uh-huh	11704	45%	yeah	4773	59%	yeah	1596	56%
right	1074	11%	yeah	6961	27%	uh-huh	1402	17%	yes	497	17%
yes	613	6%	right	2437	9%	right	603	7%	uh-huh	401	14%
that's right	553	6%	oh	974	3%	okay	243	3%	oh yeah	125	4%
no	489	5%	yes	365	1%	oh yeah	199	2%	uh yeah	50	1%
uh-huh	443	4%	oh yeah	357	1%	yes	162	2%	oh yes	31	1%
that's true	352	3%	okay	274	1%	(LAUGH) yeah	88	1%	well yeah	29	1%
exactly	299	3%	um	256	1%	oh	79	<1%	uh yes	25	<1%
oh yeah	227	2%	sure	246	<1%	sure	58	<1%	yeah (LAUGH)	24	<1%
i know	198	2%	huh-uh	241	<1%	no	49	<1%	um yeah	18	<1%
sure	95	1%	huh	217	<1%	well yeah	47	<1%	yep	18	<1%
it is	95	1%	huh	137	<1%	really	41	<1%	yes (LAUGH)	11	<1%
okay	94	1%	uh	131	<1%	huh	34	<1%			
absolutely	90	<1%	really	114	<1%	oh really	31	<1%			
i agree	73	<1%	yeah (LAUGH)	110	<1%	oh okay	31	<1%			
(LAUGH) yeah	66	<1%	oh uh-huh	102	<1%	huh-uh	27	<1%			
oh yes	58	<1%	oh okay	92	<1%	allright	25	<1%			

Table 6: Most common lexical realizations for the four DAs

Hirschberg and Litman also looked at the difference in cues between text transcriptions and complete speech.

We followed a similar line of research to examine the effect of prosody on DA identification, by studying how DA labeling is affected when labelers are able to listen to the soundfiles. As mentioned earlier, labeling had been done only from transcripts for practical reasons, since listening would have added time and resource requirements beyond what we could handle for the JHU workshop. The fourth author (an original labeler) listened to and relabeled 44 randomly selected conversations that she had previously labeled only from text. In order not to bias changes in the labeling, she was not informed of the purpose of the relabeling, other than that she should label after listening to each utterance. As in the previous labeling, the transcript and full context was available; this time, however, her originally-coded labels were also present on the transcripts. Also as previously, segmentations were not allowed to be changed; this made it feasible to match up previous and new labels. The relabeling by listening took approximately 30 minutes per conversation.

For this set of 44 conversations, 114 of the 5757 originally labeled Dialog Acts (2%) were changed. The fact that 98% of the DAs were unchanged suggests that DA labeling from text transcriptions was probably a good idea for our purposes overall. However, there were some frequent changes which were significant for certain DAs. Table 7 shows the DAs that were most affected by relabeling, and hence were presumably most ambiguous from text-alone:

Changed DA	Count	%
continuers → agreements	43/114	38%
opinions → statements	22/114	19%
statements → opinions	17/114	15%
other	32	(< 3 % each)

Table 7: DA changes in 44 conversations

The most prominent change was clearly the conversion of **continuers** to **agreements**. This accounted for 38% of the 114 changes made. While there were also a number of changes to **statements** and **opinions**, the changes to **continuers** were primary for two reasons. First, **statements** have a much higher prior probability than **continuers** or **agreements**. After normalizing the number of changes by DA prior, **continuer → agreement** changes occur for over 4% of original **continuer** labelers. In contrast, the normalized rate for the second and third most frequent types of changes were 22/989 (2%) for **opinions → statements** and 17/2147 (1%) for statements → **opinions**. Second, **continuer → agreement** changes often played a causal role in the other changes: a continuer which changed to an agreement often caused a preceding statement to be relabeled as an opinion.

There are a number of potential causes for the high rate of **continuer → agreement** changes. First, because **continuers** were more frequent and less marked than **agreements**, labelers were originally instructed to code ambiguous cases as **contin-**

uers. Second, the two codes often shared identical lexical form: as was mentioned above, while some speakers used lexical form to distinguish **agreements** from **continuers**, many others used prosody.

We did find some distinctive prosodic indicators when a continuer was relabeled as an agreement. In general, continuers are shorter in duration, less intonationally marked (lower F0, flatter, lower energy (less loud)) than agreements. There are exceptions, however. A continuer can be higher in F0, with considerable energy and duration, if it ends in a continuation rise. This has the effect of inviting the other speaker to continue, resembling question intonation for English. A high fall, on the other hand, sounds more like an agreement than a continuer.

Another important prosodic factor not reflected in the text is the latency between DAs, since pauses were not marked in the SWBD transcripts. One mark of a dispreferred response is a significant pause before speaking. Thus when listening, a DA which was marked as an **agreement** in the text could be easily heard as a **continuer** if it began with a particularly long pause. Lack of a pause, conversely, contributes to an opposite change, from **continuer** → **agreement**. The SWBD segmentation conventions placed *yeah* and *uh-huh* in separate units from the subsequent utterances. Listening, however, sometimes indicated that these *yeah*s or *uh-huh*s were followed by no discernible pause or delay, in effect "latched" onto the subsequent utterance. Taken as a single utterance, the combination of the affirmative lexical items and the other material actually indicated agreement. In the following example there is no pause between A.1 and A.2, which led to relabeling of A.1 as an **agreement**, based mainly on this latching effect and to a lesser extent on the intonation (which is probably colored by the latching, since both utterances are part of one intonation contour).

Spk	Dialog Act	Utterance
B	**Opinion**	*I don't think they even realize what's out there and to what extent.*
A	**Agree**	<Lipsmack> Yeah, /
A	**Opinion**	I'm sure a lot of them are missing those household items <laugh>.

5 Syntactic Cues

As part of our exploratory study, we have also begun to examine the syntactic realization of certain dialog acts. In particular, we have been interested in the syntactic formats found in evaluations and assessments.

Evaluations and assessments represent a subtype of what Lyons (1972) calls "ascriptive sentences" (471). Ascriptive sentences "are used...to ascribe to the referent of the subject-expression a certain property" (471). In the case of evaluations and assessments, the property being ascribed is part of the semantic field of positive-negative, good-bad. Common examples of evaluations and assessments are:

1. That's good.
2. Oh that's nice.
3. It's great.

The study of evaluations and assessments has attracted quite a bit of work in the area of Conversation Analysis. Goodwin and Goodwin (1987) provide an early description of evaluations/ assessments. Goodwin (1996:391) found that assessments often display the following format:

Pro Term + Copula + (Intensifier) + Assessment Adjective

In examining evaluations and assessments in the SWBD data, we found that this format does occur extremely frequently. But perhaps more interestingly, at least in these data we find a very strong tendency with regard to the exact lexical identity of the Pro Term (the first grammatical item in the format): that is, we found that the Pro Term is overwhelmingly "that" in the Switchboard data (out of 1150 instances with an overt subject, 922 (80%) had *that* as the subject). Moreover, in the 1150 utterances included in this study (those displaying an overt subject), intensifiers (like *very, so*) were extremely rare, occurring in only 27 instances (2%), and all involved the same two intensifiers — *really* and *pretty*. Of the 1150 utterances used as the database for this exploratory study, those utterances that showed an assessment adjective displayed a very small range of such adjectives. The entire list follows: *great, good, nice, wonderful, cool, fun, terrible, exciting, interesting, wild, scary, hilarious, neat, funny, amazing, tough, incredible, awful.*

The very strong patterning of these utterances suggests a much more restricted notion of grammatical production than linguistic theories typically propose. This result lends itself to the notion of "micro-syntax" — that is, the possibility that partic-

ular dialog acts show their own syntactic patterning and may, in fact, be the site of syntactic patterning.

6 Conclusion

This work is still preliminary, but we have some tentative conclusions. First, lexical knowledge clearly plays a role in distinguishing these five dialog acts, despite the wide-spread ambiguity of words such as *yeah*. Second, prosodic knowledge plays a role in DA identification for certain DA types, while lexical cues may be sufficient for the remainder. Finally, our investigation of the syntax of assessments suggests that at least some dialog acts have a very constrained syntactic realization, a per-dialog act 'microsyntax'.

Acknowledgments

The original Switchboard discourse-tagging which this project draws on was supported by the generosity of many: the 1997 Workshop on Innovative Techniques in LVCSR, the Center for Speech and Language Processing at Johns Hopkins University, and the NSF (via IRI-9619921 and IRI-9314967 to Elizabeth Shriberg). Special thanks to the rest of our WS97 team: Rebecca Bates, Noah Coccaro, Rachel Martin, Marie Meteer, Klaus Ries, Andreas Stolcke, Paul Taylor, and Carol Van Ess-Dykema, and to the students at Boulder who did the labeling: Debra Biasca (who managed the labelers), Marion Bond, Traci Curl, Anu Erringer, Michelle Gregory, Lori Heintzelman, Taimi Metzler, and Amma Oduro. Finally, many thanks to Susann LuperFoy, Nigel Ward, James Allen, Julia Hirschberg, and Marilyn Walker for advice on the design of the SWBD-DAMSL tag-set, and to Julia and an anonymous reviewer for Language and Speech who suggested relabeling from speech.

References

J Allen and M Core. 1997. Draft of DAMSL: Dialog act markup in several layers.

J Carletta, A Isard, S Isard, J. C Kowtko, G Doherty-Sneddon, and A. H Anderson. 1997. The reliability of a dialogue structure coding scheme. *Computational Linguistics*, 23(1):13–32.

J Carletta. 1996. Assessing agreement on classification tasks: The Kappa statistic. *Computational Linguistics*, 22(2):249–254, June.

M. G Core and J Allen. 1997. Coding dialogs with the DAMSL annotation scheme. In *AAAI Fall Symposium on Communicative Action in Humans and Machines*, MIT, Cambridge, MA, November.

K Drummond and R Hopper. 1993a. Back channels revisited: Acknowledgement tokens and speakership incipiency. *Research on Langauge and Social Interaction*, 26(2):157–177.

K Drummond and R Hopper. 1993b. Some uses of *yeah*. *Research on Langauge and Social Interaction*, 26(2):203–212.

J Godfrey, E Holliman, and J McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of ICASSP-92*, pages 517–520, San Francisco.

C Goodwin and M Goodwin. 1987. Concurrent operations on talk. *Paper in Pragmatics*, 1:1–52.

C Goodwin. 1996. Transparent vision. In *Interaction and Grammar*. Cambridge University Press, Cambridge.

J Hirschberg and D. J Litman. 1993. Empirical studies on the disambiguation of cue phrases. *Computational Linguistics*, 19(3):501–530.

G Jefferson. 1984. Notes on a systematic deployment of the acknowledgement tokens 'yeah' and 'mm hm'. *Papers in Linguistics*, (17):197–216.

G Jefferson. 1993. Caveat speaker: Preliminary notes on recipient topic-shift implicature. *Research on Langauge and Social Interaction*, 26(1):1–30. Originally published 1983.

D Jurafsky, R Bates, N Coccaro, R Martin, M Meteer, K Ries, E Shriberg, A Stolcke, P Taylor, and C Van Ess-Dykema. 1997a. Automatic detection of discourse structure for speech recognition and understanding. In *Proceedings of the 1997 IEEE Workshop on Speech Recognition and Understanding*, pages 88–95, Santa Barbara.

D Jurafsky, E Shriberg, and D Biasca. 1997b. Switchboard SWBD-DAMSL Labeling Project Coder's Manual, Draft 13. Technical Report 97-02, University of Colorado Institute of Cognitive Science. Also available as http://stripe.colorado.edu/~jurafsky/manual.august1.html.

J Lyons. 1972. Human language. In *Non-verbal Communication*. Cambridge University Press, Cambridge.

M Meteer et al. 1995. *Dysfluency Annotation Stylebook for the Switchboard Corpus*. Linguistic Data Consortium. Revised June 1995 by Ann Taylor. ftp://ftp.cis.upenn.edu/pub/treebank/swbd/doc/DFL-book.ps.gz.

M Nagata and T Morimoto. 1994. First steps toward statistical modeling of dialogue to predict the speech act type of the next utterance. *Speech Communication*, 15:193–203.

N Reithinger, R Engel, M Kipp, and M Klesen. 1996. Predicting dialogue acts for a speech-to-speech translation system. In *ICSLP-96*, pages 654–657, Philadephia.

H Sacks, E. A Schegloff, and G Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

E Schegloff. 1968. Sequencing in conversational openings. *American Anthropologist*, 70:1075–1095.

E. A Schegloff. 1982. Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. In D Tannen, editor, *Analyzing Discourse: Text and Talk*. Georgetown University Press, Washington, D.C.

J. R Searle. 1969. *Speech Acts*. Cambridge University Press, Cambridge.

E Shriberg, R Bates, P Taylor, A Stolcke, D Jurafsky, K Ries, N Coccaro, R Martin, M Meteer, and C. V Ess-Dykema. 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? *To appear in Language and Speech Special Issue on Prosody and Conversation*.

A Stolcke, E Shriberg, R Bates, N Coccaro, D Jurafsky, R Martin, M Meteer, K Ries, P Taylor, and C. V Ess-Dykema. 1998. Dialog act modeling for conversational speech. In *In Papers from the AAAI Spring Symposium on Applying Machine Learning to Discourse Processing*, pages 98–105, Menlo Park, CA. AAAI Press. Technical Report SS-98-01.

B Suhm and A Waibel. 1994. Toward better language models for spontaneous speech. In *ICSLP-94*, pages 831–834.

P Taylor, S King, S Isard, and H Wright. 1998. Intonation and dialogue context as constraints for speech recognition. *Language and Speech*. to appear.

V. H Yngve. 1970. On getting a word in edgewise. In *Papers from the 6th Regional Meeting of the Chicago Linguistics Society*, pages 567–577, Chicago.