

Representing Discourse

John W. Du Bois

Fall 2010

Linguistics Department

University of California, Santa Barbara

*NB: This document represents selected chapters and appendices from a manuscript in progress.
Please do not circulate without permission.*

©2004-2010 John W. Du Bois
dubois@linguistics.ucsb.edu

<http://www.linguistics.ucsb.edu/projects/transcription/representing>

Table of Contents

Notes

1. This is a work in progress. New material will be added during successive offerings of the course on “Discourse Transcription” (Linguistics 212, University of California, Santa Barbara). This manuscript is not currently intended for distribution beyond the participants in Linguistics 212.
2. Items listed below in *italics* are “virtual” for now.

Chapters

Part 1: Introduction

Part 2: Levels

Part 3: Categories

Part 4: Theory & Practice

Appendices & Projects

A. Appendices

Note: These page numbers are "relative" only

1.	Transcription Delicacy Hierarchy	203
2a.	Basic Transcription Symbols (by Level)	204
2b.	Basic Transcription Symbols (by Topic)	206
3.	Reference List of Discourse Transcription Symbols	208
4.	Comparison of Transcription Symbols	221
5.	Transcription Convention Updates	224
6.	Unicode Characters for Discourse Transcription	227
7.	Transcribing Procedures	237
8.	Transcription Format	244
9.	Templates	
	9.1 Transcription Template	n/a
	9.2 Intonation Unit Cue Template	n/a
10.	Ethnography	
	10.1 Ethnographic Documentation	249
	10.2 <i>[Participant Questionnaire]</i>	n/a
	10.3 <i>[Event Description]</i>	n/a
	10.4 Setting Diagram	252
	10.5 Recording Log	253
	10.6 Voice Identification Table	254
11.	Consent Form	255
12.	Computer-Assisted Transcription: Working with audio recordings (using Sound Forge)	262
13.	VoiceWalker: A Discourse Transcription Utility	265
14.	Transcriber: A Transcription-Audio Alignment Tool	276
15.	Acoustic Analysis	277
16.	Preparing a Transcription for Computer Analysis	278
17.	Using a Concordance for Discourse Research	281
18.	Resources: Access to Computational Tools for Discourse Research	293
19.	File naming conventions for media excerpts	296

P. Projects

1. Transcription	297
2. Ethnographic Recording	298
3. Developing Examples for Teaching Transcription	306
4. Intonation Unit Boundary Cues	310
5. Intonation Analysis	311
6. Folk Transcription Practices	313
7. Transcription System Comparison	316
8. Politics, Ethics, & Orthography	320
9. Concordance Analysis	321
10. Data Publication	325

[rev. 27-Sep-2009]

Part 1: Foundations

How to Learn about Spoken Discourse by Representing It

If you want to learn something about how spoken language is used—something that more and more people are taking as a central concern of linguistic research these days—there are two mediating practices you have to come to terms with. One is recording, and the other is transcription. This is true whether you record instances of language use yourself, or use someone else's recordings; and it is true whether you transcribe the recordings yourself, or use someone else's transcriptions. Even if you prefer to look at spoken language only after it has safely been rendered into writing—the spoken-language-should-be-seen-and-not-heard approach—it is still important to consider what practices, and what assumptions, went into creating the representations of spoken language that form the foundation of your future research. At the other end of the spectrum of contact with spoken language, if you prefer to directly review for yourself the sounds and sights of the original audio or video recording—perhaps because it remains your best access to the events as originally recorded—transcription still retains its central importance, to the extent that transcription mediates access to the original recording in one way or another, if only as a rough guide or index to the recorded media representation of the speech event. For all kinds of users, then, both recordings and transcriptions play a crucial role in shaping how we perceive and interpret the reality of spoken language in action.

The critical role of recording and transcription as mediating practices is grounded in the fact that both are necessarily selective. The person who wields the instruments of recording makes choices—about microphone and camera placement, for example, and these choices will shape the perceived point of view of the recording and ultimately mediate our access to the original events. In much the same way, the person who wields the pen or keyboard in transcribing makes choices about transcription that will further mediate our access to the original events. Of the many ways to go about transcribing a recording, all require selectivity, and this selectivity raises both practical and theoretical issues. Selectivity is neither to be avoided nor lamented—it's part of life as a researcher. But it must be done mindfully, in ways that serve the purposes that the transcription is meant to serve. As researchers we ask that a transcription be informative, accurate, readable, and, nowadays, searchable by computer.

Assuming that we have in our hands a reliable recording of an occasion of spoken language use, whether of naturally occurring conversation or some other kind of speaking, we face a number of decisions that need to be made about how to approach the data. Before transcribing begins, there is the initial strategic decision, at a global level, of selecting a transcription system to suit one's purposes, or designing one's own. Once the actual transcribing begins, there are numerous individual decisions to be made at the local level, as the various discourse phenomena are interpreted, categorized, and represented within the transcription. The level of informativeness should be appropriate to the purposes of the current research, which means that deciding on the appropriate degree of delicacy or granularity of transcription detail must take into account the researcher's goals. For each individual observable event on the recording, the transcriber must select the appropriate analytical category to classify the event, and then choose the appropriate convention for representing it symbolically. Making selective decisions, both at the global level of choosing a system and at the local level of choosing a category, is what it means, in effect, to transcribe.

The local level of transcribing is unavoidably dependent on decisions made at the global or strategic level of transcription system design. Transcribers face the task of creating an actual transcription of the

observable events captured in a recording. This always involves an implementation of the representational possibilities (and limitations) of the transcription system. Thus local transcription decisions are shaped by initial strategic decisions regarding transcription system design. As a rule, transcribers are successful in representing a given local phenomenon only when they, or others on whom they depend for the design of their transcription system, have appropriately provided for a suitable representation of the phenomenon within the repertoire of available notational resources. When the transcriber is also a designer of transcription systems, these activities may be worked out in tandem, but usually the transcriber is making use of analytical resources previously developed.

In general terms, a transcription system should be understood as comprising a set of transcription symbols, conventions, practices, and interpretations. While the production of a research transcription virtually always invokes a transcription system of some kind, explicitly or implicitly, this system may or may not be well-conceived and well-defined. Transcription systems normally posit a number of distinct features or categories that are to be represented, presented as a list of symbols accompanied by glosses to indicate their meanings. But more is needed in the way of conceptual infrastructure than a mere list of symbols and glosses, if a transcription system is to be taken seriously as a means of representing the richness and complexity of language in action. This book proposes to provide a detailed introduction to and explication of a system of transcription suited to a wide range of purposes, including what is in many ways the most challenging, the representation of naturally occurring conversational interaction. A discourse transcription system that can meet this challenge is likely to be able to deal with almost any kind of discourse the researcher will encounter.

[Note: This section is a revised version of a document originally prepared for a panel co-organized with Mary Bucholtz on the theme of “Transcription Issues in Current Linguistic Research”, presented at the annual meeting of the Linguistic Society of America, Albuquerque, New Mexico, on January 7, 2006.]

Representing Discourse: What, Why, How?

<i>What?</i>	<i>Why?</i>	<i>How?</i>
What does it mean to “represent discourse”?	Why represent discourse?	
What is discourse?	Why discourse?	
What is spoken discourse?	Why spoken discourse?	
What is it to represent something?	Why represent?	

1. What does it mean to represent discourse?

What is discourse?

What is it to represent something?

What, specifically, is the act of “representing discourse”?

2. Why do we want to represent discourse?

Representing (spoken) discourse is hard work

Why not just study written language? Written language is:

- plentiful
- easy to find
- easy to collect
- doesn’t seem to require representation
-it’s just there, ripe for the taking.

But with written discourse, are we missing something? In the process of producing written discourse, are there things that are eliminated or filtered out? Do these things matter?

3. We may want to look at spoken discourse for many reasons—what are they?

4. But to “look at” spoken discourse, literally, we may need to *represent* it in some way. (Do we really need to do this?) The most widely used way to represent discourse is to *transcribe* it. What is “discourse transcription”? One answer: to create a written representation of spoken discourse in order to analyze it.

5. *How* do we represent discourse? (Objectively? Selectively? Interpretively?)

Recording (e.g. electronic recording of audio/video information)

Transcription

Coding/analysis

Description/generalization

Corpus selective “representativeness”

“Oral transcription”: quotation, reported speech, represented discourse

6. How do we represent discourse? What are the means/methods/techniques that respond to our “why”?

What Is (Natural) Discourse?

Implications for Spoken Corpus Research

Corpus linguists have a strong tradition of studying discourse – or at least, of collecting instances of discourse, and then studying certain aspects of what they have collected. Most often the corpus at hand has been a collection of written texts, beginning (in the computer era) with the seminal Brown Corpus. There have been prominent exceptions to the written corpus bias, of course, notably the London-Lund Corpus of Conversational English, which stood alone as an inspiration for so long. More recently, we have seen a rising tide of interest in creating and exploiting spoken language corpora, for a wider and wider circle of languages around the world. With the exponential growth of new spoken corpus research activity there have come certain questions, questions which in principle have always been there, but which have now become more urgent. What *is* a corpus of spoken language to be? How do we *build* such a corpus? How do we *work* with it, once we have built it? Can we apply the methods that have long proven their worth in the building and exploitation of written corpora? Or do we need new or supplementary methods and perspectives, in order to engage in a truly *spoken* corpus linguistics – to attend specifically to the features that are unique to spoken language?

I suggest that in order to understand our task as builders and users of any corpus – but especially of a spoken corpus – we need to go back to the beginning and ask a very simple question: What is *discourse*? When we pose this question in the context of spoken language, we may also wish to insert one adjective: What is *natural* discourse? –because the need to come to terms with the unique characteristics and contexts of spoken language use soon enough forces us to examine our own assumptions about naturalness. When corpus-oriented linguists come to the point of seeking out spoken language, what they mostly seem to have in mind to collect is *naturally occurring* spoken discourse – whatever that might mean (and the meaning doubtless varies considerably according to the orientation of the seeker). For written language, the question of what minimum standards or qualifications a text should meet if it is to be included in a corpus can easily be overlooked, given that written texts are largely preselected and cleaned up for us through the editorial and publication process. Broadcast speech continues to lull us into complacency, because again there is a powerful institutional framework that largely preselects and refines our data for us. But as soon as we undertake to make our own recordings of raw spoken language, the language of everyday conversation and other forms of dialogic interaction, we find ourselves engaged in something new. Whether we wanted this role or not, we quickly find ourselves doing many of the things that anthropologists and sociologists traditionally do: observing, recording, trying to capture the natural way of life that people who speak to each other are engaged in, without disturbing it too much by our recorder's presence.

To truly represent spoken language, we need to consider why people speak, and then record them as they do it. In effect: To take a slice of language, we must take a slice of life. In this talk I will present extensive examples from the Santa Barbara Corpus of Spoken American English, in order to raise the question as to what natural spoken discourse is, and what a corpus seeking to represent it should look like. I will further touch on the question of what we should do with spoken language data, once we have it in our hands. Can we avoid recreating spoken language according to our image of written language? Can we see what is uniquely interesting about spoken language? Can our spoken corpus research even make us see new things when we turn back to doing written corpus research?

What, after all, is discourse? What do we hope to find in it? Natural discourse encompasses any authentic instance of language use in all its manifestations, which is to say, any organically motivated act of producing or apprehending language, via any medium, spoken, written, or otherwise, subsuming the full scope of the situated utterance in all aspects of form, meaning, and contextualization. Leaving aside artificial discourses, such as might be constructed for purposes of experimental control, we face an almost unlimited variety of genres or types of language use: newspaper editorials, novels, horoscopes, comics, boilerplate legalese, grocery lists, love letters, lectures, credos, religious rituals, roll call, sporting news, enthusiastic recitations of bad poetry, long-winded jokes, chit-chat, put-downs, the clipped commands of workplace talk. For natural discourse there can be no a priori requirement of truth, sincerity, beauty, respectability, coherence, or even grammaticality in the normative sense. Each of these uses of language is natural discourse, and each merits study on its own terms.

A corpus of discourse

1. What is discourse? What is *natural* discourse?
2. Why do we ask these questions?
3. *What do we want to know, and why?*
4. What kinds of research questions will we want to pursue? Using what data, tools, methods, theories?
5. “A text may be spoken or written, prose or verse, dialogue or monologue. It may be anything from a single proverb to a whole play, from a momentary cry for help to an all-day discussion on a committee.

A text is a unit of language in use. It is not a grammatical unit, like a clause or a sentence; and it is not defined by its size. ... **A text is not something that is like a sentence, only bigger; it is something that differs from a sentence in kind.**” (Halliday and Hasan 1976:1-2)

6. “A corpus is a collection of **naturally-occurring** language text, chosen to characterize a state or variety of a language.” (Sinclair, 1991) p. 171
7. “... a corpus is a large and principled collection of **natural** texts.” (Biber et al., 1998) p. 12

Evaluating naturalness

8. “The Switchboard component of the ANC [American National Corpus] First Release includes the transcriptions of the LDC Switchboard corpus. It consists of 2320 **spontaneous conversations** averaging 6 minutes in length...” (emphasis added)
[<http://americannationalcorpus.org/FirstRelease/contents.html#callhome>, 23-October-2003]
9. (American National Corpus, First Release: Switchboard) [timestamps for overlap etc. not given]
B: okay hi
A: hi um yeah i'd like to talk about how you dress for work and and um what do you normally what type of outfit do you normally have to wear

B: well i work in uh corporate control so we have to dress kind of nice so i usually wear skirts and sweaters in the winter time slacks i guess and in the summer just dresses
 A: um-hum
 B: we can't even well we're not even really supposed to wear jeans very often so it really doesn't vary that much from season to season since the office is kind of you know always the same temperature
 A: and is
 right right is there is there um any is there a like a code of dress where you work do they ask

[Adapted from: <http://americannationalcorpus.org/FirstRelease/encoding.html>, 23-October-2003]

10. “Activity”:

THE TOPIC IS CLOTHING. PLEASE FIND OUT HOW
 THE OTHER CALLER TYPICALLY DRESSES FOR WORK.
 HOW MUCH VARIATION IS THERE FROM DAY TO DAY?
 HOW MUCH VARIATION IS THERE FROM SEASON TO SEASON?

11. “The **striking discrepancies** in the [heavy NP] shifting rates **between the Brown corpus and the Switchboard corpus cries out for explanation**. There are several differences between the two corpora that could contribute. An obvious one is modality: Brown is written, whereas Switchboard is spontaneous speech. Writers may use HNPS [heavy NP shifting] to make their prose more easily parsable in contexts where speakers would not have the time to reflect on how alternative orderings might affect the listener. Another potentially relevant difference is **the speakers’ lack of investment in the Switchboard conversations**: they were being **paid to chat** on the phone with strangers, talking about **topics from a prescribed list**. This made for rather boring conversations, in which the speakers **may have not had much motivation** to attend to the information structure of the discourses.” (Wasow, 2002) p. 94, fn. 7

12. (*Hey Cutie Pie* SBC028: 0-36.22) +!

1	PHONE;	((RING))
2		(3.8)
3		((RING))
4		(2.0)
5	JILL;	((PICK-UP)) Hello,;
6	JEFF;	.. How’s my favorite girl in the world.
7	JILL;	= (GASP) Hey (0.3) ba:by.
8		(0.3)
9	JEFF;	Who’s —
10		(H) Who’s the girl that (0.2) I love?
11	JILL;	@@@@@@[@]
12	JEFF;	[Who’s] the girl that I’ll do anything [2for]?
13	JILL;	[2@@@]@@@@ .. [3(H)]
14	JEFF;	[3I’ll] wash her feet with my [4mouth].
15	JILL;	[4@] @@@@ @ Oh,;
16		@honey,
17		[(H) how] are you[2:].
18	JEFF;	[@@@]

19 [2after] a ten mile run.
 20 JILL; = %E::w.
 21 (0.2)
 22 @[@]
 23 JEFF; [I] would.
 24 JILL; .. @E:w.
 25 JEFF; I would.
 26 JILL; E:w.
 27 JEFF; .. @@[@ @]
 28 JILL; [@ @] @ @ @
 29 JEFF; [2 @ (H)]
 30 JILL; [2 @]
 31 @ (H) ::
 32 How are you,
 33 ho[ne]y?
 34 JEFF; [#]

13. “Interviews were usually conducted by a pair of European American fieldworkers, in most cases a male and a female. Conversations simply followed topics of interest to the subjects, such as hunting, the care of Christmas trees [the main local source of income], or mountain life – **topics within the parameters of the spontaneous natural conversation interview in sociolinguistics** (Labov 1966, Wolfram & Fasold 1974).” (Mallinson and Wolfram, 2002) p. 748-749

14. NB: “the *spontaneous natural conversation interview* in sociolinguistics”
 “*relatively* natural conversation”

15. Natural *for what interactional context?* (– interviews with university-educated strangers?)

16. Who decides whether a conversation is natural and spontaneous? By what criteria?

Non-spontaneous, naturally

17. What about *non-spontaneous* discourse? Can this be *natural*?

18. (STRANGE DREAMS)

1 DANIEL: ... (H) <READ> *Cato the younger.*
 2 *Cato the Utican.*
 3 *Marcus Portius Cato,*
 4 *ninety-five to forty-six B C. </READ>*
 5 ... (H): #My,
 6 **he didn't ... live very long.**
 ((12 LINES OMITTED))
 19 <READ> (H) *When Caesar #clearly had gained power,*
 20 *Cato committed suicide,*
 21 ... (H) *bidding his people make their peace with Caesar. </READ>*
 22 ... **Jesus.**
 23 ... (H) .. #I #don't #guess Cato the Elder would do that.

24 ... (H) (HAWK) (THROAT) (THROAT) (THROAT)
 25 ... (SPIT) (THROAT)
 26 ... (SNIFF)
 27 ... #well,
 28 ... (H) ... #I don't live in this body,
 29 I just snee:ze and cou:gh and shit out of it.
 30 .. (H) (THROAT) (Hx)

Conclusions: Natural discourse

19. Spontaneity is neither necessary nor sufficient for naturalness.
20. "Pure conversation" is a chimera, not an appropriate criterion for spoken corpus construction.
21. "What, after all, is discourse? What do we hope to find in it? **Natural discourse encompasses any authentic instance of language use in all its manifestations**, which is to say, **any organically motivated act of producing or apprehending language**, via any medium, spoken, written, or otherwise, subsuming the full scope of the situated utterance in all aspects of form, meaning, and contextualization. (Du Bois, 2003) p. 52
22. *Natural discourse* is language use—
 - which social actors participate in for their own reasons
 - which is motivated by the exigencies of social life, not by researchers' agendas
 - which has consequences in the lives of participants

References

- Biber, Douglas, Conrad, Susan, and Reppen, Randi. 1998. *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.
- Du Bois, John W. 2003. Discourse and grammar. In *The new psychology of language: Cognitive and functional approaches to language structure*, vol. 2, ed. Michael Tomasello, 47-87. Mahwah, NJ: Erlbaum.
- Mallinson, Christine, and Wolfram, Walt. 2002. Dialect accommodation in a bi-ethnic mountain enclave community: More evidence on the development of African American English. *Language in Society* 31:743-775.
- Sinclair, John M. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Wasow, Thomas. 2002. *Postverbal behavior*. Stanford, CA: CSLI Publications.

Part 2: Levels

Chapter 4**.

Delicacy

Introduction

Delicacy concerns the amount of informational detail about discourse phenomena—the fineness of discriminations of categories of events—that is represented in a given transcription. (The term *granularity* is also used for a similar concept.) It is possible to define a hierarchy of transcription features according to the level of transcription in which they should be included. A *low delicacy* (or *low granularity*, or *broad*) transcription is one which represents relatively few details about a discourse, confining itself to the essentials (e.g. the words and who spoke them). In short, a low delicacy transcription marks fewer distinctions. Relative to this, a *high delicacy* (or *high granularity*, or *narrow*) transcription marks more distinctions. What are sometimes called “broad features” or “narrow features” in a transcription are not really broad or narrow in themselves, they are simply those features which typically are represented in a certain level of transcription.

The following example illustrates a low delicacy transcription, with relatively little transcriptional information. This delicacy of transcription, which can be called *Level 1* or *Foundation*, displays just a few very basic aspects of what happened:

1. (*Hey Cutie-Pie* SBC028: 66.550-73.225)

JEFF; it's been a long time
JILL; it's been so long
I know
you missed like all the drama here
JEFF; no drama

Perhaps the most obvious point is that almost any transcription will give a representation of the *words* that were spoken. This transcription also indicates who the *speaker* of each utterance is, via a speaker label at the beginning of the line. Implicitly this notation also indicates the change of speakers, since the speaker label is written only where the speaker is different from that of the previous line. One key piece of information about how the speakers produced their words that is often overlooked, but probably should not be, concerns the basic unit of speech production. The *intonation unit*—defined as a spurt of speech uttered under a single intonation contour—is indicated by the start of a new line for each unit. Finally, this transcription incorporates information about the *sequence* of events: the temporal order in which words were spoken, intonation units produced, turns taken. The sequence of events if written according to the usual conventions, with the progression of time implicitly represented from left to right and top to bottom. While temporal sequence might seem like an obvious thing to include, almost inevitable, it is still useful to recognize that it is present in, and indeed forms an essential part of, even the most basic transcriptions.

In the next example, the same recorded stretch of interaction is represented again, but this time reflecting a slightly more delicate (narrower) transcription. This can be called *Level 2* or *Basic* transcription:

2. (*Hey Cutie-Pie* SBC028: 66.550-73.225)

1 JEFF; It's been a long time.
2 JILL; It's been so long.

3 ...
 4 I know.
 5 ...
 6 You missed like all the drama here.
 7 JEFF; No drama.

This should be considered an alternative way of representing the same recorded interaction, which is appropriate for somewhat different purposes (e.g. allowing slightly more sophisticated analysis of the interaction—though much is still missing). It incorporates all of the transcription features of the previous, lower delicacy version, and adds several more. For example, it marks the location of *pauses* (at least for silences which are more than a minimum duration), and indicates intonational *closure* at the end of the line (here the period marks each line as ending with final intonation).

The next version of the same exchange is still more delicate:

3. Delicacy A (*Hey Cutie-Pie* SBC028: 66.550-73.225)
 1 JEFF; It's been a [long time].
 2 JILL; [(H) It's been so: long.
 3 ...
 4 I know: .
 5 You missed like all the drama here.
 6 .. (H)[::
 7 JEFF; [No drama.

As always, this level of delicacy incorporates all of the transcription features of the previous two versions, while introducing several additional features. These include a distinction between short pauses (less than two tenths of a second, marked by two dots) vs. medium to long pauses (marked by three dots). It also adds indications of in-breath (lines 2 and 6). This now makes salient the marking of prosodic lengthening on the second in-breath (line 6). Moreover, once the marking of in-breath has been introduced, it becomes meaningful to show its sequential relationship to the words spoken. Since the breath in line 2 overlaps with some words in line 1, the narrower transcription must introduce some overlap markings that were not relevant in the broad transcription. The same holds for line lines 6 and 7. Note that at this level of delicacy, we have only indicated where the overlap begins, omitting to say where it ends.

The next version of this recorded segment shows a transcription delicacy which adds information about the duration of pauses, measured in seconds. It also adds an indication of where overlapping speech ends, signaled by a right-hand square bracket.

4. Delicacy A (*Hey Cutie-Pie* SBC028: 66.550-73.225)
 1 JEFF; It's been a [long time].
 2 JILL; [(H) It's been so: long.
 3 (0.2)
 4 I know: .
 5 You missed like all the drama here.
 6 .. (H)[:]
 7 JEFF; [No dr]ama.

The following transcription incorporates in addition time stamps, indicating the start time and end time for each intonation unit. The time is measured in seconds from the beginning of the recording. The two numbers are written at the beginning of each line, separated by a tab. Another tab is inserted before the speaker label.

5. Delicacy A (*Hey Cutie-Pie* SBC028: 66.550-73.225)
 66.550 67.730 JEFF; It's been a [long time].
 67.320 69.020 JILL; [(H)] It's been so: long.
 69.020 69.235 (0.2)
 69.235 70.075 I know: .
 70.075 72.155 You missed like all the drama here.
 72.155 72.940 .. (H)[:]
 72.465 73.225 JEFF; [No dr]ama.

Although this is a significantly higher delicacy than where we began, one can go much farther than this. This will become evident in the chapters to follow.

Relativity

Since the scale of delicacy is relative, and effectively continuous, the number of delicacy levels which could be distinguished in principle is limited only by the number of kinds of information that can be represented in the narrowest of transcriptions. However, the concept of delicacy is more useful—and easier to apprehend—if a relatively small set of levels is defined. The number of delicacy levels in this approach is defined, somewhat arbitrarily, as seven.

What is considered important to be included in a transcription at a given delicacy level is a theoretical issue. For some features there will be widespread agreement about the delicacy classification, while for other features the classification will be more sensitive to the specific theoretical orientation of the transcriber, or even to the particular topic under investigation at the moment. Other transcribers would no doubt define different levels of delicacy.¹ The levels defined below represent one perspective on the natural development of a transcription, from beginning stages to later more elaborated stages.

Delicacy Hierarchy and Transcription Features

The *delicacy hierarchy* presented in Table 1 (and Appendix 1) defines seven levels of delicacy. Each level of the hierarchy lists the major transcription features comprising it. The descriptive labels given to the seven delicacy levels serve as rather arbitrary mnemonics for the key categories at that level. For the first two levels, the name describes the stage of transcription (*foundation*, *basic*). For the remaining five levels, the name reflects the most numerous or distinctive type of transcription feature characteristic of that level (*boundary*, *interaction*, *intonation*, *prosody*, *multimodal*).

In each column of the delicacy table, the transcription features are organized in two groups. The first group consists of the “core” features, representing a more or less cohesive group of transcription features, which are reflected in the name for that delicacy level. The second group is the “meta” features, representing* a set of transcription features which are somewhat independent of the main focus of that delicacy level. The meta features listed under a particular delicacy level are often only loosely related the name for that name. In principle the meta features can enter into the transcription process at any time,

and they can be shifted between levels if necessary.

In fact, the entire hierarchy should be seen as simply one way of organizing the conceptualization of transcription information. It is hoped that it will prompt further consideration of alternative ways to organize the content, process, and learning of transcription and its categories.

Table 1. *Transcription Delicacy Hierarchy (Intermediate)* [Version 2.3]

		1	2	3	3+
		Foundation	Basic	Intermediate	Intermediate+
1	Core	words speakers intonation units sequence	closure/finality appeal/question pause laughter marginal words	overlap truncation/cut-off breath pause duration lag/lengthening hold/micropause utterance start latching vocalism	manner/quality vox accent non-vocal event co-action gaze gesture body
2					
3					
4					
5					
6					
7					
8					
9					
11	Meta		comments unintelligible uncertain pseudograph header	index recording alignment timestamp	
12					
13					
14					
15					

INSERT TABLE 1 ABOUT HERE

["Table 1. Delicacy Hierarchy for Discourse Transcription"; cf. Appendix 1.]

Appendix 1a

Transcription Delicacy Hierarchy

[Version 2.3]

Table 1. *Transcription Delicacy Hierarchy for Discourse Transcription*

	Level	1	2	3	4	5	6	7	
	Type	Foundation	Basic	Boundary	Interaction	Intonation	Prosody	Multimodal	
1	Core	words	closure/finality	overlap	vocalism	accent	declination unit	audio waveform	1
2		speakers	appeal/question	truncation/cut-off	manner/quality	appeal onset	prosodic sentence	timeline/score	2
3		intonation units	pause	breath	vox	onset/head/tail	paratone	pitch curve	3
4		sequence	laughter	pause duration	affect	tone	isotony/tone concord	amplitude curve	4
5			marginal words	lag/lengthening	non-vocal event	tune/contour	boundary attenuation	spectrogram	5
6				hold/micropause	co-action	chant/stylization	rush/anacrusis	energy curve	6
7				utterance start	gaze	terminal pitch	stress	duration curve	7
8				latching	gesture	key/baseline	accent unit	intonation stylization	8
9				linking	body	pitch reset	isochrony	indexed semiosis	9
10				discontinuous IU	event duration	booster	register/range	indexed context	10
11	Meta		comments	index	non-interactive source	code-switch	phonetic variation	transcription tool link	11
12			unintelligible	recording alignment	participation frame	disfluency/repair	phonetic segments	annotation graph	12
13			uncertain	timestamp	recipient/addressee	pause attribution	segment release	coding links	13
14			pseudograph		participant gloss	turn/backchannel	pre-phonation	participant database	14
15			header		alternative hearing			event database	15

[revised 26 September 2010]

Discussion

Delicacy defines a sequence or hierarchy of levels, which can be viewed from three distinct perspectives:

- (1) *Transcription content*: hierarchy as an ordered scale of alternative levels of detail which is to be represented in the completed transcription
- (2) *Transcribing sequence*: hierarchy as successive stages in the development of a given transcription over time
- (3) *Transcription learning*: hierarchy as successive stages in the process of learning how to transcribe

The first perspective is more relevant for relatively broad or low delicacy transcriptions, in which a decision is made to limit the amount of detail which will be transcribed to a certain level of delicacy. The second perspective is relevant for narrower or high delicacy transcriptions, in which additional levels of detail are likely to be added gradually over time, as successive passes are made through the recording, each time incorporating new kinds of more detailed transcription information. Thus the transcription gradually tends to move to higher and higher levels of delicacy. It must be said that actual transcription practice, and actual transcriptions, need not fall neatly within the seven discrete levels. As the transcriber adds more detail to a transcription, a more gradual transition between levels of delicacy is to be expected, perhaps with one or two new features at a time being introduced. And as the transcriber gains in experience, features will tend to be written down whenever they are noticed, even if they belong to a delicacy level which is several notches higher than the current stage of the transcription in progress. Still, the concept of delicacy level as transcription stage remains useful as a rough indication of the general direction of development of a transcription, as it is elaborated and unfolds over time. The third perspective, that of the learning process, reflects the fact that some transcription categories are easier to learn, while others require more knowledge, sophistication, or experience. Some of the narrower features thus reflect the need to build on the prior mastery of the more foundational categories. To a large extent these three perspectives converge on the same ordering of transcription features into delicacy levels. However, the most effective transcription sequence is sometimes different from the most effective learning sequence. In such cases the ordering of features represents a compromise between the two. In any case, as they gain experience, transcribers will work out their own preferred transcription sequence, and may wish to reformulate the Transcription Delicacy Hierarchy to meet their needs.

Going Deeper

As a transcriber works with a recording it often becomes necessary to modify the delicacy level of the transcription. Typically the transcription will move toward increased delicacy. For example, an initial rough draft of a transcription is often superseded by subsequent drafts which elaborate a more delicate rendering of the recording. Transcriptions sometimes need to be modified in the direction of decreased delicacy, as when the presentation of a transcribed example is simplified for a non-specialist audience. (In such cases the narrower transcription would not be discarded, of course, but is kept for use by the researcher and other specialist users.) For these reasons, it is important that the transcription system should allow for a relatively painless and straightforward transition from one delicacy level to another {Du Bois, 1991 #294}. For example, the basic meaning of the symbol for any low delicacy transcription

category should remain constant even when further categories from a higher level of delicacy are added to the transcription, to the extent that this is possible. But inevitably the delicacy level of a given transcription will strongly shape the kinds of interpretations that are accessible to the user of that transcription, to the point of precluding certain types of understandings and analyses if the transcription is very broad. (This can be seen by comparing the examples of low vs. high delicacy transcription cited above, each of which gives a different impression of what happened in the transcribed interaction.) A low delicacy or broad transcription can still be quite valuable, however, if it is paired with simultaneous access to the original recording. A broad transcription which is accompanied by the playback of the corresponding video or audio recording can allow even non-specialist audiences to follow some of the more complex aspects of interaction. As part of Computer-Assisted Transcription (CAT) and the resulting multimodal transcriptions (or MAP. Media-Aligned (re)Presentation) a relatively broad transcription can allow specialist researchers to navigate through the data, perhaps selecting a portion for more intensive analysis, and a further transcription to a narrower delicacy level. In sum, broad and narrow transcriptions both have their uses.

Delicacy and Significance

It should be noted that delicacy levels do not represent degrees of importance of the transcription features they subsume. Some features are necessarily part of the most basic levels of transcription because they are needed to flesh out the skeleton of the interaction – for example, the words spoken. Other quite important features, including much of the prosody, are most naturally transcribed only after the basic outlines of the talk has been laid out. In a study of intonation, for example, the recurrent intonation contours may be considered more important than the particular words spoken, in some sense, and yet transcription of the words is in many cases likely to be done first. In this light, the levels of delicacy can be taken as a guide to the sequential practice of transcribing, as much as to levels of informational detail.

Delicacy and Accuracy

Delicacy is independent of accuracy. A broad transcription can be as accurate and rigorous as a narrow one, each for what it purports to represent. The transcriber chooses a delicacy level appropriate to the task at hand, which implies a commitment that a certain set of features will be consistently represented in the transcription. The transcriber then presumably strives for accuracy with respect to the specified standard, that is, writes a symbol token representing a specified feature whenever it occurs.

Delicacy and Commitment

The delicacy hierarchy can be thought of as offering a range of transcription alternative which are not merely options for transcribers, but also imply a certain kind of commitment once the delicacy level is chosen. Consistency in the commitment to delicacy is important. If a transcription is designated as representing a certain delicacy level, or, equivalently, purports to represent a certain set of features, its users should reasonably be able to expect that if a symbol is absent from a given point in the transcription, then the corresponding feature is absent as well. On the other hand, if the delicacy level claimed is low, this effectively implies that for transcription features belonging to a higher level, no representation was attempted. In this case the absence of a symbol does not imply anything about the absence of the corresponding feature from the discourse. (This is in effect the Gricean notion of a scalar implicature, as applied to the communicative practice of transcription.)

It can readily be seen that it is important for the users of a transcription to know what the claimed delicacy level is, if they wish to draw conclusions about the presence or absence of particular discourse

phenomenon. Unfortunately, most transcriptions do not provide this information in any explicit form. To be sure, an experienced transcriber who has access to a few pages of the transcription in question can often deduce what the general delicacy level is, and what categories are apparently being represented, by noticing which symbols are included and which are not. But it would be helpful to the reader to just tell them the transcription delicacy level, rather than making them guess.

In effect, to mark a given transcription feature once becomes a commitment—an implied promise—to mark that feature consistently whenever it comes up, throughout the transcription. This is another reason for transcribers to think carefully about what transcription delicacy they want to commit to. Once a delicacy level is chosen, and the set of transcription features specified, the transcriber is implicitly committed to represent those features consistently and accurately throughout.

Delicacy and Reality

The delicacy of a transcription can affect how the reader understands the reality of what happened. This is obviously a sensitive issue, and needs to be considered carefully when transcribing. It is just as important when reading transcriptions created by other people. For example, the following excerpt illustrates a fairly narrow transcription (Level 4+). (This includes some symbols for details that will be introduced in later chapters, but which should not distract from the point being made here.) Notice the presence of vocalisms in line 25, and the pauses which precede and follow it (lines 24 and 26):

(Strange Dreams)

23	#It's like it's the ^middle of the damn `day.
24	(2.0)
25	(GULP) (TSK)
26	(3.4)
27	`Needed that (Hx).
28	(2.7)
29	(H):: (1.3) ^Hm_(Hx)::
30	#@Hm.

But suppose this event is represented via a broader transcription, such as e.g. Level 3. At this level, the convention is to not write vocalisms. Once the vocalism disappears, the two pauses on either side of it seemingly merge into one long pause—according to the following representation:

(Strange Dreams)

23	#It's like it's the middle of the damn day.
24-26	(5.4)
27	Needed that.
28	(2.7)
29	Hm.
30	#@Hm.

So, what really happened just prior to the speaker's utterance of "Needed that" in line 27? There is a pause, to be sure, but is it a pause lasting 5.4 seconds, or only 3.4 seconds? The answer seems to depend on the delicacy of the transcription. This may seem troubling, even problematic. But what it reminds us is that we need to always approach transcriptions as representations, and understand that the events they purport to document can only be understood in relation to the conventions—and commitments—of

representation that have been chosen. It becomes clear that one key factor in the choice of representation is the level of delicacy. It is incumbent on both producers and interpreters of transcriptions to take into account the issue of transcription delicacy, as an essential guide to both creating and reading transcriptions. Understanding a transcription depends critically on knowing what delicacy conventions—and commitments—are implicit in the representation at hand.

[Rev. 12-Oct-2009]

Delicacy Level Transitions: *Level 0-4*

Objective

This handout explores the notion of *levels of delicacy*, by examining transitions from lower to higher levels. Transitions between the following levels of delicacy are illustrated:

- Level 0: *Lexical*
- Level 1: *Foundation*
- Level 2: *Basic*
- Level 3: *Boundary*
- Level 4: *Interaction*

Examples

For the example given, the same passage of discourse is transcribed at several different levels of delicacy. Features which are newly introduced at each level are indicated in boldface.

23. *Level 0: Lexical* (Risk SBC024: 299.283-310.952)

look at you being smart I'm not smart you're stupid don't call me stupid alright

24. *Level 1: Foundation*

1	JENNIFER;	look at you being smart
2	DAN;	I'm not smart
3	JENNIFER;	you're stupid
4	DAN;	don't call me stupid
5	JENNIFER;	alright

25. *Level 2: Basic*

1	JENNIFER;	Look at you being smart.
2		...
3	DAN;	@
4		...
5		I'm not smart?
6		...
7	JENNIFER;	You're stupid.
8		...
9	DAN;	Don't call me stupid.
10	JENNIFER;	Mm.
11		...
12		Alright.

26. *Level 3: Boundary*

1 JENNIFER; Look at you being smart.
 2 (1.0)
 3 DAN; (H) @
 4 (0.7)
 5 I'm not smart?..
 6 (0.3)
 7 JENNIFER; You're stupid.
 8 (0.9)
 9 DAN; Don't call me stupid.
 10 JENNIFER; Mm.
 11 (3.4)
 12 Alright.

27. *Level 4: Interaction*

1 JENNIFER; <VOX> Look at you being smart.
 2 (1.0)
 3 DAN; (H) @
 4 (0.7)
 5 I'm not smart?
 6 (0.3)
 7 JENNIFER; You're stupid </VOX> .
 8 (0.9)
 9 DAN; Don't call me stupid.
 10 JENNIFER; Mm.
 11 (1.3)
 12 (KISS)
 13 (2.1)
 14 Alright.

[rev. 24-Aug-2008]

What Is To Be Transcribed?

Some questions

You are about to hear some recordings of naturally occurring conversational interactions. In order to help us make sense of what is happening in these recordings, we would like to have a way of *representing the discourse in writing*. That is, we would like to know how to *transcribe* the conversation. This will make it easier to talk about the conversation: to “slow down” the flow of time so we can look at the interaction more closely, and analyze what is happening in it in more detail.

But as we observe any recorded conversation, before we can even begin transcribing, there are a number of questions that arise:

1. What kinds of things are *happening* in this conversation?
2. In what sense are the participants *interacting*, that is, *affecting each other* by their actions?
3. How do the participants manage to do what they do in this conversation? What are the participants doing with language and interaction, in order to accomplish what they do?

These are big questions, and we won't try to answer them fully today. (And there are many other questions we could ask, some of which are even more fundamental than those just cited.) But we can formulate some basic questions that will help us pursue some useful answers:

4. What are some of the specific indications, which you can observe within the recorded interaction, that show that the participants are in fact interacting with and responding to each other?
5. What kind of a list can you come up with of things that participants do which affect the course of the interaction?
6. What aspects of the speaking and interacting do we want to be able to represent?
7. Given a list of things that participants do in conversational interaction, can we group them into coherent categories?
8. How can we create a transcription that will help us to see and hear how the participants are interacting? In short: What is to be transcribed?

Some recordings

As a way of starting to think about these questions, we will listen to some recordings, think about what's happening in them, and talk about it in groups. You should take notes on paper, trying to represent what is going on in the recorded interaction. (You may consider this practice to be a very informal, ad hoc sort of transcription.)

9. Use whatever means of representing the discourse you find useful—invent your own as you go along.
10. You can write words, symbols, diagrams, pictures, descriptions, or anything you want.
11. Think about what kinds of things you would find it useful to be able to represent: What kinds of things do you need symbols for?

As you listen to the recordings, begin by keeping in mind questions 4 and 5 above. Questions 6-8 will be taken up in group discussion, after we have finished listening to and transcribing the recordings.

Lexical transcriptions. Because we want to devote most of our learning time to working on the more challenging aspects of discourse transcription, we will sometimes take a little shortcut regarding the words themselves. The shortcut is that I will supply a very rough transcription that indicates only what words were spoken, and in what order they were spoken, but which doesn't tell you anything about how they were said, or even who said them. These can be called “lexical” transcriptions. It should be kept in mind that a lexical transcription leaves out a very large part of what fills talk with meaning, motivation,

emotion, and interactional significance. In fact, it leaves out most of what's interesting for discourse transcription.

Please do NOT treat the lexical transcription as a *starting point* for your own transcription, but only as a reference point for help in hearing the words. You should do your own transcription on a separate piece of paper, using your own ideas about how to transcribe. You should refer to the lexical transcriptions below only after you've started your own transcription, if you need some help in getting the English words.

1. [sounds]
2. laughing
3. I'm laughing
4. hm
5. a child that really got me grounded
6. course that may be what happens prior to the big the big nothing so why are you reading a book about dying

Transcription Level 1: *Foundation*

Objective

The objective at this level is to learn how to begin transcribing a recorded interaction. Transcribing at this stage may be termed Level 1 transcription, representing a “*Foundation*” level of transcription of spoken discourse. It is so named because it constitutes the foundation for building a well-structured representation of the discourse. For most transcribers it will serve simply as a preliminary stage in the process of transcription, which will subsequently be enriched by the addition of more detailed features.

Categories

Level 1 transcription is a very basic form of transcription that seeks to represent only some of the more essential features of the recorded interaction, such as the words spoken, the order in which they were spoken, and who spoke them. These are the transcription categories that are needed to “flesh out” a rough sketch of the basic overall organization of the speech event. Typically, a Level 1 transcription is only a first stage on the way to the production of a more detailed transcription. For the most part, categories included at this level are those which virtually all research practices would consider essential for even the most limited transcription.

The one exception is a relatively challenging feature whose inclusion at the most fundamental level of transcription level is not widespread and is even, perhaps, controversial. This is the intonation unit. Reasons for including the intonation unit at this level are two-fold. First, on a theoretical level the intonation unit represents a fundamental organizing principle of spoken language. Intonation units are important because they give structure to spoken discourse, with broad implications for aspects of language extending from phonetics and syntax to cognition and interaction. Second, on a practical level it turns out to be easier to organize the transcribing process, and the transcription itself, if the identification of intonation units is addressed right from the beginning. The intonation units define the skeleton of the transcription as it emerges over time. At the Foundation level of transcription, what is needed is at least a rough initial approximation of the intonation unit boundaries, which may be further refined at higher levels of transcription. (Among other things, higher levels of transcription will include attention to additional transcription categories which may help in identifying some of the less obvious forms of intonation unit boundary.)

Conventions

The following conventions are those used for the transcription categories in Level 1 (“*Foundation*”) transcription, as described in the Du Bois “Transcription Delicacy Hierarchy”.

MEANING	SYMBOL	COMMENTS
(1) words	<u>SPACE</u>	space before and after marks word
(2) intonation unit	<u>LINE</u>	one line for each Intonation Unit
(3) speaker label	JILL;	full first name in all capitals, plus semicolon

Implicit representations. Beyond what is indicated explicitly by the above symbols, several kinds of information are provided implicitly. This kind of information is automatically incorporated via the general transcription conventions used, without having to write it using overt symbols. Most of the implicit representation conventions may seem fairly obvious, perhaps too obvious to bother mentioning -- but it is still worth taking note of them. For example:

1. *word sequence* -- the sequence of words as they are spoken -- is indicated in the conventional manner, using the conventional order of words. For many languages, including English, the conventional order of words in a line of print is left to right; for others, including Arabic and Hebrew, the conventional order is right to left.
2. *unit sequence* -- the sequence of production of units, such as intonation units-- is indicated in the conventional manner, using the conventional order of lines on a page. Here sequence is conveyed iconically by vertical space on the page or screen, with the direction from top to bottom corresponding in most writing systems to earlier to later.
3. *speaker change* -- the shift from one speaker to a different speaker -- is marked implicitly by the beginning of a new line, and also by the introduction of a new speaker label.
4. *turn* -- when a new speaker takes the floor, beginning their turn at speaking -- is marked implicitly by the introduction of a new speaker label at its beginning.²
5. *turn sequence* is implicitly marked by the speaker labels at each change of speaker plus the top to bottom order of the speaker's utterances on the page.

Exercise A

Objectives

1. This session introduces the practice of Level 1 Transcription, also known as “Foundation” transcription.
2. At Level 1 the focus is on very basic aspects of transcription. This includes the words spoken, in what sequence they were spoken, who spoke them, and the intonation units in which the words were produced.
3. Of the categories introduced at Level 1, only one represents a significant challenge: the intonation unit. This initial exercise will be limited to fairly simple cases of intonation units. (We will take up more complex instances of intonation units in a separate exercise.)

Task

1. Your task is to create a “Level 1” (Foundation) transcription . At this level you should indicate the words, the speaker labels, and the intonation unit boundaries (using a new line for each intonation unit).
2. If you need help in identifying which words were spoken, you may refer to the “Level 0” (“Lexical”) transcriptions provided below as a starting point. These indicate just the words in sequence, without any additional information.
3. As a first step, simply indicate where you hear an intonation unit boundary. Using the Level 0 transcription provided below, mark a vertical stroke between words where you hear the boundary.
4. Note that the Level 0 or “Lexical” level (containing just words and word sequence) is a somewhat artificial level, being used primarily for the creation of exercises for learning transcription. For pedagogical purposes it may be useful to provide a Lexical transcription, in order to relieve students from having to spend time identifying what words are being uttered, allowing them to concentrate on the main transcription categories being taught. It should be kept in mind, however, that in normal transcription practice there is usually no stage at which one would produce a purely lexical transcription. Rather, a more useful beginning stage in actual

transcription practice is represented in the features specified for Level 1 (“Foundation”) transcription.

Starting Point: Level 0

1. (*Risk* SBC024: 299.283-310.952)

((1st speaker = JENNIFER; 2nd = DAN))

look at you being smart I’m not smart you’re stupid don’t call me stupid alright

2. (*Zero Equals Zero* SBC009: 731.70-744.11)

((1st speaker = KATHY; 2nd = NATHAN))

I don’t know this one so you don’t know how to do this one so we in trouble well you apparently knew how to do it did I get it right well you didn’t get the whole thing right

3. ((*Household* SBC019: 105.220-113.674))

((1st spkr= FRANK; 2nd = BRETT; 3rd = MELISSA))

get your pen back from that yeah it’s erasable and I am not marking on it I don’t care if it’s erasable don’t touch it I didn’t I know don’t that’s actually very good lettering sir I know

Target: Level 1

4. (=1) (*Risk* SBC024: 299.283-310.952)

1	JENNIFER;	look at you being smart
2	DAN;	I’m not smart
3	JENNIFER;	you’re stupid
4	DAN;	don’t call me stupid
5	JENNIFER;	alright

5. (=2) (*Zero Equals Zero* SBC009: 731.70-744.11)

1	KATHY;	I don’t know this one so
2	NATHAN;	you don’t know how to do this one
3		so we in trouble
4	KATHY;	well you apparently knew how to do it
5	NATHAN;	did I get it right
6	KATHY;	well you didn’t get the whole thing right

6. (=3) ((*Doesn't Work in this Household* SBC019: 105.220-113.674))

1 FRANK; get your pen back from that
 2 BRETT; yeah
 3 MELISSA; it's erasable
 4 and I am not marking on it
 5 BRETT; I don't care if it's erasable
 6 don't touch it
 7 MELISSA; I didn't
 8 BRETT; I know
 9 don't
 10 MELISSA; that's actually very good lettering sir
 11 BRETT; I know

[rev. 24-Oct-2008]

Exercise B

Starting Point: Level 0

7. ((*A Tree's Life* SBC007: 290.20-308.85)) ((1st speaker = ALICE; 2nd = MARY))

they're never gonna have hard times hard times do train you yep they do like I came over here to
 work with Danae which is what I'm going to do

8. ((*A Tree's Life* SBC007: 711.13-729.96)) ((1st speaker = ALICE; 2nd = MARY))

where'd you go to get em you know where Sarah and Arvela live mhm just around the corner
 remember that first cattle guard you go over unhunh I didn't even go over that you mean kinda like
 by the by the tunnel right below the tunnel oh

9. ((*A Tree's Life* SBC007: 736.66-745.62)) ((1st speaker = ALICE; 2nd = MARY))

did you get grandma a tree too hunh-unh does she already have one hm-m that pickup could only
 hold like three mm

Target: Level 1

10. (=1) ((*A Tree's Life* SBC007: 290.20-308.85))

1 ALICE; they're never gonna have hard times
 2 MARY; hard times do train you
 3 ALICE; yep
 4 MARY; they do
 5 ALICE; like I came over here to work with Danae
 6 which is what I'm going to do

11. (=2) ((*A Tree's Life* SBC007: 711.13-729.96))

1 ALICE; where'd you go
 2 to get em
 3 MARY; you know where Sarah and Arvela live
 4 ALICE; mhm
 5 MARY; just around the corner
 6 remember that first cattle guard you go over
 7 ALICE; unhunh
 8 MARY; I didn't even go over that
 9 ALICE; you mean
 10 kinda like by the
 11 by the tunnel
 12 MARY; right below the tunnel
 13 ALICE; oh

12. (=3) ((*A Tree's Life* SBC007: 736.66-745.62))

1 ALICE; did you get grandma a tree too
 2 MARY; hunh-unh
 3 ALICE; does she already have one
 4 MARY; hm-m
 5 that pickup could only hold like three
 6 ALICE; mm

[rev. 24-Oct-2008]

Chapter **X: Transcription Level 2: Basic

Objective

The objective at this level is to learn how to create a more informative, but still basic, transcription. The practice brings new levels of detail to the transcription, increasing its delicacy from a Level 1 or “Preliminary” transcription to a Level 2 or “Basic” transcription of spoken discourse.

Core Categories

Categories: *boundary intonation, end tone, closure, appeal, pause, laughter, marginal words*

The “core” set of transcription categories newly introduced at Level 2 includes pause, which is the noticeable non-use of the vocal tract, as well as one very noticeable but non-verbal use of the human vocal-tract, namely laughter. Along with the words spoken (already introduced at Level 1), laughter is perceived as so salient and relevant to human interaction leaving it out of a transcription would be felt as omitting the obvious, not to say the essential. In addition to these transcription categories, the transcriber should make sure that any hesitation words (*uh, um, etc.*) are marked (if not already introduced at the Level 1 stage of transcription).

The following symbol conventions are those used for the transcription categories in Level 2 (“Basic”) transcription, as described in the “Transcription Delicacy Hierarchy” (Du Bois 2005). Level 2 transcription also subsumes all transcription categories from Level 1 (including words, speaker attributions, and intonation units).

MEANING	SYMBOL	COMMENTS
End Tone		
13. continuing	,	intonational category marks continuity (comma)
14. final	.	intonational category marks finality (period)
15. appeal	?	intonational category marks appeal (question)
16. utterance start	<u>Capital initial</u>	Capitalize when speaker begins a new utterance
General		
17. pause, untimed	...	0.2 seconds or more
18. laughter	@	one symbol per pulse or particle of laughter
19. laughing word	@you’re @kidding	laugh symbol marks laughter during word
20. hesitation words, etc.	uh, um, mm, etc.	a.k.a. “filled pause”

Intonational meaning: End tones

Intonation is meaningful for its users. It is deployed for communicative purposes by speakers, and interpreted as significant by listeners. One way to describe the meaning of intonation is in the same terms as for any other meaningful element in language: as a linguistic *sign*. In general, the linguistic sign unites two aspects: *form* and *meaning* (or *signifier* and *signified*, in Saussure’s terms).

In the context of intonation, the *form* concerns the recognizable, audible character of the intonational signal. For example, a given intonational sign (or intonational *morpheme*) may be signaled by a particular sequence of tones, or a class of intonational contours. The task of identifying the relevant

intonation contours cannot be considered a purely acoustic one, but must consider both aspects of the intonational sign, form and meaning, in relation to each other.

While intonational signs must be recognized to convey *meaning*, these meanings are not necessarily as concrete as the meanings associated with the words linguists cite as typical signs or morphemes, works like *tree*, *duckling*, *farmer*, or *salami*. Instead, they tend to be rather abstract and even elusive in their meaning. Often an intonational sign will point, in a very general way, to some aspect of the informational or interactional context, including the cognitive and social relations between the speaker and addressee.

The intonational categories described as continuing, final, and appeal can all be thought of as intonational signs, each representing a distinctive combination of prosodic form and indexical meaning:

<i>Continuing</i>	Form	mid level, mid rise, mid fall
	Meaning	'I'm not finished with my current discourse activity'

<i>Final</i>	Form	low fall (fall to bottom of speaker's pitch range)
	Meaning	'I'm finished with my current discourse activity'

<i>Appeal</i>	Form	high rise
	Meaning	'I'm seeking an (implicit) response from you'

Each intonational morpheme is distinguished from the others by its form, that is, by the intonational contours that signal it. Note that it is not enough to describe the intonation for any category as simply 'rising' or 'falling'; the contours are more subtle than this, and require a more precise description. (Ultimately, these intonational descriptions will need to be made more precise; but they are sufficient for our purposes at this stage.)

The meanings are very general, even vague – but this may be just what speakers need from them. If speakers have only a small number of end tones (as seems to be the case), it is important that each end tone should prove itself useful in a wide variety of contexts. Thus the messages associated with intonational contours tend to quite broad but also highly versatile, adaptable in useful ways to a variety of contexts. According to the present analysis, both *continuing* and *final* intonation contours make reference to 'my current discourse activity', without specifying what it is. While this might seem too vague to be of use, it is precisely its vagueness that makes it so useful in so many contexts. The speaker relies on the hearer's capacity to infer what is the most salient current discourse activity (launching into a long utterance or turn, completing a point in an argument, asking for assistance), and the intonation functions to index this activity and provide a commentary on what its current completion status is. In effect it signals to the hearer what role she expected play in the ensuing moments, with respect to the speaker's turn: to keep listening, to take the next turn, to provide information or assistance, and so on. Despite its rather abstract meaning – or perhaps because of it – the end tone is well suited to make a critical contribution, in conjunction with other components of meaning supplied by the words, grammar, and context, to the hearer's task of constructing an on-line interpretation of the on-going utterance – an interpretation that is relevant to the situation and appropriate to its intended function.

Boundary Intonation Rules: Closure/Finality

Objective

A key part of the representation of spoken discourse involves characterizing that nature of the intonation at each Intonation Unit boundary. For this, the transcriber must classify the intonation as belonging to one or another category of boundary intonation (a.k.a. closure, transitional continuity).

Procedure

To analyze the intonation that is heard at the end of an Intonation Unit, there are four questions to consider:

1. *Is the unit wordless (consisting only of laugh, breath, pause, etc., but no words)?*

Yes → -INTONATION (non-intonational unit)
No → +INTONATION

If -INTONATION → Stop
If +INTONATION → Continue choosing boundary features

2. *Is the intonation contour heard as truncated (i.e. the projected contour is abandoned)?*

Yes → +TRUNCATED
No → -TRUNCATED

If +TRUNCATED → Stop
If -TRUNCATED → Continue choosing boundary features

3. *Is the intonation contour heard as final?*

Yes → +FINAL
No → -FINAL (= "CONTINUING")

4. *Is the intonation contour functionally interpreted as an appeal?*

Yes → +APPEAL
No → -APPEAL

Symbolization

Once the above criteria have been used to select the intonational features, the resulting complex of features is translated into unique symbols as in Table 1.

Table 1. Feature analysis of categories for boundary intonation (transitional continuity)

Symbol	Meaning	Features	Notation
.	final	+FINAL -APPEAL	period
,	continuing	-FINAL -APPEAL	comma
?.	final appeal	+FINAL +APPEAL	question mark
?,	continuing appeal	-FINAL +APPEAL	question mark followed by comma
—	truncated intonation unit	+TRUNCATED	em dash
	non-intonational unit	-INTONATION	no symbol (for wordless IUs only)

[rev. 14-Oct-2008]

Pause

Pause can be defined as the noticeable absence of speech (or other vocal tract activity) by any of the ratified participants in the current communicative interaction. As such, pause is a social fact, not a merely physical one.

Though pause is typically marked by silence, the two cannot be equated: not all silences are pauses. For example, the brief silence introduced into the middle of a word by a voiceless stop (such as *p*, *t*, or *k*) is not ordinarily interpreted by participants as a pause, even if it is considerably lengthened.

Conversely, not all pauses are silent. A speaker may make a great deal of noise by some means other than speech (for example, by the activity of preparing food), and yet still be interpreted as pausing in mid-sentence. Or in a crowded restaurant, there may be many speakers at nearby tables who fill the background of one's recording with speech sounds, without, however, being taken to be participants in the current conversation. It is the absence of speech by ratified co-participants in the current discourse, not the absence of sound, that is interpreted by participants as a pause.

At this transcription level, the location of pauses is indicated, but their duration is not timed. (At the next higher level, Level 3, the duration of pauses will be measured in seconds.)

Laughter

Laughter is a very important component of many occasions when people interact, whether formally in a performance context or informally in a casual conversation among friends or acquaintances. Whenever laughter does occur in a recorded event, to leave it out of a transcription would leave a serious gap in the record, making it difficult to interpret why the participants react the way they do. A words-only transcription would be deficient for any situation when laughter plays a significant role—as it usually does whenever it occurs. Thus it is crucial to have an effective way of representing laughter.

One way to represent laughter at a basic delicacy level is to just write (LAUGHTER). This should be accompanied by an indication of who is doing the laughing, using the standard speaker attribution label. This way of transcribing laughter is minimally informative, as it ignores details such as how long the person laughed, or the structure of the laugh, but for some purposes it may be sufficient.

(At higher levels of delicacy, e.g. Level 3 and above, the transcriber may choose to represent laughter in more detail, noting the occurrence of each individual *pulse* of laughter. Laugh pulses are sometimes referred to as “particles” or “syllables” of laughter. The present transcription system uses the @ sign for such higher-delicacy transcriptions, inserting one token for each pulse of laughter. The @ sign may be seen as vaguely iconic, reminiscent of a smiley face.)

Meta Categories

Categories: *comment, unintelligible, uncertain, header*

In addition to the “core” features noted above, a certain number of “meta” transcription categories are introduced at this level. Typically these involve metatranscription, i.e. the transcriber's commentary about the transcription proper. In the metatranscription function, a transcriber makes an observation about some aspect of the transcription itself: for example, indicating that the transcriber cannot discern what words are being uttered in some portion of the recording, or representing a low degree of certainty regarding the hearing of some word or phrase in the transcription, etc. Metatranscription commentary allows the transcriber to indicate, for example, that the transcription of a certain word is uncertain because the word was not clearly audible, or that the word cannot even be guessed at. Finally, a general comment notation allows the transcriber to comment on anything that seems necessary for clarifying the

transcription.

Metatranscription

21. comment	((COMMENT))	transcriber's comment on any topic
22. unintelligible	###	one per syllable
23. uncertain	#you're #kidding	transcribed words are uncertain
24. pseudograph	¬Jill	pseudonym etc. for anonymity (negation sign)
25. header information	((SBC002: 173-211))	e.g. recording source cited as a "comment"
26. header information	((Transcription: DT2))	e.g. transcribing conventions cited as "comment"

[rev. 16-Aug-2009]

Inserting Comments

The "comment" notation (signaled by double parentheses enclosing a word or sentence of commentary) is a useful and highly flexible notation which allows the transcriber to insert notes or observations about any aspect of the discourse being transcribed. Phenomena which can be transcribed using the comment notation include odd sounds that are hard to represent, physical actions by participants, the context of an action, or anything else the transcriber deems relevant.

Because the comment notation gives the transcriber the flexibility to insert an observation on any topic, it may be used to deal with phenomena in the recording for which there is no standard transcription symbol. As an interim measure, it can also be used by new transcribers when they have not yet learned the standard transcription symbol for some phenomenon. In either case, an ad hoc description of the sound, action, or other phenomenon at hand is created as needed, and introduced into the transcription using the comment notation.

Although the comment notation gives the transcriber great flexibility, it should be used in moderation, precisely because it is a relatively ad hoc form of representation. Excessive reliance on the general comment notation, instead of more specific transcription symbols, can undermine the attempt to provide a systematic representation of the discourse events and their temporal relationship to each other.

[rev. 24-Oct-2008]

Level 2 Transcription: *Exercise A*

Objectives

The goal of this session is to learn how to transcribe at Level 2 (“*Basic*” level of delicacy).

We will focus on aspects of transcription which are newly introduced in Level 2. In addition, we will continue to pay special attention to the identification of intonation unit boundaries. (We continue to focus on relatively simple cases of intonation units, most of them being indicated by a convergence of several strong cues.)

Task

The following examples are given in a simple “lexical” transcription—what we call a “Level 0” transcription, which contains only the words (as transcribed in conventional standard orthography), plus the temporal sequence in which the words were uttered.

Your task is to listen to the recording, and produce a Level 2 transcription, representing the relevant features for this level.

Features to be added: *Intonation units* (from Level 1) plus Level 2 features, especially *boundary intonation* (*continuing, final, appeal*), *pause* (without timing), *laughter*, and *unintelligible* or *uncertain* hearings.

Starting Point (a): Level 0

((*A Book about Death* SBC005: 173.55-211.02))

((1st speaker = PAMELA; 2nd speaker = DARRYL))

I mean books words I mean I mean they just become handbooks you distill them and use them in your own way no no no I don't I don't I I come up with my own ideas about that stuff and where do you get the ideas thought where do you get those thoughts processing what goes on around me well isn't a book part of what goes on around you well more from an oblique source

Starting Point (b): Level 1

((*A Book about Death* SBC005: 173.55-211.02))

PAMELA; I mean
 books
 words
 I mean
 I mean
 they just become handbooks
 you distill them
 and use them in your own way

DARRYL; no
 no
 no I don't
 I don't
 I
 I come up with my own ideas about that stuff

PAMELA; and where do you get the ideas

DARRYL; thought

PAMELA; where do you get those thoughts

DARRYL; processing what goes on around me

PAMELA; well
 isn't a book part of what goes on around you

DARRYL; well
 more from an oblique source

Target: Level 2

((A Book about Death SBC005: 173.55-211.02))

1		...
2	PAMELA;	I mean,
3		.. books,
4		...
5		wor:ds.
6		I mean,
7		...
8		#I #mean,
9		they just become handbooks.
10		You distill them,
11		and use them in your own way.
12		...
13	DARRYL;	No,
14		...
15		No,
16		.. No I don't.
17		...
18		I don't.
19		...
20		(H) I,
21		...
22		I come up with my own ideas about that stuff.
23		...
24	PAMELA;	And where do you get the ideas.
25		...
26	DARRYL;	Thought.
27		...
28	PAMELA;	#And where do you get those thoughts?
29		...
30	DARRYL;	Processing what goes on around me.
31		...
32	PAMELA;	Well?
33		...
34		Isn't a book part of what goes on around you?
35	DARRYL;	Well,
36		...
37		more from an oblique sou:rce.

[rev. 24-Oct-2008]

Chapter **: Transcription Level 3 (Boundary)

Objective

The objective at this level is to learn how to incorporate more information into a transcription, especially with respect to boundary features. The practice brings new levels of detail to a transcription, increasing its delicacy from a Level 2 (*Basic*) transcription to a Level 3 (*Boundary*) transcription.

Core Categories

For this transcription level, the transcription categories in the “core” of new features are mostly those that mark some aspect of the boundary between one unit and the next. Although the marking of boundaries for words and intonation units was introduced from the beginning (already at Level 1), the present level of transcription provides a more specific characterization of the nature of the intonation unit boundary, along with an indication of some of the most important cues which mark or correlate with intonation unit boundaries.

Specifically, new transcription categories introduced at this level which are directly related to unit boundaries include *truncation* (including *intonation unit truncation* and *word truncation*). In addition, other new transcription categories at this level represent cues which are correlated with the existence of a boundary, or with the strength of a given boundary. These include *breath* (i.e. *inhalation* and *exhalation*), *pause duration*, *lag* or *prosodic lengthening*, *micropause* or *hold*, *latching*, and *linking*.

A key transcription category introduced at Level 3 is *speech overlap*, which occurs when there is simultaneous speech by two or more participants in a given interaction. Overlap concerns the temporal relationship of simultaneity between utterances produced by two or more participants. Speech overlap is an aspect which carries fundamental information about the organization of interaction between speakers. For this reason, the Boundary level of transcription should include at least a general indication of where the overlap between two speakers begins and ends. In fact, using readily available computer-assisted transcription tools, it is a fairly simple matter to ascertain exactly where simultaneous speech begins and ends. Whenever possible, it is worthwhile investing the time and effort to accurately represent this information.

To avoid confusion when multiple overlaps occur in close succession, overlap brackets that belong together are indexed using numbers adjacent to the brackets. In addition, left brackets are aligned vertically for iconic clarity.³

Finally, Level 3 adds some additional detail about the nature of one of the *end tones* introduced at Level 2, the *appeal* contour. An *appeal* contour may be heard as conveying information about closure as well, simultaneously signaling either *continuing* or *final* along with appeal. In other words, the intonational features of *closure* (*continuing/final*) and *appeal* may be combined, for higher delicacy transcriptions.

Conventions

The following conventions are those used for transcription categories in Level 3 (“Boundary”) transcription. In addition, Level 3 naturally incorporates all transcription categories from lower levels (i.e. Level 1 and Level 2 categories).

Meaning	Symbol	Comments
Overlap		
overlap start & end (1st set)	[]	align left square brackets vertically
overlap start & end (2nd set)	[₂]	left square brackets indexed with subscript
Breath		
in-breath	(H)	audible inhalation
out-breath	(Hx)	audible exhalation
Truncation		
truncated intonation unit	—	aborting projected intonation unit (em dash)
truncated/cut-off word	wor—	aborting projected word (en dash)
Pause/Tempo		
pause, timed	(1.2)	number marks pause duration in seconds
lag/lengthening	:	prosodic slowing of current tempo (colon)
hold/micro-pause	..	less than 180 milliseconds; brief silence
Boundary Attenuation		
latching	=	no pause between speaker turns
linking (absence of break)	—	merging of words in rapid speech (underscore)
Boundary Intonation (complex)		
continuing appeal	?,	combines with closure feature
final appeal	?	combines with closure feature (usu. unmarked)
discontinuous IU	&	used only when 2nd speaker intervenes

Pause Attribution

Whose pause? Whose turn?

172b	PETE;	I stuck up for you today at that store.
173	HAROLD;	That's true.
174a		(0.3)
174b	JAMIE;	You did.
175		[You made me get] the u ₂ :m,
176	PETE;	[Mhm],
177	PETE;	[₂ Tha][₃ t's right].
178	#X;	[₃ ##].
179	JAMIE;	the green scarf.
180a		(0.9)
180b	JAMIE;	That's right.
181a		(0.9)
181b	JAMIE;	He was my fashion consultant today.
182	PETE;	Ri:ght.
183a		(0.6)
183b	PETE;	@@@

184 MILES; Hm.
 185 PETE; @(H)
 186a (3.3)
 186b HAROLD; I wasn't?
 187a (0.3)
 187b PETE; That's right.
 188 JAMIE; Well you%- —
 189 .. I don't know.

Attribution Units

The phenomenon of pausing presents a challenge to transcribers, in large part because of complications it introduces regarding the issue of attribution. Attribution is a key dimension of discourse transcription. When words are spoken, we generally want to know who spoke them.

((LSAC 1682-02))

2835; **Who said that?**

2833; I did.

((LSAC 1398-01))

1833; [Wait wait wait

who who **who said this to who?**]

1830; Chris did to Matthew.

But when the discourse phenomenon is a pause – the *absence* of speaking by any participant – how are we to decide who the pause should be attributed to? While it would seem useful to be able to indicate who is pausing, at the same time it can be difficult to make a definitive attribution of a particular pause to a particular participant. There is a level of ambiguity in the attribution of pauses that does not arise in attributing, for example, speech. The ambiguity as to who “owns” the pause is a not just a problem for analysts: participants as well may be uncertain as to who is responsible for a particular pause. The question may arise whether pauses should be attributed at all.

One step toward a resolution of this quandary is to recognize the importance of defining appropriate “attribution units” for discourse transcription. An attribution unit is a salient stretch of the ongoing stream of discourse phenomena (including speech, pauses, gestures, gaze and other kinds of events) which is suitable for attribution to some participant, or group of participants. In general, the clearest and least problematic case of an attribution unit is one involving speech. Because each speaker’s voice is distinctive, we generally have no trouble in identifying who is speaking. So the words spoken within one intonation unit constitute a natural and obvious attribution unit.

For pauses, on the other hand, the picture is less clear. There are no cues comparable to voice quality that can help us in assigning a pause to a particular individual. Thus it is most prudent to recognize pauses (or at least boundary pauses) as separate attribution units in their own right, distinct from the speech they occur next to. This allows one to take special measures to deal with the ambiguity of attribution of boundary pauses, without compromising the clarity of the attribution of neighboring speech. If boundary pauses and speech are combined together in a single attribution unit, the transcriber’s hands are tied, and it becomes impossible to represent a difference in attribution, or even a difference in the certainty of the attribution.

Classifying pauses for attribution

In order to facilitate the meaningful attribution of pauses, it is necessary to recognize that there are different kinds of pauses. One of the most important dimensions along which kinds of pauses may be distinguished is based on the contextual conditions which govern their potential attribution to a participant.

When should a pause be recognized as an attribution unit in its own right? In general, whenever it occurs at a well-defined boundary. For present purposes, the relevant boundary is the intonation unit. Because of the importance of pause attribution, the first distinction to make among pauses is that between an internal and a boundary pause (with boundary pauses subject to further differentiation).

Internal Pause. A pause which occurs in the middle of a speaker's intonation unit is considered an internal (or hesitation) pause. An internal pause should be written "where it happens", that is, as part of the same attribution unit as the surrounding words within the same intonation unit (that is, as located in the linear stream of words spoken by the speaker).

Boundary Pause. A pause which occurs between two intonation units (that is between words which can be attributed on prosodic grounds to separate intonation units) is considered a boundary pause.

Boundary pauses can be further divided into two categories, the medial pause and the transitional pause.

Medial/Continuation Pause. A pause which occurs between two intonation units, both of which are uttered by the same speaker, is considered a medial (or continuation) pause.

Transitional/Turn/Change Pause. A pause which occurs between two intonation units, each of which is spoken by a different speaker, is considered a transitional (or turn, or change) pause.

Functional Interpretation: Whose pause?

Responsibility for a given pause is determined primarily by what overarching activity continues in progress through the pause, that is, what activity is understood as beginning prior to the pause and continuing after it.

In analyzing a particular conversation between Marcia and Donny, Schegloff observes:

"Although everyone is silent (which silence as a state requires), someone in particular—Marcia—is "relevantly not talking." Donny has produced a possibly complete turn, one that implicates some responsive action next—by Marcia. Absence of talk is then, in the first instance, attributable to Marcia. So although the effect of her silence is that no action seems to get done, what she is specifically and relevantly "not doing" is registering some uptake of what has been told, and/or some assessment of it—for it is these that Donny's announcement has made conditionally relevant."

Schegloff (1995:197)

Schegloff, Emanuel A. 1995. Discourse as an interactional achievement III: The omnirelevance of action. *Research on language and social interaction* 28:185-211.

Pause Types and Pause Transcription

Principles

Table 1. Pause Types and Pause Transcription: Placement Rules

Pause Type	Locus	Comments	Transcription
boundary pause	at the boundary between two intonation units	<ul style="list-style-type: none"> the most common type of pause in spontaneous speech serves as a cue for intonation unit boundaries to iconically represent the fact that the pause represents a discourse moment that is relatively “open” for appropriation by any participant, boundary pause is written on a line by itself 	on a line by itself, between two intonation units
turn-boundary pause	at the boundary between two turns	<ul style="list-style-type: none"> basically the same as a boundary pause, but the boundary it defines is (also) between two turns (or at least between the words of the prior speaker and next speaker) carries greater ambiguity as to “whose pause” it is: prior, next, or some other speaker 	on a line by itself, between two turns
medial pause	between the words of a single intonation unit	<ul style="list-style-type: none"> occurs in the midst of an in-progress intonation unit, often as part of a word search the in-progress intonation unit may also represent an in-progress turn-constructual unit assumes that the overall intonational cues suggest the absence of an intonation unit boundary, despite the pause 	on the same line with surrounding speech, as part of the same intonation unit
marginal pause	between in-breath and the body of the intonation unit	<ul style="list-style-type: none"> occurs following an in-breath, considered to be a strong cue for claiming the start of an intonation unit and/or the start of a turn occurs at the left margin of the intonation unit, before the first word positioned like a boundary pause, but not open for appropriation by other participants 	on the same line as the surrounding vocalizations (e.g. breath and speech), as part of the same intonation unit
hold/micropause	between words, as very short pause or break in rhythm	<ul style="list-style-type: none"> normally considered to be a rhythmic aspect of speech involving (a) a momentary lapse or interruption of ongoing speech, or (b) a momentary delay of uptake in a new turn typically functions as a feature, often more or less predictable, of the speech it immediately precedes 	immediately before the following speech, as part of the same intonation unit
breath-filled pause	when only sound is audible breath	<ul style="list-style-type: none"> would be treated as a pause except for audible in-breath may function for participants partly like a pause, partly like a turn, turn bid, or turn start 	typically on same line as following speech by same speaker
filled pause	when only sound is a hesitation word	<ul style="list-style-type: none"> hesitation words (e.g. <i>uh</i>, <i>um</i>, etc.) are often produced with stylized/chant steady pitch controversial whether should be termed pauses 	spelled according to conventions for marginal words

[rev. 24-Oct-2008]

Meta Categories

The meta categories at Level 3 mostly involve establishing links between the transcription and the corresponding portions of the audio/video recording. This can be achieved by inserting timestamps in the transcription, whether for every line or only intermittently, and using computer software to align the transcription with the original recording, allowing audio/video playback synchronized with the transcription. The time codes also serve as a form of indexing the transcription, for reference and access to specific locations within it. A more modest means of indexing a transcription is to insert line numbers at the beginning of each line. Line numbers can also be combined with timestamps, providing a more convenient choice of ways to reference portions of the transcription.

A metatranscription category of an entirely different sort that is introduced at Level 3 is the pseudograph. The pseudograph notation indicates places where the transcription has been modified in order to obscure sensitive information uttered by participants. For example, if a speaker on the recording utters the name of another participant who wishes to remain anonymous, a pseudonym is invented and substituted for the actual uttered name. The term pseudograph applies at a more general level, including not only pseudonyms but also other kinds of substitutions for potentially sensitive information uttered by conversational participants, for example if they say their address, telephone number, credit card number, or such.

Supplementing the conventions introduced at Level 2 for the transcription of uncertain hearings, a further convention allows transcribers to indicate two alternative hearings, when they are uncertain which of two words was spoken.

Metatranscription

index/line numbers	<LINE=1>	each line timestamped and/or numbered
recording alignment		transcription linked to audio/video recording
timestamp	<TIME=154.7>	time in seconds from start of recording
alternative hearing	#can_((#can't))	uncertain which word is heard (no space)
alternative hearings	<#1> words </#1>	uncertain which of 2 word sequences is heard
(complex)	<#2> words </#2>	(no space)

[FN For presentation transcripts, it may be possible to rely on the vertical alignment, and thus dispense with the numbered bracketing. However, for database transcriptions, it is preferable to index overlap brackets by numbering them, to avoid unrecoverable disruption of overlap information during editing.]

[rev. 26-Oct-2008]

Utterance start

The category of *utterance start* indicates that the speaker is beginning new utterance. It is marked by capitalizing the first letter of the first word of the new utterance. The signaling of utterance start represents an aspect of speaking as it is found in naturally occurring discourse. The category of utterance start should not be interpreted in a normative way, imposing a literary conception of the written sentence onto spoken discourse. Certainly, when the speaker begins uttering a “proper” sentence, she is likely to use the prosody that signals utterance start, but she will also use it for backchannels (*Unhhunh*, *Mhm*), discourse particles (*Well*, *Okay*), new turns, repetitions, stutters, false starts, and more. The category of utterance start corresponds to the beginning of a new utterance of any kind, regardless of whether or not the previous utterance was completed (whether in a normative or any other sense of completeness). In spoken discourse, it is not realistic to expect that the prior utterance will necessarily always be completed before a new utterance is begun. Thus it is not useful to impose a criterion of completeness, nor of grammaticality. Moreover, the focus of the category is on the beginning of something, not on a whole unit. As long as the speaker is starting something new, it qualifies for capitalization by this convention.

Strictly speaking, it may be debated whether the utterance start category should be considered part of transcription per se. It may be argued that audible cues for utterance start are difficult to specify

precisely. The argument can be made that that in transcribing utterance start, transcribers may be attending in part to other, non-prosodic cues, such as grammatical cues for the beginning of a syntactic sentence. Evidence in support of this position would be that the location of utterance start can often be guessed at simply by reading over the transcription, without the need to go back and listen again to the original recording. On the other hand, the argument in favor of utterance start as a viable transcription category in the strict sense is that there are in fact many audible cues that contribute to its recognition (including final contour at the end of the previous utterance and pitch reset at the start of the new utterance.) Some of the relevant cues are in effect incorporated into the transcription by virtue of marking, for example, final intonation, using a period. There's a lot we don't know about the perception of utterance start, and further research is called for. In any case, taking the time to mark utterance start—whether on prosodic or syntactic grounds, or a mixture of the two—serves an important function in making transcriptions easier for users to read, by indicating when the speaker is perceived as starting something new in the discourse.

Level 3 Transcription: Exercise 1

Objectives

The goal of this session is to learn how to transcribe at Level 3 (“*Boundary*” level of delicacy). We will focus on aspects of transcription which are newly introduced in Level 3.

Task

The following example is given in “Level 2” (Basic) transcription. You should bring the transcription up to Level 3 (Boundary).

Features to be added: Selected Level 3 features, especially *speech overlap*, *breath*, *intonation unit truncation*, and *word truncation/cut-off*.

(If time allows, you may also choose to add additional Level 3 features, especially *lag/lenghtening*, *micropause/hold*, *latching*, and *pause duration*).

Starting Point: Level 2

(*A Book about Death* SBC005: 173.55-211.02)

1		...
2	PAMELA;	I mean,
3		books,
4		...
5		w words.
6		I mean,
7		...
8		#I mean,
9		they just become handbooks.
10		You distill them,
11		and use them in your own way.
12		...
13	DARRYL;	No,
14		...
15		No,
16		No I don't.
17		...
18		I don't.
19		...
20		I,
21		...
22		I come up with my own ideas about that stuff.
23		...
24	PAMELA;	And where do you get the ideas.
25		...
26	DARRYL;	Thought.
27		...

28 PAMELA; #And where do you get those thoughts?
29 ...
30 DARRYL; Processing what goes on around me.
31 ...
32 PAMELA; Well?
33 ...
34 Isn't a book part of what goes on around you?
35 DARRYL; Well,
36 ...
37 more from an oblique source.

Target: Level 3

(*A Book About Death* SBC005: 173.55-211.02)

1		(0.3)
2	PAMELA;	I mean,
3		.. books,
4		(0.3)
5		w- wor:ds.
6		I mean,
7		(3.0)
8		#I mean,
9		they just become handbooks.
10		You distill them,
11		and use them in your own way.
12		(4.5)
13	DARRYL;	No,
14		(0.6)
15		No,
16		.. No I don't.
17		(0.3)
18		I don't.
19		(1.2)
20		(H) I,
21		(2.3)
22		I come up with my own ideas about that stuff.
23		(0.7)
24	PAMELA;	And where do you get the ideas.
25		(0.6)
26	DARRYL;	Thought.
27		(0.9)
28	PAMELA;	#And where do you get those thoughts?
29		(0.3)
30	DARRYL;	Processing what goes on around me.
31		(0.5)
32	PAMELA;	Well?
33		(0.5)
34		Isn't: a book part of what goes on [around you]?
35	DARRYL;	[(H) Well],
36		(1.8)
37		more from an oblique sou:rce.

[rev. 24-Oct-2008]

Chapter **: Transcription Level 4 (Interaction)

Objective

The objective at this level is to learn how to incorporate more information into a transcription, especially with respect to aspects of the body, or paralinguage, which are relevant to interaction. The practice brings new levels of detail to a transcription, increasing its delicacy from a Level 3 or “Boundary” transcription to a Level 4 or “Interaction” transcription.

Core Categories

For this transcription level, the transcription categories in the cluster of new core features are mostly those that mark some aspect of the non-verbal use of the body, sometimes called paralinguage, many of which have particular relevance to the organization of interaction between participants. Features of the body relevant to participant interaction include gestures of the hands and face, as well as other aspects of the body overall. But it should not be overlooked that the vocal tract is part of the body too. Although much of what speakers do with the vocal tract is part of the highly specialized system of language, there are other, non-verbal uses of the vocal tract as well, such as clearing the throat, coughing, sniffing, and so on. There is an open-ended set of such vocal-tract cues which are potentially significant to interaction. When relevant, they are written in single parentheses. (The new categories at this transcription level expand on the body/paralinguage categories introduced previously, such as breathing.)

The following conventions are those used for transcription categories in Level 4 (“Interaction”) transcription, as described in the “Transcription Delicacy Hierarchy”. As always, Level 4 subsumes all transcription categories from lower levels, i.e. Levels 1, 2, and 3.

MEANING	SYMBOL	COMMENTS
Vocalism		
(4) vocalism (general)	(COUGH)	various notations: (SNIFF), (AHM), etc.
(5) click	(TSK)	alveolar click
(6) glottal vocalism	(%)	separate vocalism (separate glottal “word”)
Manner (Basic)		
(7) manner, quality	<MISC> </MISC>	brackets at start and end (various notations)
(8) piano, soft	<P> </P>	quiet speech
(9) piano, soft (alternate)	°word word word°	degree sign (alternate notation)
(10) laugh voice	<@> </@>	laugh while speaking (for seven+ words)
(11) laugh voice (alternate)	@you’re @kidding	laugh symbol marks laughter during word
(12) smile voice	£you’re kidding£	pound symbol marks smiling during word
(13) smile voice (alternate)	☺ you’re kidding ☺	smile symbol marks smiling during word
(14) glottalized word	%word	modified manner of vocal production
(15) vox: voice of another	<VOX> </VOX>	animating the voice of another person
(16) vox (alternate)	« »	alternate (double angle quotation mark)
(17) vox specification	<VOX=DUMB>	specifies interpretive evaluation of voice quality
(18) affect	<AFFECT=ANGRY>	specifies interpretive evaluation of affect/emotion
Non-Vocal Events and Co-Actions		
(19) non-vocal event	{KNOCK}	e.g. knock on door, telephone ring, etc.
(20) co-action (general)	©	act may be performed simultaneously with speech
(21) co-action (specified)	<©=DESCRIPTION>	specifies the type of co-action performed

Body

For transcribers working with video, a crucial aspect of transcription is the representation of aspects of gaze, gesture, and other aspects of body practices, as well as other aspects of action and activities as situated in context. These are complex topics which merit in-depth treatment in their own right, for which the reader is referred to the writings of Hockett & Pittenger, Birdwhistell, Kendon, Goodwin, Streeck, McNeill, and others.

Meta Categories

The so-called “meta” categories at Level 4 mostly involve aspects of the participation frame. These are not necessarily directly audible in the way that words and vocalisms are, but they may be indexed by certain audible or visible features that are associated with them or even specifically index the relevant roles. This set of categories mainly concerns the nature of the relationship between participants and their actions, or between one participant and another. Together they define the participation frame, the social framing within which roles, responsibilities, and actions in conversation are interpreted.

Participation Frame (Basic)

(22) non-vocal event	JILL; ((EXIT))	participant action, non-linguistic
(23) non-interactive agent	±GAME; ((BEEP))	marks machines, recordings, etc.
(24) participation frame	JILL/TEL;	slash symbol precedes interaction context
(25) recipient/addressee	JILL\DOG;	backslash symbol precedes addressee
(26) multiple speakers	JILL+JIM;	for simultaneous joint action of individuals
(27) group of speakers	AUDIENCE++;	for simultaneous joint action of crowd
(28) role voicing	JILL=JIM;	speaker enacts role voice or voice of another
(29) participant gloss	da_((μ=DOG))	what the speaker means (e.g. for child speech)

[rev. 26-Oct-2008]

Chapter **: Transcription Level 5 (Intonation)

Objective

The objective at this level is to learn how to represent intonation on a more comprehensive level, building on the basic intonational categories introduced at prior levels. Specifically, the practice brings new levels of detail to a transcription, increasing its delicacy from a Level 4 or “Interaction” transcription to a Level 5 or “Intonation” transcription.

Core Categories

New transcription categories introduced at this level include aspects of accent, stress, tone, tune, rhythm, key, reset.

Since the transcription of intonation is far from settled, with many competing systems, this section references several alternative notations. These representing other analyses of intonation (or other categories) that may be used at this level, including notations for intonation units, contours, and primary and secondary accents.

Core Conventions

The following conventions are the most basic versions of the main core categories used used in transcribing at Level 5 (“Intonation”) transcription, as described in the “Transcription Delicacy Hierarchy.” As usual, Level 5 also subsumes all prior transcription categories (Levels 1 through 4).

Note: This list is selective. For a complete listing of the intonation categories for Level 5 transcription, see the section on “Intonation” in the “Reference List of Discourse Transcription Symbols.”

	MEANING	SYMBOL	COMMENTS	UNICODE
Unit				
1.	intonation unit	{LINE}	one new line for each IU	2016
2.	intonation unit		double vertical line marks IU (alternat.)	
Accent & Stress				
3.	primary accent/nucleus	^	cover term (circumflex accent)	005E
4.	secondary accent	`	(grave accent)	0060
5.	stress	*	utterance stress realization	00F7
6.	chant/stylized	÷	stylized calling contour, etc.	
Heads & Tails				
7.	appeal onset/plateau	¿	alternate (inverted question mark)	00BF
8.	head onset	/	precedes first prominent syllable of IU	
9.	tail offset	\	precedes trailing unstressed final syllables	
Contour				
10.	rise	↗	notation for alternative analysis	2197
11.	fall	↘	notation for alternative analysis	2198
12.	level	→	see “sustained” tone/accent	2192
13.	rise-fall	↗↘	unitary rising-falling contour	(alternate notation)
14.	fall-rise	↘↗	unitary falling-rising contour	
15.	rise-fall	↗↘		

Part 3: Categories

Chapter **: Units

Concentric Units

A The Concentric Unit is a stretch of discourse that constitutes a unit in multiple dimensions simultaneously, at various levels of phonological, grammatical, and interactional structure.

B. Among the units at different levels which may commonly converge in discourse, reflecting the concentric unit, are the intonation unit, clause, sentence, turn, speech act/move declination unit. Less common, but well attested, are the word and syllable.

C. Concentric Units may provide evidence for the structure of one level based on the structure of another level.

E. Concentric Units are frequent in early child dialogue, and may be used by children as a source of structural information.

D. While Concentric units are not necessarily of high frequency in adult discourse, they are highly significant as a locus of maximum structural information.

(1) { SBC005 A Book About Death}

327.68 333.53	DARRYL:	... (H) So why are you reading a book about dying,
333.53 340.67	PAMELA:	... <P> I don't know </P>.
340.67 342.12	DARRYL:	.. you don't know?
342.12 344.19	PAMELA:	.. I have an interest in it.
344.19 344.72	DARRYL:	Why.
344.72 348.20		... You're alive.
348.20 350.25		Why are you r=eadng a book about dying.
350.25 354.08	PAMELA:	... I've always been interested in it.
354.08 354.88	DARRYL:	% <W> Why </W>.

(2) { SBC005 A Book About Death}

190.83 195.28	DARRYL:	... I come up with my own ideas about that stuff.
195.28 197.33	PAMELA:	... And where do you get the ideas.
197.33 198.69	DARRYL:	... Thought.
198.69 200.69	PAMELA:	... And where do you get those thoughts?
200.69 202.80	DARRYL:	Processing what goes on around me.
202.80 203.73	PAMELA:	... Well?

(3) { SBC002 Lambada}

1024.04 1024.89	JAMIE:	Do you need a partner?
1024.89 1026.70	MILES:	... To go there?
1026.70 1027.14	JAMIE:	.. Yeah.
1027.14 1027.59	MILES:	.. No.

(4) { SBC007 A Tree's Life}

1310.55	1311.75	MARY:	... Did she talk to you lately?
1311.75	1312.24	ALICE:	Hm-m.
1312.24	1314.93	MARY:	... They had a cyst on her ovary.
1314.93	1318.27		... And they had to go in and take it out?
1318.27	1318.84	ALICE:	... Mhm.
1318.84	1320.57	MARY:	... Take out that cyst?
1320.57	1320.99	ALICE:	[Mhm].

Intonation Unit Suspension

In the phenomenon of intonation unit suspension, the speaker momentarily pauses in the middle of an intonation unit—without either completing the unit as such, nor abandoning/truncating it either—while another speaker intervenes. The intonation unit is simply suspended for the moment, hanging there in mid-air, as it were. Finally, the first speaker completes the original intonation unit, more or less as they had projected to produce it in the first place. In principle, if you could edit out or mask the intervening (second) speaker, the first speaker would sound like they had produced a single continuous (“normal”) intonation unit, with the expected intonational cues marking the usual beginning, middle, and end trajectory of an intonation unit. Consider the following example:

27. (*Tape Deck* SBC016: 1126.035-1130.205, ln. 1211)

1	TAMMY;	(H) You don't sell that kind of thing here.
2		Right?
3		Cause this is an audio &
4	BRAD;	.. That's right.
5	TAMMY;	& place.

Here, the intonation unit that Tammy is in process of producing is *Cause this is an audio place* (lines 3 and 5). But in the middle of Tammy's projected intonation unit, Brad jumps in to agree with what he takes to be the meaning of her projected utterance (*That's right*, line 4). Though Tammy's intonation unit is yet complete in line 3, neither is it truncated at that point, nor abandoned, but simply suspended momentarily, before being taken up again in line 5. This is what is analyzed as intonation unit suspension, marked with an ampersand (&). It is important to note that the claim being made by this representation is that Tammy's *Cause this is an audio place* is a single intonation unit, even though it is written on two lines. In general, this is one of very few exceptions to the rule that an intonation unit is always written on a single line. (The other main exception is when an intonation unit is very long and cannot be written on a single line for purely typographical reasons, but in that case, no carriage return is inserted; the extra-long line is simply allowed to wrap.)

Although it may be said that “in principle” the two parts of the suspended intonation unit would fit back together as if they had been produced without an intervention, the reality is somewhat more complex. The initial speaker who has suspended her intonation unit seems to monitor the production of the second, intervening speaker, and may display this orientation by waiting until they finish their intervention before resuming the suspended intonation unit.

The symbol which marks suspension of an intonation unit, the ampersand (&), is written in two places. The first indicates the point where the intonation unit is first suspended, and the second indicates the

point where the utterance of the intonation unit is resumed by the same speaker. Note that the suspension analysis is NOT applicable when one speaker's sentence or clause is completed by another speaker. Such cross-speaker collaborative completions {Lerner, etc.} are different in a variety of ways, which we don't need to go into here. Suffice it to say that the intonation unit suspension analysis applies only when the two parts of the utterance are produced by the same speaker.

Consider the following somewhat more complicated example:

28. {SBC004 *Raging Bureaucracy*, line 1076}

KATHY; You only have twelve kids?

SHARON; (H) No.

.. **Seventee:n &**

KATHY; Oh,
okay.

SHARON; **& [fourth-graders].**

KATHY; [(H) So then] what you do is,

The intonation unit that Sharon is in the middle of producing is *Seventeen fourth-graders*, with a sort of “pause” in the middle. (This is not really a pause, and hence is not written as a pause, because someone else is speaking during it.) It just happens that in the middle of Sharon's suspended intonation unit, Kathy talks, marking her early recognition of what she assumes to be Sharon's projected utterance. But Sharon's intonation unit has not been truncated or abandoned, nor has it been intonationally completed.

One could imagine another transcriptional approach: writing Sharon's *Seventeen fourth-graders* all on one line as a normal intonation unit, with Kathy's two intonation units *Oh, okay* overlapping with a “pause” by Sharon after the word *seventeen*. However, once you admit the notion that a speaker can overlap with nothing (i.e. overlap with a pause), this leads to all kinds of unexpected complications—a slippery slope, in both methodological and theoretical terms. For this reason, Discourse Transcription (DT2) conventions strongly discourage writing an overlap with a pause, preferring instead the analysis in terms of suspended intonation units.

Though the concept of intonation unit suspension is of some theoretical importance for properly understanding the interactional implications of the way intonation units unfold in discourse time—especially in contexts where recipients intervene early to display their successful projection of what is coming next—it is nevertheless a rather rare phenomenon. Transcribers are likely to encounter it only rarely, if at all, so the notation should be used sparingly. (Crucially, it is not used when a speaker suspends a clause, but brings the relevant intonation unit in question to some sort of completion.)

The symbol for suspended intonation unit is used in some places where, by the conventions of Conversation Analysis {Jefferson}, a latching symbol would be called for. This is represented in Conversation Analysis by an equal sign (=). However, the categories of suspension and latching (*q.v.*) are actually quite different in functional terms, and it is preferable that they should be represented by unambiguously distinct notations. To make the functional contrast clear, the Discourse Transcription practice is to represent suspension with & and latching with =.

Examples: Level 0 transcription

The following examples are presented in a Level 0 (Lexical) transcription, which crucially lacks speaker attributions, intonation units, and of course suspended intonation units. The apparent incoherence of the transcribed excerpts gives some idea of the confusion that is generated by failing to represent the key items of missing information.

29. {SBC023: *Howards End*, l. 1334}
there is a conjugal mhm oneness

30. {SBC007: *A Tree's Life*, l. 409}
there's a lot of (DRINK) smoke coming out

31. {SBC003: *Conceptual Pesticides*, l. 716}
his question was what impact is this gonna have on the sort of fucks up the animists cultural growth

32. {SBC011: *This Retirement Bit*, l. 527}
that was just and that was in the winter weakness exertion well maybe that's what this was

Examples: Level 3+ transcription

The following examples, versions of those presented in the previous section, are here presented in a standard transcription (Level 3+). Crucially, this transcription level adds information about speaker attributions, intonation units, and suspension of intonation units, the latter marked by a & at the beginning of the suspension and another & at the end, when the suspended intonation unit is resumed.

33. {SBC023: *Howards End*, l. 1334}
EVELYN; [there] is a conjugal &
...
LINDA; Mhm.
EVELYN; & oneness,

34. {SBC007: *A Tree's Life*, l. 409}
MARY; there's a lot of &
ALICE; (DRINK)
MARY; & smoke coming out,

35. {SBC003: *Conceptual Pesticides*, l. 716}
ROY; his question was,
what impact is this gonna have,
on the &
MARILYN; .. Sort of fucks up the animists.
ROY; & cultural growth,

36. {SBC011: *This Retirement Bit*, l. 527}
SAM; That was just &

DORIS; And that was in the winter.
 SAM; & .. weakness.
 ...
 Exertion.
 ...
 DORIS; Well maybe that's what this was.

37. {SBC019: *Doesn't Work in This Household*, l. 200}

RON; (TSK) (H) It's not pleasant to talk about politics with anyone who doesn't believe --
 .. % d- not believe,
 but doesn't &
 MELISSA; .. Agree with you.
 RON; & vote in the same way that you vote,
 or whatever.

[rev. 12-Jun-2007]

[Draft – do not circulate without permission]

Intonation Unit Cues in Context

John W. Du Bois
University of California, Santa Barbara
dubois@linguistics.ucsb.edu
September 4, 2008

Introduction

Intonation Units are prosodic units which contribute in a fundamental way to the organization of the structure and function of language. The Intonation Unit is apparently a universal category of spoken language, taking its place alongside the phoneme, syllable, and other basic phonological units. From a functional point of view, the Intonation Unit can be argued to represent the fundamental basic-level unit of language, at the level at which speakers realize linguistic conceptualization, communication, and action. The Intonation Unit constitutes a pivotal unit at the intersection of form (phonetic, phonological, syntax), meaning (semantics, pragmatics), and function (cognitive process, social interaction).

complex units of language use in the real-world of communication and social interaction.

The reality of Intonation Units is attested in the following observations:

- (1) Speakers indexically mark the start and end (or better, the transition into and out of) each Intonation Unit, as well as its internal structure and overall gestalt unity, in more or less consistent ways, within any given language;
- (2) Hearers regularly attend to and interpret these indexes of Intonation Unit boundaries and structure;
- (3) The phonetic form and prosodic structure of Intonation Units tend strongly to be marked in the same or similar ways across languages, constituting a fundamental language universal of prosody. Thus for both analyst and theorist of language, knowing how to identify the Intonation Unit within the ongoing stream of discourse, and understanding its structure, gestalt unity, and sequential development, is an essential part of the task of representing, analyzing, and theorizing language as produced and interpreted by participants.

Discourse and especially prosody are produced and experienced by participants as an ongoing stream of actions which partly overlap with and blend into one another, exhibiting fuzzy boundaries par excellence. Thus it is important for analysts to acknowledge that in discourse, neither units nor their boundaries are ever given for free. (The methodology of inventing sentences and producing them in isolation makes it appear as if the boundary is given in advance, but this is an illusion of the methodology.) For any unit of importance, participants as well as analysts are faced with the task of recognizing what kind of unit it is and where it starts, continues, and ends. Intonation units are, in the first place, carved out of the ongoing stream of discourse, through processes which modulate fluctuations in certain omnipresent audible signals, including tempo, pitch, and amplitude.

FLUX is key in understanding how Intonation Units are perceived. The idea of flux implies that a more or less continuous signal (be it tempo, pitch, amplitude, etc.) is modulated up or down at a given point in the ongoing stream of discourse. The point of flux, where there is a significant shift in the baseline or expected value for a given percept, helps to define a cue for the boundary between one unit and the next. What matters is not so much the absolute value of, say, duration, as the structured change in tempo, especially if this correlates with other changes such as those of baseline pitch (i.e. on unstressed syllables).

When several such cues are synchronized with other such perturbations they may together index, among other things, the presence of an Intonation Unit boundary.

Intonation unit realizations are characterized by prototype structure. That is, prototypical (or “best”) exemplars are marked by a full complex of well-defined multiple cues, other instances are marked by a reduced complement of cues. Thus there is considerable intrinsic variation in cue realization across instances of Intonation Units. Moreover, the various cues for the presence of an IU boundary differ in significance. Their importance in a given context is subject to a complex process of interpretation, which in principle should correspond to the interpretive process used by speakers and hearers as they produce and perceive Intonation unit boundaries. This paper represents my current hypothesis as to how the cues are weighted, and how their value and significance are modified according to the context they appear in, within the ongoing processes of discourse production and comprehension in interaction.

Identifying Intonation Units means essentially locating their boundaries in the stream of speech. In the end this comes down to making an essentially binary decision as to whether an Intonation Unit boundary is present or not at any given moment in the speaker’s ongoing stream of speech. But there is a lot that goes into answering this seemingly simple yes-no question. There are many different cues that must be considered, and while an experienced transcriber comes to find the process to be a fairly fast and intuitive one, it takes some time and study to get to that point. To move toward that level of transcription fluency, it may be useful if the complex process of evaluating Intonation Unit cues can be given some systematicity. It is the goal of this paper to present at least a partial systematization of the complex process of orienting to, weighing, and evaluating the various cues that together index the Intonation Unit and its boundaries.

Intonation Unit Cues

Cues by Complex

INTRODUCTION**

INSERT TABLE 1 ABOUT HERE

38. SCOPE indicates where in the Intonation unit the cue is to be found: in INITIAL position, FINAL position, at either initial or final BOUNDARY, spread over the whole UNIT, or requiring comparison across multiple units (‘multi-unit’).

In addition to these audible cues, Intonation Unit boundaries tend to show significant correlations with various non-audible but overtly observable behavioral cues including gesture, gaze, and other aspects of body behavior. Little is known about these correlations with Intonation Units at present, but the issue is ripe for investigation.

Table 1. Principal Intonation Unit cues

	CUE	COMPLEX	SCOPE	DEFINITION
1.	lag	tempo	end	tempo lag or prosodic (non-lexical) lengthening
2.	rush	tempo	start	rapid tempo unstressed syllables (anacrusis)
3.	closure tone	pitch	end	IU-final boundary tone distinguishing intonational finality vs. continuity
4.	pitch reset	pitch	start	rise/drop in overall baseline pitch level for IU (esp. on unstressed syllables)
5.	pause	pause	boundary	noticeable absence of speech by discourse participants
6.	creak	laryngeal	end	creaky voice on final portion of Intonation Unit (not consistent)
7.	breath	laryngeal	start	breathing in (and other vocalisms: exhale, throat-clear, sniff, click, etc.)
8.	tune gestalt	tune	unit	coherent intonation contour perceived as unified (holistic) gestalt for the unit
9.	isotony	tune	unit	repeated tunes across sequence of Intonation Units (intonational parallelism)
10.	turn start	turn	start	next speaker (new voice) begins
11.	turn end	turn	end	current speaker (current voice) ends
12.	accent count	prosody	unit	IU “size” in primary accents per IU (tends to be 1, 2 , or 0—in that order)
13.	register	prosody	unit	overall register shift (of pitch and/or amplitude) for whole Intonation Unit
14.	truncation	fluency	end	truncation masks normal end cues, but sometimes is signaled by glottal stop

Correlations of Intonation Unit boundaries with non-phonetic features such as word quantity, information quantity, information structure, syntactic structure, and so on, are of considerable interest in their own right. But they do not represent *audible* cues for Intonation Units as such.

Note: *Lag* or prosodic (non-lexical) lengthening is characterized by longer than expected duration, beyond the expected duration of the phoneme sequence for a given lexical word.

Isolation vs. Continuity

Some approaches to Intonation Units (or intonational phrases, or other similar units) treat the unit in isolation, often with invented examples illustrating a single intonational phrase out of context. The picture is like this:

Figure 1. Isolated sentence examples

Example (a): Intonation Unit

Example (b): Intonation Unit

The Intonation Units seem to come with their boundaries already given. Convenient, but intellectually risky. In failing to problematize the identification of the boundaries of Intonation Units, this approach gives a false sense of the demands of the task, and may lead researchers to overlook what is at stake in doing analysis of units of intonation at this level. The use of pre-segmented “examples” invites analytical complacency, tending to underestimate the challenge—for both participants and analysts—of identifying the boundaries of prosodic units, and in particular, Intonation Units.

A more realistic picture of the what the hearer initially faces is the following:

Figure 2. Continuous stream of discourse

Example (c):	Intonation	Intonation	Intonation	Intonation
--------------	------------	------------	------------	------------

The hearer encounters a more or less continuous stream of speech, characterized by a rapid succession of syllables with overlapping trajectories of change in various dimensions, including rising and falling pitch and amplitude, speeding and slowing tempos, shifts in voice quality. This is interspersed with, perhaps, the occasional pause. The native participant in conversation manages to segment this ongoing stream rapidly and expertly into the prosodic units which will be most useful for interpreting, engaging with, and reacting to what is heard. Central among the auditory units which carry significance for the rapid organization of speech is the Intonation Unit. As the hearer identifies each Intonation Unit boundary by its audible cues, the ongoing stream of speech is segmented into a well-organized structure characterized by more or less definite boundaries between Intonation Units. This is symbolized in the following:

Figure 3. Continuous stream of discourse, with Intonation Unit segmentation

Example (d):	Intonation Unit	Intonation Unit	Intonation Unit	Intonation Unit
--------------	-----------------	-----------------	-----------------	-----------------

In discussing Intonation Units (or any other linguistic unit, for that matter), it is crucial to keep in mind the following basic principle: Units are never given for free. Units must be earned, by doing the work of identifying cues that allow recognition of their boundaries. The prosodic boundaries that demarcate the Intonation Unit correlate with cues that represent the most significant segmentation of the ongoing voicestream.

As a practical matter, no-free-unit principle has immediate consequences for the practice of identifying Intonation Units. This is addressed in the contrast between the “unit perspective” and “boundary perspective”, as developed in the following two sections.

Unit Perspective

Given a set of cues that have proved relevant to the identification Intonation Units (assuming for the moment we have a reliable list of such cues), there are still several questions that must be answered about how to go about using the cues.

On a very basic level, we may ask where in the Intonation Unit the cues are located, that is, where we should listen for them and expect to find them. If initial rush (anacrusis) and terminal lag (prosodic lengthening) are cues for Intonation Unit boundaries, it is useful to list for the first at the beginning of the Intonation Unit and the second at the end. Similarly, pitch reset and audible in-breath are cues characteristic of the beginning of an Intonation Unit, while final boundary tones such as closure tones and terminal pitch (and sometimes creaky voice) are characteristic of the end of an Intonation Unit. Pause is a little trickier as to where it resides, being defined negatively by virtue of what doesn’t happen, as absence of speech (by participants in the current conversation). Given that a boundary pause actually occurs between one Intonation Unit and the next, one can think of pause as being present (potentially) at both the starting point and endpoint of the Intonation Unit.

Figure 4. Intonation Unit Cues: Unit Perspective

... (H) This is a double au:ght,		
↕	↕	
<i>START</i>	<i>END</i>	<i>COMPLEX</i>
Rush	Lag	<i>Tempo</i>
Reset	Closure Tone	<i>Pitch</i>
Pause	Pause	<i>Pause</i>
Breath	Creak	<i>Laryngeal</i>
<i>UNIT</i>		
Tune Gestalt		
Isotony		
Turn		
Accent Count		
Register		

Boundary Perspective

One of the tasks faced in representing discourse is multiple Intonation Units. The unit perspective discussed in the previous section can be schematized as in Figure 5.

Figure 5. Unit perspective: Multiple Intonation Units



We can contrast this with the unit perspective, schematized in Figure 6.

Figure 6. Boundary perspective: Segmenting the continuous voicestream

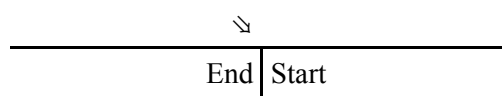


Table 3 illustrates this principle with a concrete example, presenting the same Intonation Unit as in Table 2 (*This is a double aught*), but now from the boundary perspective.

Table 3. Intonation Unit Cues: Boundary Perspective

... (H) *This is a double au:ght, this is an au:ght.*

		↕	↕	
<i>PART</i>		<i>END</i>	<i>START</i>	<i>COMPLEX</i>
		Lag	Rush	
		Closure Tone	Pitch Reset	
		Pause	Pause	
		Creak	Breath	
<i>WHOLE</i>		<i>UNIT₁</i>	<i>UNIT₂</i>	<i>Laryngeal</i>
		Tune Gestalt	Tune Gestalt	
		Isotony	Isotony	
		Turn	Turn	
		Accent Count	Accent Count	
		Register	Register	

Similarity Principle vs. Difference Principle

NB: Categorization Principles:

Similarity vs. Difference

Unification vs. Differentiation

Cue Validity: Weighting the Intonation Unit Cues

Variability of Intonation Unit cue validity

Prosodic features differ with respect to their validity as cues for the identification of Intonation Unit boundaries. Some prosodic cues are generally strong indicators of an Intonation Unit boundary, and may be taken to be highly criterial. Other cues are in themselves only weakly indicative of the presence or absence of an Intonation Unit boundary. They contribute to the criterial decision regarding boundaries only when they combine with other cues to form part of a larger pattern. Marginal cues may become more important, however, when for one reason or another the usually preferred cues are not accessible (e.g. when an Intonation Unit is truncated, obscuring any cues that would normally occur at the end of the unit). In addition to the overall patterning of observable cues, the context in which the pattern is found must be taken into account, in order to appropriately assess the contextualized implications for the presence or absence of Intonation Unit boundaries.

Table 4 summarizes the cue validity of the major Intonation Unit cues, with comments on their scope and context. For assessing the role of the various Intonation Unit cues, the concepts of cue scope, sufficiency, necessity, reliability/clarity, and neutralization of cues in context are relevant. Each of these concepts is introduced in the following paragraphs.

INSERT TABLE 4 ABOUT HERE

Scope

The scope of a cue concerns where in the utterance (or in the Intonation Unit) the cue is observable. Some cues are observable at the start of an Intonation Unit, others at the end. Still other cues can be assessed only by taking into account the whole Intonation Unit, and some even require attending to a series of Intonation Units. (Despite what is said in Table 4 about cues with “start” or “end” scope, actually it often becomes necessary to include both the end portion of one Intonation Unit and the beginning portion of the next in evaluating the significance of a pattern of prosodic cues; see the discussion of “Boundary Perspective” below.)

Necessary vs. sufficient

One way to evaluate an Intonation Unit cue is in terms of its positive vs. negative implications for existence of an Intonation Unit boundary. In any given context, the cue may be primarily *positive* in its implications, supporting the presence of a boundary, or *negative*, arguing against the presence of a boundary.

This can be understood in terms of *necessary* and *sufficient* criteria. A given Intonation Unit cue can be classified according to whether it is *necessary* for concluding that an Intonation Unit boundary is present. Independently, it can be classified as to whether it is *sufficient* for this conclusion. The interpretation of the terms necessary and sufficient in this sense is as follows:

Sufficient—high: If a “highly sufficient” cue is observed to be *present* at a certain point or region in the stream of discourse, this is strong evidence in favor of the presence of an Intonation Unit boundary.

Necessary—high: If a “highly necessary” cue is observed to be *absent* at a certain point or region in the stream of discourse, this is strong evidence against the presence of an Intonation Unit boundary.

Table 4. Intonation Unit cue validity. For cues with a high value under SUFFICIENT, cue presence strongly suggests Intonation Unit boundary presence. For cues with a high value under NECESSARY, cue absence strongly suggests Intonation Unit boundary absence. Cues with a high RELIABILITY value are generally easier to hear, and more likely to be agreed on by all observers. CONFOUND refers to the potential alternative interpretations of a given cue, which must be ruled out before the cue can be given its full weight as an Intonation Unit cue. NEUTRALIZING CONTEXT indicates a discourse environment in which the cue is irrelevant or indeterminate, and thus neither supports nor refutes the presence of an Intonation unit boundary.

	CUE	SCOPE	SUFFIC.	NECESS.	RELIABLE	CONFOUND	NEUTRALIZING CONTEXT
1	<i>lag</i>	end	high	high	high	primary accent, phonemic length	truncated IU, hyperfluency (rush-through)
2	<i>rush</i>	start	mid	mid	mid	internal unstressed syllables	
3	<i>closure</i>	end	variable	mid	mid	primary accent	level continuing contour vs. truncated IU
4	<i>pitch reset</i>	start	high	mid	variable	initial primary accent or high tone	sequence of unaccented & accented syllables
5	<i>pause</i>	end/start	mid	low	high	voiceless stop, silent activity	word search, disfluency
6	<i>breath</i>	start	high	low	high	laughter	bodily exertion, respiratory difficulty
7	<i>creak</i>	end	low	low	low	truncated word or Intonation Unit	speaker vocal style
8	<i>tune gestalt</i>	unit	mid	low	variable	unit-internal (foot) or supra-unit (declination unit, paratone) gestalt	flat or shapeless intonation contours
9	<i>isotony</i>	multi-unit	high	low	high	unit-internal (foot) or supra-unit (declination unit, paratone) isotony	flat or shapeless intonation contours
10	<i>turn</i>	end/start	high	low	high	backchannel, continuer	speech overlap
11	<i>accent count</i>	unit	low	mid	mid	secondary accents	unstressed one-word IU; hesitations, particles
12	<i>register</i>	start	low	low	mid	primary accent	amplitude: microphone distance variability
13	<i>truncation</i>	end	high	low	mid	disfluency	level continuing contour

Sometimes a cue is sufficient (positive implication), but only weakly so. In the same vein, a cue may be necessary (negative implication), but only weakly so, thus generating only a weak implication of absence. In both cases, this has the consequence of weakening the importance of these cues in supporting or rejecting the argument for presence of an Intonation Unit boundary at a given point in a discourse.

Sufficient—low: If a “weakly sufficient” cue is *present*, this is weak evidence in favor of the presence of an Intonation Unit boundary.

Necessary—low: If a “weakly necessary” cue is *absent*, this is weak evidence against the presence of an Intonation Unit boundary.

The assessments of a given cue as sufficient and/or necessary may interact in salient ways to determine its significance. For example, if a cue (such as *breath*) appears in Table 4 with a high value under SUFFICIENT but a low value under NECESSARY, this means that when the cue is present at a given location, the presence of an Intonation Unit boundary at that location is strongly supported; however, if the cue is absent, this does not significantly detract from the likelihood of an Intonation Unit boundary there.

Reliability/Clarity

Cues differ not only in their relative importance in various contexts, but also in how easy they are to recognize. This can be described as CLARITY, with implications for RELIABILITY. Some cues are easier to distinguish in the stream of speech than others. In addition, the quality of the recording may affect the audibility of some cues more than others. Thus the various cues for Intonation Units differ according to the likelihood that a transcriber will be able to obtain good auditory evidence for the presence or absence of the cue. If a cue is assigned a *high reliability* value, this represents a claim that the judgment of cue presence can regularly be made in such a way as to secure a high degree of inter-transcriber reliability. For example, identifying a breath is relatively straightforward (assuming a high quality recording with close microphone placement), but recognizing the presence and scope of a gestalt tune, or counting primary accents, may be subject to transcriber uncertainty. This is likely to yield lower reliability for the cue, as reflected in reduced inter-transcriber agreement. (This is of course subject to verification by psycholinguistic experiment.)

Confound

For some cue categories, there is a risk of mistaking one value for another. Alternatively, a given phonetic value may be interpreted as pertaining to more than one linguistic category or dimension. This can be described in terms of potential *confounds* for a given cue, acknowledging the existence of alternative interpretations (values, categories, dimensions) which can be confused with it. CONFOUND indicates a potential ambiguity, i.e. another function or phenomenon marked by the same linguistic, which can represent an alternative source for the cue (and hence an alternative explanation for its presence).

Neutralizing Context

Some cues have clear significance in one context but not in another. It may be possible to identify specific prosodic, structural, or functional contexts in which a cue’s significance becomes neutralized. This occurs especially when a given phonetic value is subject to alternative interpretations. For example, a cue like high pitch is governed by more than one linguistic system (e.g. boundary intonation and pitch accent or lexical tone), only one of which implies an Intonation Unit boundary. Many prosodic cues (including pitch, duration, amplitude) are notoriously multifunctional, with the result that their

interpretation is highly context-dependent.

Functional Contexts for Intonation Unit Cues

Conditions affecting Intonation Unit production and perception

TRUNCATION. In a truncated Intonation Unit, the end of the unit as originally projected is not actually realized – this is what “truncated” means, after all. As a result, all final cues normally disappear, as well as any cues which pertain to the unit as a whole. This leaves only the initial cues intact. The initial cues allow identification of the beginning of the truncated Intonation Unit in accordance with the usual criteria. But to identify the end of a truncated Intonation Unit requires reliance on the following Intonation Unit’s initial cues (e.g. pitch reset). In addition, most but not all truncated Intonation Units have a truncated word at the end. If present, the cues for word truncation (e.g. missing or shortened phonemic segments, final glottal stop, absence of any prosodic lengthening or lag) may be an indicator of Intonation Unit truncation. However, word truncation is not a sufficient cue for Intonation Unit truncation, since speakers sometimes truncate a word but persist in completing the Intonation Unit anyway.

REPAIR. One common occasion for word truncation is when a repairable is followed by a repair by the same speaker within the same turn, i.e. a self-initiated self-repair. (The term “false start” is sometimes applied here.) But whether the same-turn self-repair is performed within the same Intonation Unit is another matter. Speakers do it both ways. Sometimes they continue in the same Intonation Unit as they introduce the repair. Other times they abandon the problematic (false-start) Intonation Unit – that is, they truncate the repairable unit – and then begin a new Intonation Unit, within which the utterance is redone (“repaired”). It is not uncommon for both strategies to be invoked in immediate succession. The speaker strives to pronounce a particular word, then tries again one or more times within the same Intonation Unit, but finally truncates the unit and then begins a new one, in which the repair is usually done successfully the first time. If a speaker repeats a repairable word or sequence multiple times, how does one tell where (and whether) there is a new Intonation Unit? In a truncated unit, few cues remain viable. The most consistent mark turns out to be pitch reset at the beginning of the next Intonation Unit. In other words, if three repetitions of the same word are performed each at the same pitch, the words are usually being performed as one Intonation Unit. But if between the second and third repetition there is the slight rise in pitch that indicates a reset, we normally conclude that the first two repetitions constitute part of the first (repairable, abandoned) unit, while the (reset) third repetition represents the initiation of a new unit.

Another cue affected by repair is pause. Often a speaker produces a pause between the repairable and the repair. Such pauses are usually very short (often less than two tenths of a second), which helps to distinguish them from two other functional types, the boundary pause and the search pause.

OVERLAP. During simultaneous speaking by two or more speakers, participants often produce their Intonation Units and cues in just the same way as they do when only one speaker at a time is speaking. And discourse participants appear to attend to Intonation Unit cues in the usual way. Transcribers, on the other hand, tend to get distracted by the complexity of interpreting the simultaneous speech, and the task of locating the precise onset and offset of overlap. As a result they may neglect to attend to the Intonation Unit cues which are present, as standardly produced by participants, and may fail to transcribe even boundaries which are marked by a full complement of Intonation Unit cues. In such cases, all that is

required is that the analyst simply attend to the usual cues with unusual care.

On the other hand, overlap tends to motivate restarts by one or more speakers (e.g. seeking to be heard and receive a recipient response under competitive conditions). To the extent that overlap multiplies restarts, possibly multiple times, simultaneous speech presents an increased likelihood of phenomena like truncation, of both words and Intonation Units. In this case the difficulty of boundary identification is real, and does call for careful attention. But in the end it is the standard principles for dealing with truncated units which become relevant in this context, and the overlap doesn't really change anything here (except to make it harder to hear).

Note that simultaneous speech often makes the words of one or more of the overlapping speakers unintelligible (especially in monaural or otherwise poorly miked recordings), but this does not mean that their prosody, and hence their Intonation Unit cues, are inaudible. A common error, upon deciding that the words are indecipherable, is to ignore the prosodic cues which may be reasonably clear. Even when words are obliterated it is usually possible to clearly recognize that, say, the first six syllables were spoken in one Intonation Unit, and the next three syllables were spoken in a second unit.

The one cue that does become less useful in simultaneous speech, however, is turn change. Although it is often said that the entering speaker comes in with "precision timing" at a point where the prior speaker is at a point of possible completion, the exceptions are frequent enough that this cannot be relied on for realistically determining the first speaker's Intonation Unit boundaries.

HYPER-FLUENCY. Sometimes speakers exhibit an especially high degree of fluency, which may be correlated with various functions including extended narration and floor-holding tactics such as rush-through of a point of possible completion. When speakers are hyperfluent for whatever reason, they commonly rush from one Intonation Unit to the next without producing any boundary pause, leaving final lag as an important tempo cue. In the more extreme cases, not even a final lag is produced. This leaves all syllables at the putative boundary uniformly rushed, rendering the lag-rush flux unusable. When all temporal cues have been washed out, what remain are the pitch cues. Thus, the most reliable cue during hyperfluent production is pitch reset. Sometimes this is buttressed by other pitch cues like riff and isotony.

DISFLUENCY. In many ways the greatest challenge for Intonation Unit recognition is disfluent, hesitant speech. We are not talking here about truncation and repair in general, because these are often fluently produced and managed, yielding Intonation Unit boundaries which are more or less unambiguous, given the circumstances. But when a speaker makes multiple pauses for word searches, filled hesitations, self-repairs, and so on (possibly as part of an uncertain speaking style), this can make the recognition of Intonation Unit boundaries more difficult. (This kind of disfluency is likely to be especially prevalent in cases when speakers speak a language which is not their native language, for which they may lack proficiency and/or confidence. However, disfluency is by no means restricted to non-native speakers. It seems likely that what is difficult for the analyst is also difficult for discourse participants, who may have just as much trouble discerning where the intonation boundaries are in such speech. One common case of disfluency which can affect Intonation Unit boundaries is the search for a word or other element, as described below.

SEARCH. This is a common form of disfluency, in which a speaker indicates by various means, including pause, that s/he is taking time to search for a suitable word or other expression or action. It is probably

the commonest reason for the presence of a pause of more than minimal duration *within* an Intonation Unit. Because pause can signal presence of an Intonation Unit boundary, some care must be taken in justifying that a given pause is a search pause rather than a boundary pause. The best evidence for this usually lies in the word immediately preceding the pause in question. In the case of a search pause, the word preceding the pause often lacks the expected cues that would show it to be the final element in its Intonation Unit, such as lag or a final boundary tone. Moreover, for the stretch of speech preceding a search pause, cues which characterize the unit as a whole are not complete, such as tune, accent quantity, and isotony. One common configuration is as follows: the putative search pause is preceded by a series of words which remain on a level pitch, with no boundary tone, no lengthening of the final word in the series, and perhaps no heavily accented word in the series. All of these are normal characteristics of the *initial* and possibly *medial* portions of an Intonation Unit which is still in progress. After the search pause, there is often a heavy accent on the word which fulfills the search. Because the Intonation Unit has been produced in two parts, the overall contour sometimes lacks the coherence that might otherwise have been expected in the absence of the interruption introduced by the search.

BREVITY. Some Intonation Units are very short, consisting of a single word, which may represent a single syllable. This is not particularly uncommon, especially for discourse particles (*well, oh, yes, no*). Even convergent units may consist of a single one-syllable word representing an entire turn (*yes, no, Darryl's thought*). Under such circumstances some cues necessarily disappear, including those which involve a differentiation of the initial and final tempo within the Intonation Unit. Does a one-syllable unit exhibit initial rush, final lag, or neither? With everything compressed onto one syllable, it is often difficult to tell. In such cases, the pitch cues come to the fore, and reset (or tune, isotony, etc.) may be relied on to signal presence (or absence) of a boundary.

Minor Contexts

The contexts listed above arise relatively often, and so have a fairly pervasive consequences for the expression of cues marking Intonation Unit boundaries. In addition there are several less common contexts, which nevertheless may significant, if occasional, effects on the cues.

FINALITY. While the finality effect is relatively minor and infrequent one, it can affect one of the important cues which normally marks an Intonation Unit boundary, namely lag. In some cases where the BOUNDARY TONE/CLOSURE value is FINAL (i.e. a final fall to low [period] or final rise to high [appeal]), it seems that the lag cue sometimes disappears. One possible explanation for this tendency is that where the final contour is strongly displayed, the presence of a boundary is thereby made sufficiently clear, so that there is no need for the use of lag to show the boundary. (Lag is especially useful when there is no pause and the contour is a level continuing one, all of which is easily confounded with the middle of an Intonation Unit.) The observation of a finality effect is preliminary and bears further investigation. (Also worth investigating in this connection is whether such +final, -lag units are regularly marked by +pause.)

CREAKY STYLE. Some speakers have a tendency to use creaky voice pervasively, and this tendency may even characterize speech communities as a whole (e.g. for many speakers of Finnish). If creak is pervasive throughout the speech of an individual, then it loses much of its capacity to mark a particular location as an Intonation Unit boundary.

PITCH RANGE. **

Typology and Universals

The Intonation Unit appears to be a linguistic universal, and even the specific cues which mark its boundaries show strong evidence of cross-linguistic convergence. But the tendency to universal Intonation Unit cues necessarily interacts with language-specific local phenomena within any given language. This is especially true for cues which are capable of marking a contrast between words, that is, for cues which participate in the lexical contrastive system of phonetics/phonology in a given language. The importance of this issue is heightened by the multifunctionality of the key prosodic dimensions of pitch, duration, and amplitude. Not only are these dimensions exploited for multiple functions within the domain of intonation (including Intonation Unit boundary marking), but they are also often exploited, for example, to mark phonemic contrasts between words.

Thus, in addition to the interactions noted as “confounds” noted above, there are larger interactions with the typology of the overall phonological system of any given language, including at least two relevant contrasting types: lexical tone vs. intonation, and phonemic length vs. prosodic lengthening.

Tone vs. Intonation

It is said that more than half of the world’s languages make use of tone to distinguish words or grammatical categories. For such languages, this has sometimes raised the question of whether such languages are thereby precluded from using pitch for intonation, with some scholars attempting to make a typological divide between tone languages and intonation languages. But the evidence now seems clear that there is no conflict between the presence of tone and intonation in a given language, and hence no typological division along these lines is to be expected. **

(Cf. also so-called “Pitch accent” languages [Hyman, LSA 2007])

Phonemic vs. Prosodic Length

**

Universals of Intonation Units

The discussion of cues and their weightings in context is based primarily on my research on American English conversation. However, informal analysis of various other languages suggests that many of the cues are cross-linguistically valid and likely to be universal. Despite their universal tendencies, Intonation Unit cues interact in complex ways with other linguistic systems in any given language (including lexical-phonemic tone, lexical-phonemic length, pitch accent, and so on). The universal tendencies here presented must thus be qualified, for some languages, with additional language-specific statements. **

Conclusions

**

[Revised 4-Sep-2008]

Rhythm and Tunes:

The Intonation Unit in the Structure of Dialogic Engagement

John W. Du Bois
University of California, Santa Barbara
dubois@linguistics.ucsb.edu

Conference on Prosody in Interaction
Potsdam
September 14-18, 2008

Introduction

What's an Intonation Unit?

- (1) The Intonation Unit can be defined as a stretch of speech uttered under a single coherent intonation contour (Chafe, Du Bois)
- (2) Although this characterization gives a glimpse of the big picture as to what Intonation Units are about, it's too simple and too general to represent the whole story.
- (3) (What exactly is a coherent intonation contour? And how do we know when we're hearing one?)
- (4) The devil is in the details. So we look to a more detailed inventory of audible cues for Intonation Units.
- (5) Currently we use more than a dozen audible cues to determine whether an Intonation Unit boundary is present.

Goals

- (6) To specify the cues that mark the boundaries of Intonation Units
- (7) To examine the role of Intonation Units in organizing the structure of dialogic engagement between speakers, specifically in creating prosodic frames that may guide participants' orientation to resonance across utterances.

Transcription

- (8) Each Intonation Unit is written on a line by itself.
- (9) The delicacy level of the transcriptions presented in this talk is *intermediate*, or Level 3+ in the Transcription Delicacy Hierarchy (see Appendices 1 and 2).
- (10) The intonationally simplified transcription allows us to focus on identifying Intonation Unit *boundaries*, setting aside the interesting problem of distinctive *tunes*.
- (11) For more information on the transcription system known as "Discourse Transcription" (e.g. "DT2", the current version as used in this handout), see the Appendix 3 and the web page <http://www.linguistics.ucsb.edu/projects/transcription/representing>
- (12) All examples in this presentation are taken from the Santa Barbara Corpus of Spoken American English (Du Bois et al. 2000-2005), <http://www.linguistics.ucsb.edu/research/sbcorpus.html>
- (13) Most transcription examples have been updated and revised by the present author.

Units and Tunes

- (14) The present paper focuses on identifying the boundaries of *units*, rather than on the array of distinctive *tunes* that speakers may choose from in producing a particular realization of a unit.
- (15) The separation of representational goals (*units* vs. *tunes*) is (1) possible, (2) practically convenient, (3) analytically perspicuous, and (4) theoretically significant.
- (16) The claim is that speakers attend to the recognition and interpretation of unit boundaries as a partly distinct process from the recognition and interpretation of distinctive tunes.

Units

- (17) Notation: one Intonation Unit per line

- (18) (*Deadly Diseases* SBC015: 862.800-880.935)

1	JOANNE;	(H) And she's totally happy now.
2		I mean there's no%— —
3		I mean she's got no excuse.
4		(0.7)
5	KEN;	(YAWN) [(YAWN):: (YAWN)::]
6	JOANNE;	[Absolutely no excuse to be miserable,
7		but] she's miserable.
8		Cause that's just the way she is.
9		(H) It's kind of like you,
10		~Ken.
11		(0.8)
12	KEN;	That's .. not at all like [me ~Joanne].
13	LENORE;	[@:]
14	JOANNE;	No reason to be miserable.
15		(1.0)
16	KEN;	(GULP)
17		(0.7)
18	JOANNE;	You have no reason to be [mise]rab[₂ le].
19	KEN;	[I'm] —
20		[₂ First o]f all I'm not miserable,
21		and se[₃ condly],

Tunes

- (19) The above excerpt is characterized not only by a unit structure (e.g. Intonation Units), but also by distinctive tunes (intonational melodies) on each unit, which may differ from one unit to the next.
- (20) The variety of possible distinctive tunes is suggested by the following excerpts from the above interaction. Each represents a single Intonation Unit, and each can be heard as carrying a distinctive tune. (Many more tunes are possible, of course.)
- (21) (*Deadly Diseases* SBC015: 866.945-868.520)

6	JOANNE;	Absolutely no excuse to be miserable,
---	---------	---------------------------------------
- (22) (*Deadly Diseases* SBC015: 874.225-875.585)

14	JOANNE;	No reason to be miserable.
----	---------	----------------------------

- (23) (*Deadly Diseases* SBC015: 876.995-878.965)
 18 JOANNE; You have no reason to be miserable.
- (24) (*Deadly Diseases*: SBC015: 878.580-880.190)
 20 KEN; First of all I'm not miserable,
- (25) The representation of tunes (intonational melodies) is a topic that deserves a full treatment in its own right, but given time limitations I will not address it today.
- (26) Tunes are not marked in the present transcriptions.
- (27) The only feature in these Level 3 transcriptions that can be considered tune-oriented is the *boundary tone* marked at the end of the line (e.g. comma vs. period, marking continuing vs. final intonation, or *closure* in Bolinger's terms). But this is NOT by any means intended as a complete statement of the tune.
- (28) The "Discourse Transcription" (DT2) transcription system provides for full representation of the distinctive character of tunes for any given intonation Unit; for more information, see Appendix 3; also Appendix 1 (Transcription Delicacy Hierarchy, Levels 5 and 6).
- (29) Even though the representation of tunes is not treated in detail in this paper, tunes remain implicitly relevant to the extent that Intonation Units defines a sequence of unit boundaries, which creates a metrical framework that may contribute to the perception of a "rhythm of tunes", and to resonance between utterances (see below).

Boundaries are meaningful

- (30) (*New Yorkers Anonymous* SBC051: 1519.175-1530.481)
 1 FRAN; I was the world's oldest bride.
 2 I mean,
 3 I waited,
 4 and waited.
- (31) (*New Yorkers Anonymous* SBC051: 1519.175-1530.481)
 1 FRAN; I was the world's oldest bride.
 2 I mean,
 3 I waited,
 4 and waited.
 5 @And @waited.
 6 @and @waited, @@
 7 (H)@ @and @everyone had given up.
 8 (H)@ @@
 9 (H)@: Bu:t,
 10 (0.5)
 11 when I finally did it,
 12 #I'm —
 13 .. I'm (%) —
 14 .. I'm: delighted that I waited.
- (32) (*You Baked* SBC059: 873.918- 878.002)
 2 JO; [2Cause you know you got]ta wait <@> and wait and wait,
 3 for tha:t </@> @@@.
 4 CAM; What do you mean wait [3and wait].

(33) (*Appease the Monster* SBC013: 421.10-422.60)

- 1 MARCI; Now wait,
- 2 wait wait wait wait wait.

Isolation vs. Continuity: Perspectives on the Intonation Unit

(34) *Figure 1.* Isolated sentence model => Isolated Intonation Unit model (“examples”)

Example (a):

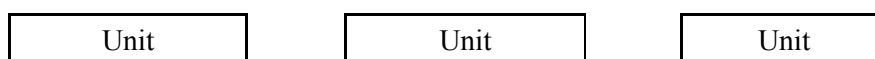
Intonation Unit

Example (b):

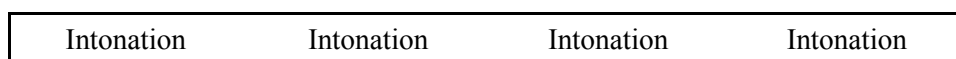
Intonation Unit

(35) The “examples” seem to come with their boundaries already given—how convenient! (but perhaps unreal).

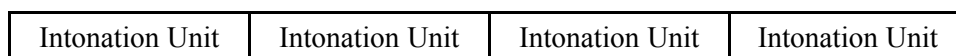
(36) *Figure 2.* Projecting the isolated unit model onto discourse:



(37) *Figure 3.* Towards a more realistic view of participant experience: continuous discourse stream



(38) *Figure 4.* Continuous stream of discourse, with orientation to Intonation Unit segmentation



Some examples

(39) (*A Book about Death* SBC005: 173.55-211.02)

((1st speaker = PAMELA; 2nd speaker = DARRYL))

I mean books words I mean I mean they just become handbooks you distill them and use them in your own way no no no I don't I don't I I come up with my own ideas about that stuff and where do you get the ideas thought where do you get those thoughts processing what goes on around me well isn't a book part of what goes on around you well more from an oblique source

(40) (*A Book About Death* SBC005: 173.55-211.02)

- 1 (0.3)
- 2 PAMELA; °I_mean°,
- 3 .. books,
- 4 (0.3)
- 5 (%) w%- wor:ds.

6 I_mean,
 7 (3.0)
 8 #°I_mean°,
 9 they just become handbooks.
 10 You distill them,
 11 and use them in your own way.
 12 (4.5)
 13 DARRYL; °No°,
 14 (0.6)
 15 No,
 16 .. No I don't.
 17 (0.3)
 18 I don't.
 19 (1.2)
 20 (H) I,
 21 (2.3)
 22 I come up with my own ideas about that stuff.

(41) (*Actual Blacksmithing* SBC001: 302.390-320.880)

when we put em on a horse's hoof all we do they're already made they're round we pick out a size you know we'd like look at the horse's hoof and say okay this is a double aught this is an aught

Unit vs. Boundary Perspective

Unit Perspective

(42) (*Actual Blacksmithing* SBC001: 311.98-315.220)

this is a double aught

(43) *Figure 5*. Intonation Unit Cues: Unit Perspective

		... (H) This is a double au:ght,		
		↑		↓
		START		END
PART	Rush	Lag		COMPLEX Tempo Pitch Pause Laryngeal
	Reset	End-Tone		
	Pause	Pause		
	Breath	Creak		
		UNIT		
WHOLE	Tune Gestalt			
	Isotony			
	Turn			
	Accent Count			
	Register			

COMPLEX
 Tempo
 Pitch
 Pause
 Laryngeal

- (44) *Figure 6. Unit perspective on discourse: Multiple Intonation Units (isolated? autonomous?)*

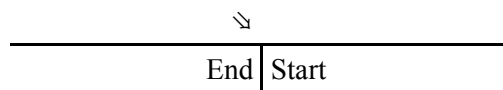


Boundary Perspective

- (45) (*Actual Blacksmithing* SBC001 311.98-315.220)

this is a double aught this is an aught

- (46) *Figure 7. Boundary perspective: Segmenting the continuous voicestream*



- (47) *Figure 8. Intonation Unit Cues: Boundary Perspective*

... (H) *This is a double au:ght, this is an au:ght.*

		↕	↕	
		<i>END</i>	<i>START</i>	
<i>PART</i>		Lag	Rush	<i>COMPLEX</i> <i>Tempo</i> <i>Pitch</i> <i>Pause</i> <i>Laryngeal</i>
		End-Tone	Pitch Reset	
		Pause	Pause	
		Creak	Breath	
		<i>UNIT₁</i>	<i>UNIT₂</i>	
<i>WHOLE</i>		Tune Gestalt	Tune Gestalt	
		Isotony	Isotony	
		Turn	Turn	
		Accent Count	Accent Count	
		Register	Register	

Summary of Intonation Unit cues

- (48) Table 1 summarizes the major cues for Intonation Unit boundaries, organizing them according to the cue complexes they participate in^{4,5}
- (49) SCOPE indicates the position in the Intonation Unit where a given cue is heard:
- INITIAL position (START),
 - FINAL position (END),
 - at a BOUNDARY (either initial or final),
 - spread over the whole UNIT,
 - or requiring comparison across multiple units ('MULTI-UNIT').
- (50) The temporal distribution of the different cues gives internal structure to the Intonation Unit, and hence unit-internal projectability (progressivity; cf. C. Goodwin, Lerner, Raymond)

(51) *Table 1. Principal Intonation Unit Cues*^{6,7}

CUE	COMPLEX	SCOPE	DEFINITION
1. lag	tempo	end	tempo lag or prosodic (non-lexical) lengthening
2. rush	tempo	start	rapid tempo unstressed syllables (anacrusis)
3. end tone	pitch	end	IU-final boundary tone distinguishing intonational finality vs. continuity
4. pitch reset	pitch	start	rise/drop in overall baseline pitch level for IU (esp. on unstressed sylls)
5. pause	pause	boundary	noticeable absence of speech by discourse participants
6. creak	laryngeal	end	creaky voice on final portion of Intonation Unit (not consistent)
7. breath	laryngeal	start	breathing in (and other vocalisms: exhale, throat-clear, sniff, click, etc.)
8. tune gestalt	tune	unit	coherent intonation contour perceived as unified (holistic) gestalt for unit
9. isotony	tune	unit	repeated tunes across sequence of Intonation Units (intonational parallelism)
10. turn start	turn	start	next speaker (new voice) begins
11. turn end	turn	end	current speaker (current voice) ends
12. accent count	prosody	unit	IU “size” in primary accents per IU (tends to be 1, 2, or 0—in that order)
13. register	prosody	unit	overall register shift (of pitch and/or amplitude) for whole Intonation Unit
14. truncation	fluency	end	truncation masks normal end cues, but sometimes is signaled by glottal stop

(52) Although these cues all merit further discussion and exemplification, I will mention one of them beyond those already discussed: *isotony*, which is particularly relevant to dialogic resonance

Isotony

(53) *Isotony* can be defined as the realization of the same tune on successive Intonation Units

(54) Isotony is comparable to, but independent from, *isochrony* (Couper-Kuhlen).

(55) (*Raging Bureaucracy* SBC004: 250.12-269.21)

- 1 SHARON; cause I] made the whole class learn,
- 2 like,
- 3 (H): good morni:ng,
- 4 goodbye:,
- 5 (H) can I go to the bathroo:m,
- 6 can I stand u:p,
- 7 (0.3)
- 8 (H) can I go to the —
- 9 er —
- 10 Time to go to the cafeteria:,
- 11 (0.8)
- 12 (H) an:d,
- 13 (0.2)
- 14 the office,
- 15 you know,
- 16 certain st— terms that related to school.
- 17 (1.0)
- 18 (TSK) (H) And she:,
- 19 like,

(56) (*Constitution*)

- | | | |
|---|-------|-------------------------------------|
| 1 | LONI; | I mean that's not the only one too, |
| 2 | | there was that Paula: Jones? |
| 3 | | (0.6) |
| 4 | | and there was Watergate? |
| 5 | | #and wasn't there another gate? |
| 6 | | (0.3) |
| 7 | | and another: something or other? |

(57) (*Cuz SBC006: 55.43-59.23*)

- | | | |
|---|--------|---|
| 1 | ALINA; | So he got another radio this summer, |
| 2 | | but of course that got ripped off also. |
| 3 | | But never mind. |

What are Intonation Units for?

- (58) Intonation units are complex units of language, which are put to functional use by speakers in processes of cognition, communication, and social interaction.
- (59) The Intonation Unit constitutes a pivotal unit at the intersection of form (phonetic, phonological, syntactic), meaning (semantic, pragmatic), and function (cognitive process, social action, dialogic interaction).
- (60) As a putatively universal category of spoken language, the Intonation Unit represents, arguably, the fundamental basic-level unit in which speakers realize social action through words.
- (61) But there is no reason to think that a fundamental unit like the Intonation Unit will have only one function; multiplicity of function is more likely.
- (62) Functions that have been proposed include:
1. cognitive: chunking for information processing (Chafe)
 2. sequential: shaping turn-construction unit (Ford, Fox, & Thompson)
 3. dialogic: framing the structure of dialogic engagement; enhancing harmonic resonance across utterances (Du Bois)
- (63) Intonation Unit function as (1) delimitative (2) metrical
- (64) Metrical alignment of Intonation Units creates a "rhythm of tunes"

Intonation Units in the Structure of Dialogic Interaction

- (65) One of the key functions of the Intonation Unit, I argue, is to organize the orientation to potential dialogic resonance.
- (66) *Resonance* is defined as the activation of affinities across utterances.
- (67) Dialogic resonance arises when speakers engage with the words of others who have spoken before them, whether in the current conversation or in a prior discourse
- (68) Resonance is always a *relation* that arises between utterances, and is not inherent in any linguistic element considered in isolation.
- (69) Resonance can arise across speakers, or between the utterances of a single individual.
- (70) Intonation Units contribute to the recognition of resonance, by demarcating sequences of units suitable for cross-mapping.

(71) (*A Book about Death* SBC005: 193.14-206.63)

- 1 DARRYL; I come up with my own ideas about that stuff.
 2 (0.7)
 3 PAMELA; And where do you get the ideas.
 4 (0.6)
 5 DARRYL; Thought.
 6 (0.9)
 7 PAMELA; #And where do you get those thoughts?
 8 (0.3)
 9 DARRYL; Processing what goes on around me.
 10 (0.5)
 11 PAMELA; Well?
 12 (0.5)
 13 PAMELA; Isn't: a book part of what goes on around you?

(72) (*diagraph 1*)

- 1 DARRYL; I come up with my own ideas about that stuff .
 3 PAMELA; and where do you get the ideas .
 5 DARRYL; thought .
 7 PAMELA; and where do you get those thoughts ?

(73) (*diagraph 2*)

- 9 DARRYL; processing what goes on around me .
 13 PAMELA; is n't a book part of what goes on around you ?

(74) (*Hey Cutie Pie* SBC028: 73.225-80.730)

- 1 JILL; Yeah,
 2 there was such drama.
 3 There was drama,
 4 and there was suspense.
 5 (H) And then there was relief,
 6 (H) and [then there was] ecstasy.

(75) (*diagraph 3*)

- 2 JILL; there was such drama .
 3 there was drama ,
 4 and there was suspense .
 5 and then there was relief ,
 6 and then there was ecstasy .

(76) (*Ancient Furnace* SBC029: 1507.370-1531.340)

- 1 SETH; So the registers that we're connecting to,
 2 that are existing,
 3 how many of those **we have got**.
 4 **We got the one in the kitchen**,
 5 (2.3)

6 SETH; and then **we've got:** uh,
 7 **one in the d- .. living room here,**
 8 (5.0) ((RADIO TALK IN BACKGROUND))
 9 LARRY; That guy,
 10 (0.8)
 11 LARRY; and then **there's one in the bathroom,**
 12 **there's one in the s- (0.9) study,**
 13 (1.0)
 14 LARRY; and **there is one in the .. library.**

(77) (*diagraph*)

3 SETH; *how many of those* we have got .
 4 we got the one in the kitchen ,
 6-7 and then we 've got uh |one in the d- living room here ,
 11 LARRY; and then there 's one in the bathroom ,
 12 there 's one in the s- study ,
 14 and there is one in the library .

(78) (*Deadly Diseases* SBC015: 849.456-856.005)

1 LENORE; [(TSK) So your] mother's happy now.
 2 (0.2)
 3 JOANNE; (H) My mother's never happy.
 4 [My mother wouldn't be happy if] everything was g- —
 5 LENORE; [Excuse me (Hx):].
 6 JOANNE; (%) everything was great,
 7 and everything is great.

(79) (*diagraph*)

1 LENORE; so your mother 's happy now .
 3 JOANNE; my mother {never} 's happy .
 4 my mother would n't be happy if everything was g- —
 6 everything was great ,
 7 and everything is great .

(80) (*Deadly Diseases* SBC015: 198.625-204.760) (Du Bois et al., 2000, Du Bois et al., 2003)

1 JOANNE; [₂(H) I] [think Mexi]co's like the place: to go:,
 2 KEN; [I don't know].
 3 JOANNE; because **it's got everything.**
 4 (H) **It's got history,**
 5 **it's [got big] [₂cit]ies,**

(81) (*diagraph*)

3 JOANNE; because it 's got everything .
 4 it 's got history ,
 5 it 's got big cities ,

- (82) (*Deadly Diseases* SBC015: 198.625-208.130)
- 1 JOANNE; [2(H) I] [think Mexi]co's like the place: to go:,
 2 KEN; [I don't know].
 3 JOANNE; because **it's got everything**.
 4 (H) **It's got history**,
 5 **it's [got big] [2cit]ies**,
 6 KEN; [Yeah],
 7 [2it's got]—
 8 JOANNE; **it's [3got ruins]**,
 9 KEN; [3**It's got diseases**][4es],
 10 JOANNE; [4and] [5it's got]—
 11 KEN; [5**that I keep**] [6@catching].
 12 JOANNE; [6(H) and it's] got res[7o:rts].

- (83) (*diagraph*)
- | | <i>A</i> | <i>B</i> | <i>C</i> | <i>D</i> | <i>E</i> | <i>F</i> | <i>G</i> |
|------------|----------|-----------|-----------|------------|------------------------------------|----------|----------|
| 3 JOANNE; | because | it | 's | got | everything | | . |
| 4 | | it | 's | got | history | | , |
| 5 | | it | 's | got | big cities | | , |
| 7 KEN; | | it | 's | got | | | — |
| 8 JOANNE; | | it | 's | got | ruins | | , |
| 9 KEN; | | it | 's | got | diseases | | , |
| 10 JOANNE; | and | it | 's | got | | | , |
| 11 KEN; | | | | | <i>that I keep catching</i> | | . |
| 12 JOANNE; | and | it | 's | got | resorts | | . |

Conclusions

Structure in the stream of discourse

- (84) Discourse, including prosody, is produced and experienced by participants as an ongoing stream of actions which partly overlap with and blend into one another, exhibiting fuzzy boundaries par excellence.
- (85) For any unit of importance, participants as well as analysts are faced with the task of recognizing what kind of unit it is, and where it starts, continues, and ends.
- (86) Speakers indexically mark the start and end (or better, the transition into and out of) each Intonation Unit, as well as its internal structure and overall gestalt unity, in more or less consistent ways, within any given language.
- (87) Hearers regularly attend to and interpret these indexes of Intonation Unit boundaries and internal structure.
- (88) Thus for both analyst and theorist of language, knowing how to identify the Intonation Unit within the ongoing stream of discourse, and understanding its structure, gestalt unity, and sequential development, is an essential part of the task of representing, analyzing, and theorizing language as produced and interpreted by participants.

Intonation Unit Cues

- (89) Intonation Unit realizations are characterized by prototype structure, in which some (prototypical) exemplars are marked by a full complement of well-defined cues. Some Intonation Units have

weaker boundaries, marked by a reduced complement of cues.

- (90) For the participant's and analyst's task of identifying Intonation Unit boundaries, the most productive initial orientation is the *boundary perspective*, focusing on the relation between cues at the end of one unit and cues at the start of the next unit.
- (91) The temporal distribution of the different cues gives internal structure to the Intonation Unit, and hence unit-internal projectability.

Dialogic Engagement

- (92) Intonation Units contribute crucially to the perception of dialogic resonance across utterances.
- (93) Metrical alignment of Intonation Units creates a "rhythm of tunes"
- (94) Intonation Units contribute powerfully to the foregrounding and recognition of resonance, by demarcating sequences of units suitable for cross-mapping.

Chapter **: Boundary Intonation

Contrasts

The boundary intonation system involves three binary contrasts of prosodic form and meaning, realized at the end of the intonation unit. The three contrasts are:

final vs. continuing (=closure)
appeal vs. non-appeal
truncation vs. non-truncation

In many approaches to intonation, the first two pairs above (closure and appeal) are treated together as part of a single set of (3 or more) contrasts, including at least: final, continuing, and “question” intonation, and often supplemented by several additional intonational categories. But we have found that a more structured system, with just three binary contrasts, provides a simpler and more robust system for analyzing and classifying the range of prosodic phenomena which characterize the boundary zone. (We will give arguments for the three binary contrasts below.)

For representing these contrasts, four symbols are used, as follows:

<i>final</i>	.	[period, or full stop]
<i>continuing</i>	,	[comma]
<i>appeal</i>	?	[question mark]
<i>truncation</i>	—	[em dash]

These symbols are written at the end of the intonation unit. They appear always in final position, and only there. That is, for every intonation unit, at least one of them is required in final position, and they may occur in no other position. We will describe the three contrasts in turn, beginning with closure and appeal, while reserving truncation for a later discussion.

Closure

Closure is the foremost concept in the system of boundary intonation, which concerns the prosodic marking of intonational meaning associated with the boundary between the end of one intonation unit and the beginning of any following intonation unit. Specifically, the intonational category of closure signals the degree of finality of the current intonation unit in contrast to the degree of continuity in the prosodic transition to the next unit. We have previously discussed this contrast as one of *transitional continuity* {Du Bois, 1993 #293; Du Bois, 1992 #295}, emphasizing the issue of segmenting the ongoing stream of speech. Presently we prefer to adopt the more compact term *closure*, following Bolinger {, 1998 #1099 @48}. Note that Bolinger used the term somewhat more generally to encompass the broader territory of boundary intonation, including such additional categories such as question intonation. Other widespread concepts which address categories of boundary intonation and closure include the Prague school concept of the *delimitative function* (Trubetzkoy), the American structuralist concept of *juncture* (Bloch and Trager), and the related concept of *boundary tone* (Pierrehumbert).

Boundary intonation is of singular importance because of the way it contributes to the identification of units of language structure in use. In discourse, intonation units are not typically found in isolation, but

occur as part of an ongoing stream of talk. The stream of speech is characteristically fluid and is marked by a dominant perception of continuity. The continuity of the stream is from time to time punctuated or delimited by an indication of discontinuity, to a greater or lesser degree. While every intonation unit boundary represents a segmentation of this stream—it is, after all, a prosodic boundary—intonation unit boundaries are not all alike. They differ in character, with some marking more finality, and others more continuity, associated with the intonation unit boundary. Note that this is not just a matter of weak vs. strong boundaries: one can have a very strong intonation unit boundary with a strong marking of continuity. It may also happen, though less often, that a well-defined intonation boundary carries a weak indication of finality. So closure is in principle independent of the strength of the intonation unit boundary (though in practice there is naturally some correlation).

Closure can be seen as related to the function of chunking language into discrete units. While there is a popular notion that units in speech are naturally demarcated by pauses, the reality of fluent speech undercuts this notion. The fact is, the intonational system of closure allows us to demarcate units on the fly, while maintaining a rapid and uninterrupted stream of speech. According to Bolinger, “Children learn closures in association with pause, which is simply “separation” in speech corresponding to separation of ideas; adults may continue to imagine a pause where the intonation behaves as if the pause were still there.” {Bolinger, 1998 #1099 @48 chk}

As a first approximation, consider the following definitions of boundary intonation categories:

Final. A prosodically distinct class of contours characteristically used to mark **finality**.

Continuing. A prosodically distinct class of contours characteristically used to mark **continuation**.

Appeal. A prosodically distinct class of contours characteristically used to mark **appeal**.

Truncation. A prosodically distinct class of contours characteristically used to mark **truncation**.

What exactly is meant by a term like finality, or continuity? Finality of what? Continuity of what? Despite our desire for precise answers to such questions, the way speakers actually use intonational meaning is often broad and open-ended, leaving the specific interpretation of the terms to context-based pragmatic interpretation. In general, we can say that the “closure” meaning applies to whatever linguistic activity is most saliently in progress during the utterance at hand. As Bolinger observes,

Though essential to the segmentation of speech and to that extent “grammatical”, closures are basically metaphorical: a fall is “finality” in any sense (end of a series, end of a main part, “nothing more worth saying” hence positiveness); a rise or high pitch is non-finality in any sense (“not through speaking”, “answer my question”, “incomplete utterance”, “too excited to calm down”, “give me feedback”).
{Bolinger, 1998 #1099 @48}

It’s not that the prosodic configurations heard at intonation unit boundaries actually signal all of these diverse meanings, but that the closure contours distinguish a few broad meanings, each of which is sufficiently open as to be compatible with a variety of distinct uses, depending on the context. Nevertheless, what remains constant is the association of closure marking with distinct points along a

scale of finality, ranging from meanings of definite finality to unimpeded continuity. Bolinger interprets the closure scale as gradient:

Each of these pitch movements is graded: a deeper fall makes for greater finality (for example end of a paragraph as against end of a sentence), and a fall that ends on a level non-low tone may resemble a rise as a signal of non-finality. A higher rise makes for higher excitement, a more demanding question, a wider separation, *etc.*
{Bolinger, 1998 #1099 @48 chk}

Bolinger notes the use of punctuation marks in conventional written language as a crude representation of closure in spoken language, with “period for fall, question mark for rise, ... comma for slight rise or sustention...” {Bolinger, 1998 #1099 @48}. While we draw on these same conventions, we specify a strict interpretation of each symbol in terms of the distinctive class of contours it represents.

General Definitions

But first it is necessary to clarify the nature of the definitions of these categories, distinguishing “general” vs. language-specific definitions of prosodic categories.

The definitions of intonational categories presented in this section are intended as *general definitions* (in the sense of {Huddleston, 2002 #1098 @32}). QUOTE THEM That is, they are definitions which foreground functional contrasts between elements of the system, while maintaining some independence from the particular details of realization in a specific language. As such they represent categories with potential for application at two levels: first, in the functional analysis of a given system of intonational meanings and contrasts within one language; and second, in typological analysis and comparison across languages. Such general definitions must differ from language-specific ones because different languages show a degree of variability of prosodic systems, even as they mark similar contrasts in meaning. While there may be strong tendencies toward universal convergence on a small set of prosodic features marking these contrasts, it still must be recognized that in any given language the specific details of phonetic realization may vary in some degree. If nothing else, such variation in intonational realization is often motivated by competition among the several linguistic systems which govern the use of pitch. Among the systems in competition with intonation are those governing lexical tone and pitch accent. Indeed, there is probably even competition within different subsystems of intonation, such as the subsystem for boundary closure and that work tune.

Markedness

Only four symbols are needed to mark boundary intonation, for two reasons. First, all three categories (closure, appeal, truncation) represent binary contrasts pertaining to one unified issue—how the intonation unit boundary is to be heard and classified. Thus all the symbols for boundary intonation are written at the same location, at the final boundary of the intonation unit. Second, while some of the symbols may occur in combination (e.g. *appeal* may co-occur with *closure*, giving the combinations *final appeal* vs. *continuing appeal*), not all combinations of the contrasts do occur (e.g. *truncation* doesn’t co-occur with any other contrast). As a result, four symbols are sufficient to distinguish all the occurring combinations.

The issue of occurrence vs. non-occurrence of the various combinations is linked to the issue of markedness. For each of the boundary tone contrasts listed, the first member (cited here in boldface) represents the *marked* member of the opposition:

final vs. *continuing*
appeal vs. *non-appeal*
truncation vs. *non-truncation*

In saying that the boldface items are marked, we mean that each is the less *expected* of the two paired values, or the more *salient*. This tends to go along with lower frequency. The marked member also tends to have a more specific, narrow meaning than the unmarked member. Reversing the perspective, we can say that the *unmarked* (more frequent, more common) pair, has a ****

As an indication of dysfluency, truncation signals that the projected intonation contour for the current unit is not to be realized in full. “Non-truncation”, then, is the norm.

Boundary Tones: Closure/End-Tone

Continuing vs. Final

(1) (*Hey Cutie Pie* SBC028: 73.225-80.730)

- | | | |
|---|-------|-----------------------------------|
| 1 | JILL; | Yeah, |
| 2 | | there was such drama. |
| 3 | | There was drama, |
| 4 | | and there was suspense. |
| 5 | | (H) And then there was relief, |
| 6 | | (H) and [then there was] ecstasy. |

(2) (*Ancient Furnace* SBC029: 1507.370-1531.340)

- | | | |
|----|--------|--|
| 1 | SETH; | So the registers that we're connecting to, |
| 2 | | that are existing, |
| 3 | | how many of those we have got. |
| 4 | | We got the one in the kitchen, |
| 5 | | (2.3) |
| 6 | SETH; | and then we've got: uh, |
| 7 | | one in the d- .. living room here, |
| 8 | | (5.0)_((RADIO TALK IN BACKGROUND)) |
| 9 | LARRY; | That guy, |
| 10 | | (0.8) |
| 11 | LARRY; | and then there's one in the bathroom, |
| 12 | | there's one in the s:- (0.9) study, |
| 13 | | (1.0) |
| 14 | LARRY; | and there is one in the .. library. |

(3) (*Appease the Monster* SBC013: 421.10-422.60)

1 MARCI; Now wait,
2 wait wait wait wait wait.

(4) (*You Baked* SBC059: 873.918- 878.002)

2 JO; [Cause you know you got]ta wait <@> and wait and wait,
3 for tha:t </@> @@@.
4 CAM; What do you mean wait [and wait].

(5) (*New Yorkers Anonymous* SBC051: 1519.175-1530.481)

1 FRAN; [I was] the world's oldest bride.
2 I mean,
3 I waited,
4 and waited.
5 @And @waited.
6 @and @waited, @@
7 (H)@ @and @everyone had given up.
8 (H)@ @@
9 (H)@: Bu:t,
10 (0.5)
11 when I finally did it,
12 #I'm —
13 .. I'm (%) —
14 .. I'm: delighted that I waited.

[Revised 14-Oct-2008]

Chapter **: Vocalism

as a Discourse-Functional Category

Vocalism represents an important, yet surprisingly little understood, category of phenomena in the realization of spoken language. Although individual instances of vocalism have been noticed recurrently in the literature, and several have attracted considerable attention in their own right, it must be said that in general the awareness of the category's members has been scattered and unfocused; the methodological treatment and representation in transcriptions has been haphazard; and most problematically, appreciation of their categorial coherence and hence larger theoretical significance has been limited in scope. Typically, each new vocalism encountered is seen as a unique phenomenon unto itself. Conversely, there is little awareness of the commonalities of form and function that would unite vocalisms in a natural class. It remains a challenge to recognize a unified category which is capable of elucidating the general significance of vocalisms in the realization and interpretation of spoken language in use.

This paper seeks to describe in general terms the phenomenon of vocalism; to clarify its coherence as a discourse-functional category; and to explore the significance of vocalism for understanding spoken language in its own terms, as speaking that is embedded in the embodied actions of speakers engaged in interaction, and which is integrally designed to fit with their capabilities, needs, and goals. While a full exposition of a category as broad, diverse, and open-ended as vocalism is beyond the scope of this paper, and its individual members are so numerous that we will be able to do little more than mention, I will present enough examples to give an idea of the scope and significance of the class. The emphasis is on the coherence of the category as a whole, and its unified role in the functional realization and organization of language in use.

Vocalisms can be defined as sounds which are (1) produced with the same apparatus that produces speech (i.e. the human vocal tract), but (2) not part of speech per se. Because vocalisms are made where language is made, they alternate with language in the voicestream of a given individual speaker. As a consequence, discourse participants seem to pay very close attention to vocalisms. Vocalisms are interpreted as sharing the same timestream that organizes speech and other aspects of spoken interaction. This gives them a special significance for sequential organization of interaction. The rich diversity of vocalism categories, including laughter, inhalation, exhalation, clicks, sniffs, coughs, and so on, allows them to carry important information about the state of the person who is producing the vocalism. For all these reasons it is important to have an effective way to represent vocalisms so as to capture their structural and functional roles in language use.

For example, some vocalisms act as an important cues to turn-taking, as when a participant makes an audible in-breath just before starting to speak. Because this practice is so regular, audible in-breath becomes a way for speakers to actively signal that they are seeking to begin a turn. While breathing might seem to be motivated by simple necessity, it is necessary to distinguish audible from inaudible breathing, because of the the different interactional significance of each. The fact is that, barring respiratory problems or conditions of physical exertion, breathing is often effectively inaudible (or at least very quiet). Conversational participants are able get most of their breathing done without making significant audible noise—except when they are about to take a turn, when the in-breath is often done in a way that makes it audible. This gives special meaning to the audible in-breath when it is performed.

The fact that the prospective speaker is *choosing* to make his or her in-breath audible often signals a cue for impending entry into speaking. Breathing has other roles as well (as a reliable indicator of intonation unit boundaries, for example). This is just one example; other vocalisms may also turn out to have functions which go beyond their obviously utilitarian ones (think of the potential for the significant cough).

Frequency

The following tables presents the most important vocalisms by frequency, based on those used in transcribing the Santa Barbara Corpus of Spoken American English. These conventions represent most of the vocalisms that a transcriber is likely to encounter. Table 1 lists vocalisms by overall frequency. The frequencies cited below give the number of occurrences in Parts 1 & 2 of the Santa Barbara Corpus of Spoken American English (roughly 100,000 words). (Note that this table, as well as Table 2 below, represent vocalisms considered as discrete vocal events, written as separate “words”. As such they do not include vocal modifications of words, such as a glottalized cut-off of a truncated word, or laughter during a word.)

Table 1. Top ten vocalisms, by frequency

5474	@	laugh (all forms)
4934	(H)	in-breath (audible inhalation)
549	(TSK)	alveolar click
474	(%)	glottal/laryngeal constriction
450	(Hx)	out-breath (audible exhalation)
158	(SNIFF)	sniff
156	(AHM)	hemming and hawing (“ahem”), or throat clearing (THROAT)
115	(COUGH)	cough (one per cough)
54	(LAUGH)	generalized laughter (e.g. of a crowd)
52	(GULP)	swallowing sound

Type

It’s useful to classify vocalisms according to type, in order to get a more general picture of how they function. One such classification distinguishes the following categories, according to the manner or location in which the sound is made (laugh, air, larynx, click, nose) or the type of iconic or indexical sign involved (onomatopoeia, expressives). . Table 2 groups the various vocalism categories into these broad types.

Table 2. Vocalisms by type

LAUGH	AIR	LARYNX	CLICK	NOSE	ONOM.	EXPRESSIVE
4538 @	4934 (H)	474 (%)	549 (TSK)	158 (SNIFF)	4 (POW)	2 (SCREAM)
882 @word	450 (Hx)	156 (AHM)	10 (KISS)	9 (SNORT)	3 (BBBB)	2 (YELL)
54 (LAUGH)	25 (SIGH)	115 (COUGH)	5 (TCHK)	7 (SNEEZE)	1 (BUZZ)	1 (HOWL)
	8 (GASP)	52 (GULP)		2 (SNORE)		1 (GROWL)
	7 (BLOW)	27 (YAWN)		2 (SNUFFLE)		3 (WHISTLE)
		4 (GROAN)				
		3 (CHOKE)				
		3 (SOB)				
		2 (GRUNT)				
		2 (BURP)				

		1 (BELCH)				
5474	5424	839	564	178	8	9

NB: This table does not include various crowd sounds, (MURMUR), of which there were altogether 43. Also excluded are actions ((EAT)) ((DRINK)), of which there are about 30, and animal calls ((BARK)), totaling 7.

Examples

The following examples illustrate some of the key issues that arise in representing vocalisms.

(1) (*A Book about Death* SBC005: 828.73-835.67)

- 1 PAMELA; [My] favorite word when I was twelve,
 2 (1.5)
 3 was par[adox].
 → 4 DARRYL; [(YAWN)]:: <YAWN> Why </YAWN>.

(2) (*Constitution*)

- 1 MAUDE; (SNUFFLE) (SNUFFLE) (SNUFFLE)
 2 (2.2)
 3 Hm.
 4 (0.9)
 → 5 BELLA; <%> A::h </%> [(Hx)],
 6 MAUDE; [Maybe I'm allergic to] shrimp.

(3) (*Doesn't Work in This Household* SBC019: 687.625-701.384)

- 1 JAN; (YAWN)::: <DUR=3.0>
 2 BRETT; You could practically do that with a quarter-sized violin.
 3 (0.6)
 4 Fit it in your hand.
 → 5 JAN; (SNIFF)
 6 (0.8)
 → 7 (SIGH)
 → 8 (TSK) (H) Bob came by this afternoo:n?
 9 FRANK; Bob who.

(4) (*Risk* SBC024: 738.960-752.380)

- 1 DAN; So it's just you and me babe.
 2 (1.3)
 3 JENNIFER; What's new.
 4 (1.2)
 5 DAN; @@@@ @
 → 6 JENNIFER; (KISS)
 7 DAN; (H):[:]

→ 8 JENNIFER; [(KISS)]
 9 (0.7)
 → 9 (KISS)
 10 (1.1)
 → 11 DAN; (KISS)
 12 (0.6)
 13 (TSK) (H)
 14 (0.8)
 15 Oh,
 16 let's see.

(5) (*Cuz* SBC006: 1560.33-1567.93)

1 ALINA; He's real wishy-washy.
 2 (0.2)
 → 3 (H) (SNIFF) He's too: uh,
 4 .. namby pamby.
 5 (0.4)
 6 You need somebody to come in and go,
 → 7 .. (POW).
 → 8 .. (POW).
 → 9 .. (POW).
 10 (0.4)
 11 LENORE; Yeah[:].

The Vocalism Event: Iconic Representations

It is useful to represent in a quasi-iconic manner the sound of vocalism events as they are experienced by participants. Each distinct audible event token in the recording should be represented by its own distinct notational token. Thus instead of writing a transcriber's comment to the reader, of the form “((Jennifer yawns three times))”, it is preferable to represent the event in a way that more iconically reflects the production of sound by the speaker, as it would be heard by participants, with a separate symbol for each yawning event:

(6) (*Risk* SBC024: 358.822-367.589)

1 JENNIFER; Every time you t]ake over a country you get cards.
 2 (2.3)
 → 3 (YAWN): (YAWN) (2.0) (YAWN)

In general, the idea is that each vocalism notation token, such as (YAWN), should represent the actual concrete sound event itself, rather than presenting merely an abstract description of the sound, or of the action that produces the sound. For this reason, a notation that captures a more iconic evocation or “image” of the sound event is to be preferred. This is one reason why we try to use words which are

themselves more or less iconic or sound-symbolic (yawn, sniff), and we avoid inflecting them grammatically (i.e. avoiding “yawning” or “yawns”). This representational practice follows the usual practice of cartoonists, who have developed a very effective set of informal notational conventions for evoking the audible form of vocalism events, which are important to represent because of what they tell you about the speaker in the moment.

Vocalisms vs. Speech

Vocalisms should be clearly differentiated from speech in a transcription (again, cartoonists follow this practice rigorously). This is because participants hear vocalisms as different from lexical words, and respond to them in very different ways. In the following example, there is a lexical word representing a (fake) laugh, written “ha”:

- (7) (*Appease the Monster* SBC013: 1064.05-1067.38)
- | | | |
|-----|---------|--------------------------------------|
| 1 | KENDRA; | you won't make] half of what I make. |
| → 2 | | Ha ha. |
| 3 | | (H) |
| 4 | KEN; | .. Wha:t? |

This fake ha-ha laugh cannot be considered a true vocalism. As such it contrasts with a real laugh, which is written as a vocalism using the @ sign (with one @ for each pulse of laughter):

- (8) (*Risk* SBC024: 1126.200-1134.154)
- | | | |
|---|-----------|-----------------------------------|
| 1 | JENNIFER; | <VOX> Aw </VOX>. |
| 2 | | (0.3) |
| 3 | DAN; | [% (H) (H)] |
| 4 | JENNIFER; | [(H) <VOX> I'm s::o hurt </VOX>]. |
| 5 | DAN; | @@@@@@@@@@ |
| 6 | | (0.3) |
| 7 | | (KISS) |
| 8 | | (0.5) |
| 9 | | (H):: |

As another example, sometimes a speaker will speak the word “cough” in a sort of pretense of coughing, but in a way which clearly involves the lexical word “cough” rather than the sound of an actual cough. The lexical word “cough” is written in normal case (the standard mix of upper- and lower-case letters), rather than in all capitals as for the actual sound of a cough (COUGH):

- (9) (*Appease the Monster* SBC013: 1056.23-1063.33)
- | | | |
|-----|---------|---|
| 1 | KENDRA; | [(H) I'll be making] well over seventy, |
| 2 | | I'll be making a lot more than hi:[m]. |
| 3 | KEN; | [O][₂ :h]. |
| → 4 | MARCI; | [₂ Cough], |
| → 5 | | choke, |
| → 6 | | sputter, |

7 ...
8 Well over [seven][,ty:]?

One piece of evidence that an act like kissing matters to participants is that participants may occasionally refer to the act overtly. Sometimes the word kiss will occur both as an action, and as a participant's description of the action. In the following example, the lexical word "kissing," written in normal case, appears shortly after an actual kiss, represented as a vocalism in all caps:

- (10) (*Risk* SBC024: 1069.186-1099.360)
- | | | |
|------|-----------|---|
| 1 | JENNIFER; | (H) (SIGH) |
| 2 | | Now, |
| 3 | | if I can just get to the next — |
| 4 | | (5.1) |
| 5 | DAN; | ((MOUSE-CLICKS)) |
| → 6 | JENNIFER; | (KISS) |
| 7 | | (1.2) |
| → 8 | JENNIFER; | (KISS) |
| 9 | | (1.1) |
| 10 | DAN; | ((MOUSE-CLICKS)) |
| 11 | | Thank you very much. |
| 12 | | (0.9) |
| 13 | JENNIFER; | For what. |
| 14 | | (0.2) |
| 15 | DAN; | I just took over Iceland. |
| 16 | | (0.3) |
| 17 | JENNIFER; | (TSK) Oh:, |
| 18 | | f:uck. |
| 19 | | (1.0) |
| → 20 | JENNIFER; | (H) Right while I was kissing you, |
| 21 | | you: traitor. |

Since the set of possible vocalisms is more or less open-ended, and even the more frequent vocalisms are quite varied, it becomes a practical necessity to distinguish them mnemonically using words, rather than giving each a unique symbol. This also allows transcribers to invent new vocalism notations as needed. But in choosing these notations, it is preferable to draw on words from very basic vocabulary which are short, monosyllabic, onomatopoeic, and uninflected. Such words are easiest to "hear" as if they represent the actual audible event they stand for.

Vocalism notations should be used consistently. Once it is decided to represent the sound of swallowing by (GULP), for example, this notation should be used throughout one's transcriptions.

One outstanding representational issue concerns the choice of vocalism notation for the transcription of languages other than English. Depending on the publication conditions and the target audience for one's transcriptions, it may be necessary to develop vocalism notations appropriate to the language you are transcribing, using the appropriate words in the respective language. For example, you could use the word in the target language that onomatopoeically captures the yawn, cough, sniff, or other vocalism

category.

[rev. 5-Jun-2006]

Chapter **: In a Manner of Speaking

When people speak, most of the time they use their “own” voice—the quality of voice which is recognized as typical for that person. This seems so normal that we are scarcely conscious of hearing anything special in their speech, anything that might stand out as noteworthy, a distinctive quality of voice. The impression is that this represents no particular manner of speaking: it’s just talk. But sometimes speakers do something different, taking on a special manner of speaking, which may carry some kind of meaning and potentially lead to interesting consequences. If we perceive that a speaker is employing some distinctive quality in their speech, we hear this as introducing an element of significance, one which is likely to prove relevant to the framing and interpretation of the utterance at hand. One reason to pay attention to an individual’s way of speaking a particular utterance is that this manner often comes as an accompaniment to something else, such as a literal message conveyed in words. The manner of speaking represents a metamessage, an indexical message which bears a significant relation to the other, literal message. In such cases the manner of speaking represents a second channel of vocal information which rides along on the main channel of the spoken word, but which can change how we understand those words. The manner modifies the message, as it were. We hear one voicestream, but the voicestream carries two layers of meaning. A single utterance is characterized by both the words and the distinctive manner in which they are spoken. It is this second layer of voice and its significance that specifically concern us here.

To understand the significance of the use of special voice qualities or other manners of speaking, we need to know how to observe when, how and why speakers use the potential of their voices in special ways. This will only have meaning if we attend equally to the consequences which the vocal shaping yields, in the domains of meaning and interaction.

At the most basic level, this study requires us first to be able to locate and describe the various manners of speaking. Beyond this, we must be able to situate them in relation to the other aspects of speaking, as they unfold simultaneously in the flux of discourse.

This implies that we should develop tools for representing the phenomena in question. How to proceed? The speaker’s freedom to speak in a distinctive manner seems at first to vary in unpredictable ways, unlimited in its diversity and open-ended in its typology. While the first impression is indeed thus, closer examination suggests that there are ways to bring this complexity into the realm of the apprehensible—what can be pointed to, categorized, represented.

Notation

The analytical objective demands notations sufficient to the task at hand. Notations are required that will allow the words and the manner in which they were spoken to be represented clearly, efficiently, and in correlation with each other. A special manner may extend over an indefinite region of the voicestream, so the symbols for representing it must be equally flexible in their application. And given that the number of manners that speakers may invoke is open-ended in principle, it is necessary that the notational conventions be generative, allowing for the innovation of new categories and new symbols as the situation demands.

The most effective way to notate manner of speaking involves a pair of symbols, written so as to bracket

the start and end of the feature in question. To make clear that the manner notation is not to be read as words spoken, by convention the manner is written in capital letters and enclosed in angle brackets.

For example, the phenomenon of *vox*, which is when a speaker adopts the “voice of another,” is written as follows: <VOX> *words* </VOX>. Note that the notations for the *start* and *end* of the indicated manner differ slightly: the end notation is written with a slash character immediately following the left angle bracket.

(1) (*Cuz* SBC006)

ALINA; Nothing was gonna help her.
 No makeup,
 no nothing.
 Cause she's the little <VOX> **girl,**
 and he's the older man,
 and he's taking care of me </VOX>.

(2) (*Conceptual Pesticides* SBC003) (Simplified transcription)

ROY;	There's a bad smell in here,	
MARILYN;	There is?	+3
ROY;	No it's the smell of fish.	+4
	<SING> The smell,	
	of fish,	+2
	is in,	
	the fridge </SING>.	
MARILYN;	(LAUGH)::	+1
	<VOX> Oh God </VOX>.	
PETE;	(LAUGH)::	

While the array of manner categories is diverse and in principle open-ended, there are some manners that recur with a relatively high frequency. It is useful to standardize the notation used for such high-frequency manner phenomena. Table 1 presents frequency counts for the Santa Barbara Corpus of Spoken American English, which gives a good indication of how often these are likely to be encountered by transcribers.

Scope

The scope of a manner notation (like <SING> *words* </SING>) is to be read as applying to the stream of speech of one particular speaker, the speaker in whose labeled turn the manner notation first appears. The manner notation does not apply to any other speakers who may happen to speak in the same stretch of conversation. In the following example, only one speaker begins to sing, the other does not:

1	490.647	492.087	BRETT;	Approximately: here,
2	492.053	494.881		... <SING> Du du du: ,
3	494.881	495.728		du du du du.
4	495.728	497.557	FRANK;	And staff meeting tomorrow after school,
5	497.540	498.574	BRETT;	Duh da-duh: ,

6 499.454 501.317 [hm hm hm hm hm hm </SING>].

Here, Brett's turn begins in a normal speaking voice with the words *Approximately here*, but he soon starts to sing beginning with the first syllables uttered in line 2 (*du du du*), as marked by the onset marker <SING> placed before them. Brett's singing continues until line 6, where its termination is marked by </SING> at the end of the line. Note that it is only necessary to use a single pair of brackets indicate the singing, one bracket at the start and one at the end. But the bracketed notation for singing does not apply to Frank, whose words *And staff meeting tomorrow after school* in line 4 are spoken in a normal, not a singing, voice.

The convention whereby a manner notation applies only to the stream of speech of the speaker within whose labeled turn it is written is based on the reality that in general, each speaker chooses their own manner of speaking. Moreover, the speaker's manner may continue for several lines and even several turns. Other speakers in the same stretch of speech (i.e. those who may interrupt with simultaneous speech, or exchange turns) are not particularly likely to choose the same manner of speaking as the first speaker.

Of course there are occasions when several people do choose the same manner at the same time, as for example in choral singing. In such cases, it is necessary to indicate the manner (e.g. singing) for each participant individually. In the following example, Kendra begins singing and then Marci joins her, while Wendy continues using an ordinary speaking voice:

1	KENDRA;	<☺><SING> My mama loves me cause she [spanks me: </☺>],
2	WENDY;	[That's why I take] [2all my al-] --
3	KENDRA;	[2Woo][3-oo,
4	MARCI;	[2<SING> Woo][3-oo,
5		(H) .. woo-oo </SING>3].
6	KENDRA;	(H) .. woo-oo </SING>3].
7	KENDRA;	.. (H) [4@@@]
8	WENDY;	[4All of our] alterations go to Edna,
9	MARCI;	(H)

Table 1. Manner Categories by Frequency. Frequency counts for Santa Barbara Corpus of Spoken American English, Parts 1 & 2. (Counts are based on the published Linguistic Data Consortium version, which use “DTI” [Du Bois et al. 1993] format.)

((NB: <@> (laughter) is treated elsewhere.))

<i>DT 1</i>	<i>Quantity</i>	<i>DT 2</i>	<i>Alternate</i>
<X>	510	#	<#>
<VOX>	248	« »	<VOX>
<Q>	134	<Q>	“ ”
<WH>	81	□	<WHISPER>
<P>	66	°	<P>
<HI>	49	↑	<HI>
<MRC>	36	<MARCATO>	
<READ>	33	<READ>	
<SM>	33	<☺>	<SMILE>

<L2>	32	<L2>	<L2=SPANISH>
<SING>	26	<SING>	
<FOOD>	20	<EAT>	
<%>	21	<%>	
<F>	19	<F>	
<PAR>	13	<PAREN>	
<YWN>	12	<YAWN>	
 	6	<BREATHY>	
<A>	5	<ALLEGRO>	
<SHOUT>	3	<SHOUT>	
<SMOKING>	2	<SMOKE>	
<ACC>	1	<ACCELERANDO>	
<CRY>	1	<CRY>	
<TALK>	1	<TALK>	
<HUMMING>	1	<HUM>	
<DRINKING>	1	<DRINK>	
<ACCENT>	1	<VOX>	
<SLUR>	1	<SLUR>	

Open-Class Categories

Manner: Voice

laugh voice	<@>	</@>	
smile voice	<☺>	</☺>	marks smiling during word (for seven+ words)
smile voice	☺ you're ☺ kidding		smile symbol marks smiling during word
smile voice	<SMILE>	</SMILE>	alternate
voice of another	<VOX>	</VOX>	
voice of another (alt.)	«	»	
voice of another (specify)	<VOX=DUMB>		
quotation quality	<QUOTE>		
whisper	☐word word☐	<☐> word word </☐>	
whisper	<WHISPER>		alternate
breathy	<BREATHY>		
creaky	<%>		
yawning	<YAWN>		
reading	<READ>		
singing	<SING>		
shouting	<SHOUT>		
crying	<CRY>		

humming	<HUM>
slurred speech	<SLUR>

Manner: Prosody

forte, loud	<F>
piano, attenuated	<P>
piano, attenuated (alt.)	<°>
high pitch (region)	<HI>
low pitch (region)	<LO>
wide pitch	<WIDE>
narrow pitch	<NARROW>
marcato speech	<MARCATO>
fast speech	<ALLEGRO>
slow speech	<LENTO>
increasingly fast speech	<ACCELERANDO>
increasingly slow speech	<RALENTANDO>
parenthetical intonation	<PAREN>

Manner: Miscellaneous

unintelligible speech	<#>	
code-switch/2nd language	<L2>	
code-switch (lg. specified)	<L2=DUTCH>	alternate notation
code-switch (lg. specified)	<LANGUAGE=DUTCH>	alternate notation
affect	<AFFECT>	
affect, specified	<AFFECT=SAD>	sad, angry, disgust, fear, happy
eating, food in mouth	<EAT>	
drinking	<DRINK>	
smoking	<SMOKE>	
talking on telephone (alt.)	<TEL>	
media speech (alternate)	<MEDIA>	

Short form/Long form: by Category

<i>Meaning</i>	<i>Short Form</i>	<i>Long Form</i>
[Manner]		
laugh voice	<@>	<MANNER=LAUGH>
laugh voice	@word @word	<MANNER=LAUGH>
smile voice	<☺>	<MANNER=SMILE>
smile voice	☺ word ☺ word	<MANNER=SMILE>
smile voice	<SMILE>	<MANNER=SMILE>
whisper	<α>	<MANNER=WHISPER>
whisper	<WHISPER>	<MANNER=WHISPER>
breathy	<BREATHY>	<MANNER=BREATHY>
creaky	<%>	<MANNER=CREAKY>
yawning	<YAWN>	<MANNER=YAWN>
reading	<READ>	<MANNER=READ>
singing	<SING>	<MANNER=SING>
shouting	<SHOUT>	<MANNER=SHOUT>
crying	<CRY>	<MANNER=CRY>
humming	<HUM>	<MANNER=HUM>
slurred speech	<SLUR>	<MANNER=SLUR>
[Vox]		
voice of another, unspecified	<VOX>	<VOX>
voice of another (alternate)	« »	<VOX>
voice of another, specified	<VOX=DUMB>	<VOX=DUMB>
voice of another, specified	<VOX=SISSY>	<VOX=SISSY>
voice of another, specified	<VOX=WHINE>	<VOX=WHINE>
voice of another, specified	<VOX=CHILD>	<VOX=CHILD>
voice of another, specified	<VOX=FLIRT>	<VOX=FLIRT>
quotation quality	<QUOTE>	<VOICE=QUOTE>
[Prosody]		
forte, loud	<F>	<AMPLITUDE=FORTE>
piano, attenuated	<P>	<AMPLITUDE=PIANO>
piano, attenuated (alt.)	<°>	<AMPLITUDE=PIANO>

Meaning**Short Form****Long Form**

high pitch (region)	<HI>	<PITCH=HIGH>
low pitch (region)	<LO>	<PITCH=LOW>
wide pitch	<WIDE>	<PITCH=WIDE>
narrow pitch	<NARROW>	<PITCH=NARROW>
fast speech	<ALLEGRO>	<RHYTHM=ALLEGRO>
slow speech	<LENTO>	<RHYTHM=LENTO>
increasingly fast speech	<ACCELERANDO>	<RHYTHM=ACCELERANDO>
increasingly slow speech	<RALENTANDO>	<RHYTHM=RALENTANDO>
marcato speech	<MARCATO>	<RHYTHM=MARCATO>
parenthetical intonation	<PAREN>	<PROSODY=PARENTHETICAL>

[Affect]

affect, unspecified	<AFFECT>	<AFFECT> </AFFECT>
affect, specified	<SAD>	<AFFECT=SAD> </AFFECT>
affect, specified	<ANGRY>	<AFFECT=ANGRY> </AFFECT>
affect, specified	<DISGUST>	<AFFECT=DISGUST> </AFFECT>
affect, specified	<FEAR>	<AFFECT=FEAR> </AFFECT>
affect, specified	<HAPPY>	<AFFECT=HAPPY> </AFFECT>

Frame (q.v.)

talking on telephone (alt.)	<TEL>	<FRAME=TELEPHONE>
media speech (alternate)	<MEDIA>	<FRAME=MEDIA>

Metatranscription (q.v.)

unintelligible speech	<#>	<META=UNINTELLIGIBLE>
code-switch/2nd language	<L2>	<LANGUAGE=SWITCH>
code-switch (lg. specified)	<L2=DUTCH>	<LANGUAGE=DUTCH>

Voice/Manner: durative background events

SEE: VOICE/MANNER

<<WHISTLE>>	4
<<SING>>	2
<<SOB>>	1
`<<COUGHING>>	1
<<LAUGHTER>>	1
<<MURMUR>>	1

<<VOMIT-SOUND>> 1
 <<TALK>> 1

[Notes: Relation to literature]

Bateson on frames, metamessage
 Paralanguage: Hockett, Trager, McQuown, Birdwhistell
 First five minutes
 Goffman on footing
 Chuck Goodwin
 Streeck
 Kendon
 Ekman
 Bakhtin: Multivocality
 etc.

Isotony

(1) (*Raging Bureaucracy* SBC004: 250.12-269.21)

1 SHARON; cause I] made the ^whole class learn,
 2 like,
 3 (H): good ^morni:ng,
 4 ^goodbye:.,
 5 (H) can I go to the ^bathroo:m,
 6 can I stand ^u:p,
 7 (0.3)
 8 (H) can I go to the —
 9 er —
 10 `Time to go to the ^cafeteria:.,
 11 (0.8)
 12 (H) `an:d,
 13 (0.2)
 14 the ^office,
 15 you know,
 16 `certain st— `terms that related to ^school.
 17 (1.0)
 18 (TSK) (H) And ^she:.,
 19 like,

(2) (*Raging Bureaucracy* SBC004: 163.59-167.94)

1 SHARON; (H) and ^they: have never learned `English.
 2 (0.8)
 3 SHARON; (TSK) And ^they have never been `educated.

(3) (*Raging Bureaucracy* SBC004: 849.09-850.47)

1 SHARON; That's ^right.
 2 (0.2)
 3 That's ^right.

(4) (*Constitution*)

1 LONI; I mean that's `not the only one ^too,
 2 there was that ^Paula: `Jones?
 3 (0.6)
 4 and there was ^Watergate?
 5 #and wasn't there ^another gate?
 6 (0.3)
 7 and ^another: something or other?

(5) (*Cuz* SBC006: 55.43-59.23)

1 ALINA; So he got ^another radio this ^summer,
 2 but of course ^that got ripped off ^also.
 3 But ^never mind.

[rev. 30-Oct-2007]

Chapter **: Questions

Questions 1 [version A, simplified]

Notation

End-tones include: closure (continuing vs. final), appeal, and truncation.

Symbol combinations for end-tones include: , . ?, ?. —

Exercise

At the end of each Intonation Unit, transcribe the end-tone.

5. (*Cuz* SBC006:2.50-6.71)
- 1 ALINA; Remember Tyke
3 .. Lived next door to Mom
4 (1.6)
5 LENORE; Yeah:
7 ALINA; .. Okay
6. (*Cuz* SBC006: 18.68-22.44)
- 1 LENORE; How do you know
2 (0.2)
3 ALINA; (H) Well
7. (*Raging Bureaucracy* SBC004: 27.75-29.51)
- 1 SHARON; Llenar
2 is to fill
3 SHANE; .. Yeah
8. (*Raging Bureaucracy* SBC004: 13.25-14.75)
- 1 KATHY; Juice in the morning
2 .. makes you sick
9. (*Doesn't Work in This Household* SBC019: 698.217-701.384)
- 1 JAN; (TSK) (H) Bob came by this afternoo:n
2 FRANK; Bob who
10. (*Raging Bureaucracy* SBC004: 875.64-879.87)
- 1 CAROLYN; But I was like
2 so what is your f:ather doing right now:
3 (H):_<0.7> Is the man d:ead yet

Target

11. (*Cuz* SBC006: 18.68-22.44)

1 ALINA; Remember Tyke?
 3 .. Lived next door to Mom?
 4 (1.6)
 5 LENORE; Yeah: .
 7 ALINA; .. Okay.

12. (*Cuz* SBC006: 18.68-22.44)

1 LENORE; How do you know.
 2 (0.2)
 3 ALINA; (H) Well,

13. (*Raging Bureaucracy* SBC004: 27.75-29.51)

1 SHARON; Llenar?,
 2 is to fill?
 3 SHANE; .. Yeah.

14. (*Raging Bureaucracy* SBC004: 13.25-14.75)

1 KATHY; Juice in the morning?,
 2 .. makes you sick?.

15. (*Doesn't Work in This Household* SBC019: 698.217-701.384)

1 JAN; (TSK) (H) Bo:b came by this afternoo:n?,
 2 FRANK; Bob who.

16. (*Raging Bureaucracy* SBC004: 875.64-879.87)

1 CAROLYN; But I was like,
 2 so what is your f:ather doing right now: .
 3 (H):_<0.7> Is the man d:ead yet?.

[Revised 14-Oct-2008]

Questions 2

39. (*Cuz* SBC006: 76.91-83.17)

1 ALINA; We had to go see ζJan: last Sunday:?,
 3 .. at her ζhouse for dinner?.
 4 (H) And—
 5 or #at ζbreakfast?,
 6 And Mom said,
 7 .. n— you know,
 8 don't bring it up to Linda,
 9 don't tell her about Mike.

40. (*Raging Bureaucracy* SBC004: 59.95-66.16)

1 SHARON; accidentally I'll,
 2 .. I'll put an el or a la,
 3 in front of a verb that ζsoun:ds masculine or feminine?.
 4 Cause I'll forget.

41. (*Raging Bureaucracy* SBC004: 120.45-126.64)

1 SHARON; And I— I like,
 2 withheld ζrecess?,
 3 from: .. s— several kids,
 4 on,
 5 .. on [Thursday],
 6 KATHY; [Well did you gi]ve candy to the ones that got .. <L2> ζexelentes </L2>?.

42. (*Doesn't Work in This Household* SBC019: 698.217-709.429)

1 JAN; (TSK) (H) Bob came by this ζafternoo:n?.
 2 FRANK; Bob who.
 3 (0.4)
 4 JAN; Um.
 5 BRETT; A[sh-tone].
 6 JAN; [Ash]ton,
 7 JAN; and,
 8 (0.4)
 9 uh,
 10 .. he has his:,
 11 .. miter thing.
 12 FRANK; <VOX> ζHe took it </VOX>?.
 13 JAN; Unhunh.
 14 FRANK; (H) Why.
 15 (0.5)
 16 MELISSA; [Cause it's] [₂hi:s].
 17 BRETT; [it's][₂:] —
 18 RON; [₂@]

43. (Cuz SBC006: 1584.95-1599.52)

1 ALINA; It's about —
 2 it has to do with the bar (0.4) generation.
 3 @_<0.5>
 4 U:m,
 5 @[₂@]
 6 LENORE; [₂The-] w- .. ɿb:a:r gen[era-]?—
 7 ALINA; [generati]on.
 8 Yeah:.
 9 (%) (%)
 10 (TSK) Oh[;,
 11 LENORE; [#ɿ What]?.
 12 ALINA; what] is it.
 13 (0.9)_(DRINK)
 14 LENORE; The ɿb:a:r generation?.
 15 In what sense bar generation.
 16 ALINA; (TSK) Oh no,
 17 (%)
 18 (0.2)
 19 <L2=FRENCH> Pub </L2>.
 20 (0.2)
 21 The Pub.
 22 (0.2)
 23 [The Pub Generation].
 24 LENORE; [@@@]
 25 ALINA; There you go.

Chapter **: Simultaneity

Objective

The objective of this section is to present consistent principles to guide the transcriber in representing the various forms of speech overlap, i.e. configurations of simultaneous speech by multiple speakers in a conversation or other spoken interaction.

Conventions

Definitions

Overlap is defined as simultaneous speech (or other vocal activity) by two or more participants in a given interaction.

An *overlap group* is a set of participants who are speaking simultaneously at any given moment.

Notations

overlap (1st set) [] align left square brackets vertically
overlap (2nd set) [₂] left square brackets indexed with subscript

Simple Overlap

Each case of overlapping speech in a conversation is marked with square brackets. For each speaker participating in a given overlap group, one pair of square brackets is used. A left bracket indicates the start of overlapping speech by the current speaker; a right bracket indicates the end of overlapping speech. An additional pair of brackets (one left and one right bracket) is used for each additional participant in the overlap.

In general, the two left brackets are aligned vertically, in order to iconically convey that they represent the same moment in discourse time. For typographical and practical reasons, no attempt is made to vertically align the right brackets. (Indeed, it is generally preferable *not* to attempt to vertically align right brackets, as this can make the transcriptions difficult to read and interpret, for a variety of reasons.)

Enclosed overlap

Enclosed overlap occurs when the words of the second speaker are entirely enveloped within the stretch of speech by the previous speaker. From another point of view, the first speaker continues speaking after the second speaker finishes.

1. (*Deadly Diseases* SBC015: **-405.205)
- 1 JOANNE; .. That wasn't Deutsch?
- 2 That [sou:nds like] Deutsch,
- 3 KEN; [No].

2. (*Appease the Monster* SBC042 201.759-215.895)

- 1 KENDRA; ... All I need is your signature so I can .. play the volleyball.
 2 On [the volleyball] team.
 3 KITTY; [No].
 4 .. No.

3. (*Appease the Monster* SBC042)

- 1 MARCI; % Listen.
 2 I figure I was doing well to have hit the steaks,
 3 which I also [didn't know w]as your favorite thing.
 4 KENDRA; [Yeah].

4. (SBC051 1434.637-1436.440)

- 1 FRAN; How [long did you] stay.
 2 BERNARD; [Loved it].
 3 Four years.

Stepped Overlap

Stepped overlap occurs when the second speaker continues speaking after the first speaker has stopped.

5. (*Conceptual Pesticides* SBC)

- 1 MARILYN; .. we don't have a com[post pile yet].
 2 PETE; [So it just goes] in the trash like this.

6. (*Lambada* SBC002)

- 1 MILES; Since he didn't know her,
 2 .. <HI> that must have been [quite a shock </HI>].
 3 JAMIE; [That was really] wei:rd.

7. (*Doesn't Work in This Household* SBC019: *-202.368)

- 1 MELISSA; And so Mom told me never ever to ask,
 2 FRANK; Maybe [I'll have tea].
 3 MELISSA; [anything about that a]gain.

Simultaneous start

Sometimes two speakers begin speaking at the same time, producing a simultaneous start. This is adequately captured by the standard square bracket notation used for overlapping speech; no special notation is required.⁸

8. (*Lambada* SBC002)

- 1 MILES; Does she have a what?
 2 JAMIE; [A ma:n].
 3 HAROLD; [A ma:n].

9. (*Deadly Diseases* SBC015: 398.735-**))

- 1 JOANNE; He is like that though.
 2 that's his attitude.
 3 KEN; [That wasn't Deutsch].
 4 JOANNE; [That's what I don't like] about him.
 5 Oh.

10. (*Doesn't Work in This Household* SBC019)

- 1 JAN; Is everyone who plays in that symphony paid.
 2 FRANK; .. [They don't get] paid much.
 3 BRETT; [Yes].
 4 ...
 5 FRANK; But they get paid a little.

Simultaneous end

Sometimes two speakers end their overlapping utterances at the same time. Again, no special notation is required for this, as it is adequately represented by the standard overlap notation.

11. (*Conceptual Pesticides* SBC)

- 1 ROY; It's [illegal].
 2 MARILYN; [it's illegal].

12. (*Conceptual Pesticides* SBC)

- 1 MARILYN; Isn't that [retarded]?
 2 ROY; [Isn't that weird]?

3-way overlap

While the examples so far have all presented cases where exactly two speakers speak at once, an overlap group may also include three or more speakers. The standard square-bracket notation works for such cases, regardless of the number of speakers in the overlap group.

13. (*Doesn't Work in This Household* SBC019)

- 1 BRETT; [Who does].
 2 FRANK; [Mhm].
 3 MELISSA; [and Young Life] is over there,

Vocalisms: Laughter

Overlap is not just about speech, but further encompasses anything that speakers do with their vocal tracts. Thus overlap occurs between speech and vocalisms such as laughter, or between two instances of vocalisms produced by different speakers.

14. (*Guilt* SBC033 31.401-36.316)

- 1 JENN; The engine of this family,
 2 is a little guilt motor.
 3 That goes vroo:m [vroo:m vroo:m].
 4 DON; [@ @ @ @]

Because of the social nature of laughter, one of the commonest types of overlap involves multiple participants laughing simultaneously, often in quite complex configurations of overlapping laughter.

15. (*Conceptual Pesticides* SBC003)

1 MARILYN; You @son @of @a @bitch.
 2 PETE; @@@[@@@]
 3 MARILYN; [@ @ @] @ @

16. (*Lambada* SBC002: 1049.16-1055.75)

1 MILES; (H) But @,
 2 JAMIE; @: [@@]
 3 PETE; [@@@@@]@ .. @ (H)

Vocalisms: Breath

One of the subtler forms of overlap involves one participant breathing while another is speaking. This can carry important interactional information, such as indicating that the breathing speaker is about to begin speaking, or is planning to continue to hold the floor. One complication for transcribers is that breathing can be difficult to hear reliably in recordings which are of poor quality or which contain a lot of background noise.

17. (*Tapedeck* SBC 261.043-263.817)

1 BRAD; that's a circuit to cut the tap:e hiss,
 2 .. [(H)] without cutting the high frequencies.
 3 TAMMY; [right].

Multiple Overlap Sets

Overlap is sufficiently frequent in conversational interaction that often one overlap event has scarcely concluded before another one begins. This is especially likely in lively conversations involving three or more speakers.

In cases where an instance of overlap between a certain group of speakers is immediately followed in the subsequent conversation by another instance involving a different set of speakers, the two overlap groups are distinguished by attaching different indexes, using subscript numerals, to the overlap brackets in the respective overlap groups. This is done to avoid confusion when multiple overlaps occur in close succession: overlap brackets that belong together are indexed using the same number, affixed to the left square brackets.

In addition, left brackets are aligned vertically for iconic clarity. For typographical and practical reasons, no attempt is made to align the right square brackets.

18. (SBC051 1431.174-1438.447)

1 BERNARD; <@> And there I was,
 2 in New York city </@>.
 3 @
 4 ...
 5 FRAN; Did [you like it]?
 6 SEAN; [#Hair #school]?

- 7 ...
 8 FRAN; How [₂long did you] stay.
 9 BERNARD; [₂Loved it].
 10 Four years.
 11 ...
 12 FRAN; Where'd you live.

Mixing 2-way and 3-way overlap

There is in principle no limit to the number of speakers in any overlap group (beyond the limitation imposed by the number of people present, of course). Instances of 3-way overlap may be interspersed with instances of 2-way overlap, for example. This does not affect the standard principle that all members of the a given overlap group are assigned a common bracket index, represented by a subscript numeral. This holds true with the speakers involved in two successive overlap groups are different speakers, or the same set of speakers overlapping yet again.

19. (*Doesn't Work in This Household* SBC019: 780.516-786.546)
 1 MELISSA; they live three blocks [away],
 2 BRETT; [(H)]
 3 [₂Who does].
 4 FRANK; [₂Mhm].
 5 MELISSA; [₂and Young Life] is over there,
 6 FRANK; Good for [₃them],
 7 MELISSA; [₃and so],
 8 it would make .. most sense for me to have gone over there.

For a presentation transcript (e.g. a version of a transcription excerpt as presented when publishing for a general audience in an academic journal), it is sometimes possible to rely on the iconic value of the vertical alignment of overlap brackets. This may be desirable when presenting examples to a relatively unsophisticated audience that may not be familiar with the complexities of spoken discourse, thus motivating the use of a visually simplified transcription for the ease of exposition. It may be possible in such cases to dispense with the numbered bracketing, though this option should be used with caution. To ensure the reliability of one's own primary data, however, it remains best practice in a transcription database to index all instances of confusable overlap brackets by numbering them, in order to avoid unrecoverable disruption of overlap information during editing.

Potential ambiguity (without bracket indexing)

As evidence of the importance of indexing bracketing, it should be noted that real misunderstandings can arise regarding the question of who is overlapping with whom, if the brackets are not carefully indexed to remove ambiguity.

20. (*Lambada* SBC002: 66.75-**)
 1 MILES; I'm just ama:zed.
 2 JAMIE;[## —
 3 Dumb mo:ve].
 4 HAROLD; [See I'd heard that it was main]ly,
 5 JAMIE;[₂Really dumb].
 6 HAROLD; [₂%in .. %t's mainly] [₃urban] areas,

7 PETE; [₃Yeah].

In the above example, if no bracket index numbers were used, the reader may be led to think of Jamie in line 5 overlapping with Harold in line 4. But since Harold's lin4 overlaps with Jamie's 3, this leads to an impossible conclusion--that Jamie is overlapping with herself. Although readers may eventually figure all of this out for themselves, they shouldn't have to do this kind of detective work just to decipher a transcript.

More seriously, there are cases where logical deductions would not be sufficient to recover the correct facts. In a case where four different participants speak, it makes a difference whether all four overlap together, or the first two produce one overlap and then the last two make a second overlap group.

21. (*Howards End* SBC023)

1 DIANE; [and],
2 SUE; [Yeah].
3 LINDA; [₂True].
4 LORI; [₂#]

In contrast to the above case of two separate overlap groups, note that it really can happen on occasion that a single overlap group involves four people starting to speak at the same time.

22. (*Howards End* SBC023)

1 DIANE; [It's so] .. [₂proper],
2 LINDA; [Ye:s].
3 EVELYN; [Mm].
4 JANICE; [Yeah].
5 LINDA; [₂True].

The possibility for true ambiguity is real, if rare. More commonly, failure to index brackets simply leads to confusion, or unnecessary work for the reader.

Complex Overlap

There is an almost unlimited array of possibilities for more complex forms of overlap. A few of the more common situations that a transcriber is likely to encounter are illustrated here.

Multiple short IU's overlap one longer IU

It often happens that while one speaker produces a single intonation unit, two (or more) other speakers produce successive intonation units which overlap with those of the first speaker. Here the standard principles for writing intonation units and for writing overlaps apply. Each intonation unit is written on a line by itself, and each overlap group receives its own distinctive index.

23. (*Raging Bureaucracy* SBC004: 908.08-910.38)

1 KATHY: .. And [that was the on][₂ly way to do it].
2 CAROLYN: [No doubt].
3 SHARON: [₂(H) crazy].
4 CAROLYN: No doubt.

In the example above, the first overlap group is Kathy and Carolyn (lines 1 and 2), while the second

overlap group is Kathy and Sharon (lines 1 and 3). Each overlap group is indexed differently (using plain brackets vs. brackets with subscript numeral 2). In the next line, Pete's intonation unit in line 3 is overlapped by three separate (short) intonation units, as produced by two separate speakers. Each overlap gets its own brackets.

24. (*Lambda* SBC002)

- 1 PETE; As compared to the other stuff,
 2 for some reason it sounded like it was out there,
 3 or kind of [.. coming] in [₂from some] [₃place],
 4 JAMIE; [Unhunh:]?
 5 HAROLD; [₂Hm:].
 6 JAMIE; [₃Hunh].

It may also happen that while the first speaker produces a single intonation unit, a single second overlapping speaker produces two (or more) intonation units. It is important to note that in such cases, the standard practice for representing intonation units is maintained: each if the second speaker's two intonation units is written on a line by itself. In such cases it advisable to use two separate sets of overlap brackets, with distinct indexing, to correspond to the two separate intonation units produced by the second speaker.

25. (*Guilt* SBC033 96.012-97.781)

- 1 JENN; [Like you're] feeling a respon[₂sibility] towards people,
 2 DON; [(TSK) Ah].
 3 [₂Mm].

Change in overlap group membership

In any lively multi-party conversation, it is likely that there will be times when the configuration of overlapping speakers changes in the middle of an intonation unit, in ways that require a general strategy for representing what is happening. Compare the following two versions of the same stretch of conversation. The first excerpt shows the utterances of two speakers who overlap, beginning in the latter portion of the first speaker's intonation unit:

26. (*Lambda* SBC002)

- 1 HAROLD; You think we have like a [jazz band next door]?
 2 MILES; [In fact I was getting ready to say],

Based on the excerpt presented so far, the speakers can be described as participating in an overlap group composed of two members, Harold and Miles. This representation looks perfectly routine, and is familiar from many other similar examples. But if we expand the context of the transcription excerpt to include what happens next, a more complex picture emerges. In the continuation, a third participant comes in half-way through Harold's and Miles' overlap portions, producing some laughter (line 3):

27. (*Lambda* SBC002)

- 1 HAROLD; You think we have like a [jazz] [₂band next] [₃door]?
 2 MILES; [In fact I] [₂was getting ready] [₃to say],
 3 PETE; [₂@@@@]
 4 (0.3)

5 MILES; is there somebody downstairs,
6 playing?

This kind of overlap is quite common in lively conversation with three or more participants. The best way to approach it is to think in terms of moments of discourse time, and to consider what is changing across a sequence of successive moments. At any given moment in the discourse, there is a certain group of (one or more) speakers who are currently speaking, whose membership may change quickly and often during the conversation. Sometimes there is just one person speaking at a time (no overlap), and sometimes there are two participants speaking at once (simple overlap). Even three or more participants speaking at once can be considered simple overlap if they all start and end together. But where it gets complicated is when there is a change in the membership in the group of simultaneous speakers (the "overlap group"): for example, a change from two people speaking at a given moment to three people speaking, or from three speaking to two. This is what is happening in the above interaction. The number of speakers participating in the first overlap group is two, but with Pete's entrance in line 3 the overlap group changes to one with a membership of three. After his brief laugh, the current overlap group shifts again, decreasing from three participants to two.

This requires a more complex representation to capture the overlap among shifting groups, as the above example shows. The general rule for representing such overlaps is the *overlap group principle*:

28. For any given moment in a spoken interaction, whenever there is a change (i.e. an increase or decrease) in the number of participants in an overlap group (i.e. the set of simultaneous speakers), the transcriber should do the following:
 - (a) close off the set of brackets for the previous overlap group (i.e. insert a right bracket to indicate the end of this overlap segment), and
 - (b) open a new set of brackets for the current overlap group (inserting a new left bracket to signal the start of the new overlap segment).

Though the resulting proliferation of brackets and their index numbers might seem at first to make the transcriptions unduly forbidding, it will be found that this rule provides a quite general solution to the bewildering diversity of situations that the transcriber attempting to represent complex overlap must face. In most kinds of cases that are likely to be encountered, this solution yields a clear basis for an effective solution to the transcription problem at hand, defining a representation which is unambiguous in its interpretation.

Continuous Stepping Overlap

The overlap group principle defined above is quite general in its application. It extends without modification to cases of chaining or continuous stepping overlap like the following. (In the following example, a waitress has just delivered food to a table of customers, who haven't started eating yet.)

29. (SBC031 1304.992-1308.874)

1	JAMIE;	Is everything alright [now].
2	ROSEMARY;	[Look][₂ s fi][₃ :ne].
3	BETH;	[₂ Look][₃ s de][₄ licious],
4	SHERRY;	[₃ Looks] [₄ good],
5	JAMIE;	Okay great.
6	BETH;	Thank you.

Overlap with lengthened sounds

30. (Guilt SBC033 527.808-535.499)

- 1 RICHARD; What does this have to do with guilt?
 2 LISBETH; @ (Hx) [@ @ @]
 3 LAURA; [(LAUGHTER)] [2:] [3:] [4:] [5:] [6:] [7:] [8:]
 4 BILL; [.. @] [2@Right.
 5 DON; [2@@@@@]
 6 BILL; .. @2]
 7 [3Right].
 8 DON; [4I don't] [5know ~Rich].
 9 LEANNE; [5Or mon] [6ogamy].
 10 BILL; [7What does it have to do with monogamy.
 11 That's what I wan] [8na know].

Style: Overlap Rules

The following is a summary of the various rules and principles governing the use of notations to represent overlapping speech. While many of these rules could be considered a matter of transcription style, it is important to note that they are designed to promote clarity and iconicity in the transcription while avoiding representational ambiguity.

Do	Don't
Use square brackets exclusively for marking overlap.	Never use square brackets for <i>anything</i> other than overlapping speech (such as comments, phonetic transcription, etc.--for each of these, an alternative notation is provided).
Use a left square bracket to indicate the start of an overlap, and a right square bracket to indicate the end.	Don't leave the end of an overlap unmarked.
For every left bracket, there should be a corresponding right bracket for the same speaker.	Don't overlook cases of unmatched or stray brackets.
Align left brackets vertically, to iconically represent the same moment of discourse time.	Don't align right brackets. Never manipulate the transcription (by inserting spaces, etc.) to try to get the right brackets to overlap. Let the right brackets fall where they may.
Insert enough spaces (never tabs) before the second speaker's left bracket to make it align vertically with the first speaker's left bracket.	Never insert extraneous spaces in the middle of a word (e.g. in an attempt to get right brackets to align vertically). Space has a meaning which should be respected--it marks word boundaries--

	and the transcriber should not introduce spurious word boundaries by breaking up words.
If it is inconvenient to align left brackets vertically, this can be dispensed with as redundant, given that the numerical indexing of brackets alone is sufficient to unambiguously specify which bracket corresponds to which. In this case, the second overlap should begin at the left text margin.	Don't align left brackets if it makes the transcription difficult to read, i.e. don't do it if it forces the current intonation unit to wrap over to the next line.
Always close off the brackets for one overlap group before opening a new set of overlap brackets for the next overlap group. Do this whenever there is a change (i.e. an increase or decrease) in the number of participants in an overlap group (i.e. the set of simultaneous speakers).	Never embed one pair of overlap brackets inside another: <i>This [should [2not] be] done.</i> Never use two left overlap brackets together: <i>[[</i> Never use two right overlap brackets together: <i>]]</i>
To avoid cases of possible confusion, index the left brackets in an overlap set with a distinctive subscript numeral, to indicate that all the left brackets with the same numeral represent the same instant of discourse time.	In general, do not index the right overlap bracket with a subscript numeral in cases where the number is redundant (i.e. when the [unindexed] right bracket occurs on the same line of the transcription as the [indexed] left bracket). However, the right bracket may be explicitly indexed when necessary, as when it appears on a different line from the corresponding left bracket.
Use plain brackets whenever possible, i.e. whenever the absence of indexing will not lead to confusion about who is overlapping with whom.	Don't number all brackets in a transcript just to have them numbered. This produces unnecessary clutter.
Index overlap brackets (when necessary) using the numerals 2-9 only. Remember that the first set of overlap brackets is always marked using plain brackets; the second set is indexed with the numeral 2, and so on.	Never index overlap brackets using the numerals 1 or 0. The numeral 1 looks like a (lower-case) letter L in many fonts, and the numeral 0 can be confused with the letter O.
If a large number of overlaps occur in a long sequence without interruption, it is sometimes necessary to start the indexing sequence over again from 2. The new sequence can be continued up to 9 again, cycling from 2 to 9 as many times as necessary.	Never use double digits for bracket indexing. On the other hand, also avoid using plain brackets in the middle of a complicated sequence of 10 or more distinctively indexed sets of brackets; the plain brackets don't stand out sufficiently among all those numbers.
As soon as it is convenient and unambiguous, the use of bracket indexing should be dropped in favor of using plain brackets. Plain brackets are to be preferred any time when there is no immediately prior case of overlap, that is, when at least one	Don't number overlap brackets that occur following one or more overlap-free intonation units.

overlap-free intonation unit has occurred since the last case of overlap.	
Note that the use of <i>subscript</i> formatting for the numerals that index overlap brackets is redundant. If the subscript format is not used (or is lost), no information is lost, since the use of a numeral adjacent to a square bracket unambiguously indicates overlap indexing.	Don't worry about preserving subscript formatting when, e.g., inserting transcription data into a database, because it is in principle redundant. The numbers are what matter, not the subscript font.

[rev. 16-Oct-2007]

Overlap Interpretation

Objective

The objective of this session is to think about the relationship between speakers and their utterances during overlapping or simultaneous speech.

In addition, the session emphasizes the practice of *reading* a transcript to gain information about the nature and organization of the interactional events it portrays.

Questions: Overlap

Regarding the simultaneous speech, consider the following questions:

1. Who said what when?
2. Who is overlapping with whom?
3. Why are they overlapping? Is it interruption, solidarity, duetting, choral production, or ...?

As always it is important to keep in mind, for any given moment in the conversation what the participants know about the discourse they are in (what they have access to).

Transcript vs. Recording

Begin by simply *reading* the transcript, without hearing the recording. Make your interpretation with respect to the above questions.

1. Based on reading the transcription, what seems to be going on?
2. How much can you tell from the discourse transcription?

Now listen to the recording. After you have heard it, consider what, if anything, changes in your interpretation of the events.

3. How well did your transcript-based interpretation hold up?
4. What kinds of new information did the recording give you, which might lead you to change your interpretation?
5. What additional features could have been added to the transcription to give you the information you needed for a better interpretation?

Examples

Q. Can a person who is speaking hear the words of someone else who is speaking at the same time as them (overlapping with them)? How can you tell? What is Joanne doing in line 5? What is Ken doing in line 8?

31. (*Deadly Diseases* SBC015: 398.735-405.205)
- | | | |
|---|---------|---------------------------------------|
| 1 | JOANNE; | He is like that though. |
| 2 | | that's his attitude. |
| 3 | KEN; | [That wasn't Deutsch]. |
| 4 | JOANNE; | [That's what I don't like] about him. |
| 5 | | Oh. |
| 6 | | .. That wasn't Deutsch? |
| 7 | | That [sou:nds like] Deutsch, |
| 8 | KEN; | [No]. |

Q. What is the sequential relation between Frank's utterance in line 5 and the surrounding discourse? What is the consequential relation?

32. (*Doesn't Work in This Household* SBC019: 196.497-202.368)
- | | | |
|---|----------|---------------------------------------|
| 1 | MELISSA; | And they just talked about it, |
| 2 | | on and on and on and on and on. |
| 3 | | (0.3) |
| 4 | MELISSA; | And so Mom told me never ever to ask, |
| 5 | FRANK; | Maybe [I'll have tea]. |
| 6 | MELISSA; | [anything about that a]gain. |

33. *Without* (1) voice segregation; *with* (2) temporal sequencing of words (only):
I threw a green you threw a pepper green down your pepper blouse down my shirt

34. (*Conceptual Pesticides* SBC003: 376.04-384.85)
- | | | |
|---|----------|--|
| 1 | ROY: | I threw a [green pepper down your blouse]. |
| 2 | MARILYN: | [You threw a green pepper down] my shirt. |

[NOTE: This is a *broad* transcription of overlaps – and a little loose. Note the “cheating” in line 9, in which Joanne's speech supposedly overlaps with Ken's pause.]

35. (*Deadly Diseases* SBC015: 198.728-208.130) ((BROAD))
- | | | |
|----|---------|---|
| 1 | JOANNE; | (H) I think Mexi]co's like the place: to go:, |
| 2 | KEN; | I don't know]. |
| 3 | JOANNE; | because it's got everything. |
| 4 | | (H) It's got history, |
| 5 | | it's [got big cities, |
| 6 | KEN; | [yeah, |
| 7 | | it's got— |
| 8 | JOANNE; | it's got ruins, |
| 9 | KEN; | (0.3) It's got diseases, |
| 10 | JOANNE; | and it's got— |
| 11 | KEN; | that I keep @catching]. |
| 12 | JOANNE; | (H) and it's] got res[₂ o:rts]. |

[NOTE: This is a *narrow* transcription of overlaps. Note that Ken's "pause" no longer exists!]

36. (*Deadly Diseases* SBC015: 198.728-208.130) ((NARROW))

- 1 JOANNE; [(H) I] [₂think ^Mexi]co's like the `place: to go:,
 2 KEN; [₂I don't know].
 3 JOANNE; because it's `got ^everything.
 4 (H) `It's got ^history,
 5 `it's [got `big] [₂^cit]ies,
 6 KEN; [Yeah],
 7 [₂it's got]—
 8 JOANNE; `it's [₃got ^ruins],
 9 KEN; [₃It's got ^diseas][₄es],
 10 JOANNE; [₄and] [₅it's ^got]—
 11 KEN; [₅that I keep] [₆@catching].
 12 JOANNE; [₆(H) and `it's] got ^res[₇o:rts].

37. (*Lambada* SBC002: 241.94-255.37)

- 1 MILES; [I thought that was the] real thi:ng.
 2 (0.2)
 3 HAROLD; You think we have like a [jazz] [₂band next] [₃door]?
 4 MILES; [In fact I] [₂was getting ready] [₃to say],
 5 PETE; [₂@@@]
 6 (0.3)
 7 MILES; is there somebody downstai:rs,
 8 playing?
 9 JAMIE; [@]
 10 PETE; [#I #mean] that was what I first [₂thought—,
 11 JAMIE; [₂(H) It's #our garage band₂].
 12 PETE; when .. when the ja:zz came on₂].
 13 (0.2)
 14 PETE; As compared to the other stuff,
 15 for some reason it sounded like it was out there,
 16 or kind of [.. coming] in [₂from some] [₃place],
 17 JAMIE; [Unhunh:]?
 18 HAROLD; [₂Hm:].
 19 JAMIE; [₃Hunh].

[rev. 7-Nov-2005]

Discourse Time: Overlap Ambiguity

Ambiguity (vs. bracket indexing)

(1) Transcription: overlap AMBIGUITY (of who's overlapping, in the absence of bracket indexing: note overlaps 4 AND 5 – could be a 4-way overlap! – instead, it's two 2-way overlaps)

(SBC023 HOWARDS END)

DIANE: [4and4],
SUE: [4Yeah4].
LINDA: [5True5].
LORI: [5X5]

(2) ACTUAL 4-way overlap – it does happen!

(SBC023 HOWARDS END)

DIANE: [It's so] .. [2proper2],
LINDA: [Ye=s].
EVELYN: [Mm].
JANICE: [Yeah].
LINDA: [2True2].

(Lambada SBC002: 660.75-**))

1 MILES; Cause there're a lot of women out there who apparently don't believe in using condoms.
2 (1.0)
3 PETE; Hm.
4 (0.6)
5 MILES; I'm just ama:zed.
6 JAMIE; [## —
7 Dumb mo:ve].
8 HAROLD; [See I'd heard that it was main]ly,
9 JAMIE; [2Really dumb].
10 HAROLD: [2%in .. %t's mainly] [3urban] areas,
11 PETE: [3Yeah].

(4) Transcription: overlap AMBIGUITY (of who's overlapping, in the absence of bracket indexing: note overlaps 2 and 3)

BRETT: [@]
MELISSA: [XX] something @like that.
BRETT: .. [2%=2]
MELISSA: [2When he awaked2]?
JAN: .. [3When he3] [4awoke4].
BRETT: [3Yeah3].
MELISSA: [4<READ I bet it was fun4] for him when he awoke,
at the silence,
BRETT: he awokeded [at the silence].
MELISSA: [and he got to hear] Cousin Simmons and the Reverend,
saying fine things about him READ>.

Chapter **: Discourse Time

Sequence & Consequence

Objective

The objective of this session is to think about the relationship between *sequence* and *consequence* in interaction.

In addition, the session emphasizes the practice of *reading* a transcript to gain information about the nature and organization of the interactional events it portrays.

Questions: Discourse Time

To find the relation between sequence and consequence, consider the questions:

- (5) What is the order of events?
- (6) How does this order affect what happens?

This requires close attention to the following, for any given moment in the conversation:

- the sequence of utterances and other events, pauses, overlapping speech, and so on
- what the participants know about the discourse they are in (what they have access to)
- what activity is currently in progress
- what consequences a participant's action has in the subsequent interaction

Going beyond the analysis of individual examples, can you find some generalizations about these data, and/or your own conversational data? Are there recurrent sequential patterns, which have consistent consequences for the discourse and for the participants?

Transcript vs. Recording

Begin by simply *reading* the transcript, without hearing the recording. Make your interpretation with respect to the above questions.

- Based on reading the transcription, what seems to be going on?
- How much can you tell from the discourse transcription?

Now listen to the recording. After you have heard it, consider what, if anything, changes in your interpretation of the events.

- How well did your transcript-based interpretation hold up?
- What kinds of new information did the recording give you, which might lead you to change your interpretation?
- What additional features could have been added to the transcription to give you the information you needed for a better interpretation?

Examples

Q. *What is happening in line 2? Why?*

Does Arnold's utterance in line 5 respond to Rickie's utterance in line 4?

38. (Tell the Jury That SBC008: 317.532-320.552)

- | | | |
|---|----------|-------------------------|
| 1 | REBECCA; | You guys newly married? |
| 2 | | (0.5) |
| 3 | RICKIE; | Well, |
| 4 | | [just a yea]r. |
| 5 | ARNOLD; | [A yea:r], |
| 6 | | two days ago. |

Q. *What is Harold doing in line 6? Is he echoing Jamie? What is she doing in line 5?*

39. (Lambada SBC002: 301.70-308.16)

1 HAROLD; [Does she even] have a b— a man—?
 2 I guess she must.
 3 (0.6)
 4 MILES; Does she have a what?
 5 JAMIE; .. [A ma:n].
 6 HAROLD; [A ma:n].
 7 PETE; @@@@

Q. *What do you think is the relationship between line 2 and line 1? How does it reflect the alignment between the speakers?*

40. (Hey Cutie Pie SBC028: 765.38-766.80)

1 JEFF; No you don't.
 2 JILL; Yes we do.

Q. *What do you think is the relationship between line 6 and line 4? How does it reflect the alignment between the speakers?*

41. (Lambada SBC002: 1026.70-1027.59)

4 JAMIE; Yeah.
 5 (0.2)
 6 MILES; No.

Q. *The following example gives three additional lines of (prior) context for the previous example. How does seeing this sequence of utterances change your understanding of the relationship between lines 6 and 4. What is the general pattern here?*

42. (Lambada SBC002: 1024.04-1027.59)

1 JAMIE; Do you need a partner?
 2 (1.2)
 3 MILES; To go there?
 4 JAMIE; Yeah.
 5 (0.2)
 6 MILES; No.

Q. *Why does Kevin repeat his question? Why does Kendra say “yes” twice?*

43. (Appease the Monster SBC013: 530.91-537.11)

1 KEVIN; [2Can you] have a [cat in your] apartment?
 2 MARCI; [(H)]
 4 KENDRA; I'm gonna have a cat in my apartment.
 5 KEVIN; .. Can you,
 6 .. by law,
 7 [have a c]at in your apartment.
 8 KENDRA; [Yeah].
 9 ..
 10 Yea:h.

Q. *What is Pamela laughing about in line 5. Why this now?*

44. (*A Book about Death* SBC005: 1160.43-1169.90)

1 PAMELA; (TSK) (H) Well(Hx),
 2 DARRYL; (TSK) Now if he's a very old soul,
 3 he should keep it to [himself].
 4 PAMELA; [(H)] you know,
 5 (%) (%) @@(Hx)
 6 (0.4)
 7 I felt l— after Gretchen died,
 8 it was all,
 9 (0.2)
 10 rather unfair.

Q. *How does Ken change his utterance between lines 3 and 4? Why? Why just now?*

45. (*Deadly Diseases* SBC015: 618.353-622.150)

1 LENORE; (H) There's `fi]fteen ^thousand units of vitamin ^A,
 2 and I think that's .. [^toxic].
 3 KEN; [^That's] #re— —
 4 That ^is toxic.

Q. *Can a person who is speaking hear the words of someone else who is speaking at the same time as them (overlapping with them)? How can you tell?*

46. (*Deadly Diseases* SBC015: 1074.820-1080.195)

1 LENORE; a few wor:ms @],
 2 JOANNE; when she get]s that bi—,
 3 .. Oh a few worms,
 4 when she gets that big,
 5 she'll need a v:at of worms.
 6 And they're ex[pensive].

[rev. 7-Nov-2005]

Chapter **: Names

Anonymity, Pseudonyms, & Naming Practices

This chapter talks about how to deal with issues of anonymity, especially in terms of changing the names (and other identifying information) of participants in a conversation.

Pseudonym practices

The transcriber should decide where pseudonyms (or other kinds of pseudographs) need to be used to protect anonymity, and where they do not. The transcriber should make this decision at the early stages of the transcription process, implementing it from the earliest drafts to avoid having real names slip through when they shouldn't. (The question of when to use pseudonyms in transcription, and when to bleep the audio, is based on ethical and legal considerations governing confidentiality in research, as well as on the specific nature of the consent agreement worked out between researchers, speakers, and the relevant overseeing agencies; these issues are discussed at length elsewhere.)

Wherever pseudonyms are called for, it is important to think about what name is chosen. This is because names carry a great deal of information about the people who bear them, including obviously their sex (for most names), but also, at least in some cases, their ethnic group membership, approximate age, and degree of formality. As for the sound of the name, each name has a particular prosodic pattern, defined by the number of syllables and the location of the stress.

Prosody. To the extent possible, the pseudonym should be matched to the original name in terms of its prosody, including:

- 44. Syllabic structure
- 45. Stress pattern

Normally the pseudonym should have the same number of syllables and the same stress pattern.

Social meaning. To the extent possible, the pseudonym should be matched to the original name for the following aspects of social meaning:

- 46. Gender (male, female, neutral, etc.)
- 47. Language
- 48. Age (grading according to birth year; note that degrees of age-grading [fads] vary considerably)
- 49. Nicknames (full name vs. nickname)
- 50. Formality (inherently formal name vs. inherently casual name)
- 51. Popularity (rare name vs. common name)
- 52. Ethnicity
- 53. Religion
- 54. Dialect/Region

Variants. One complicating factor is that sometimes one is confronted with the need create pseudonyms, not just for one original name, but for a whole family of names. This happens when there are multiple variations on one person's name, for example if a participant in a conversation is referred to by both a full name and a nickname.

Resonance. If participants in a conversation engage in forms of resonance based on a person's name (for example, making a pun on the name), this creates a real challenge for the transcriber. It may be possible to come up with an alternative name which has resonance comparable to the original, to maintain the pun or other resonance without revealing the real name. But most of the time it is unlikely that the pun can be re-created with a new name, so that some information about the interaction will be lost due to the pseudonym. In such cases it is best for the transcriber to insert a note explaining, in general terms, that a pun was invoked, again without giving away the real name.

Stereotypes. Transcribers should avoid being lazy in coming up with a pseudonyms. If transcribers just use the first name that comes to mind, this can lead to overuse of the most obvious, commonplace, cliché, or stereotypical names. There are various resources that can help in creating appropriate pseudonyms. In addition to books designed to help new parents name their baby, there are numerous web sites that deal with names, which may be helpful in coming up with an appropriate name. (Because many of these web sites are oriented to parents seeking help in the choice of a "baby name", this phrase makes a good basis for an Internet search. One useful web site, based on names in the United States, is <http://www.ssa.gov/OACT/babynames>. Another is <http://www.babycenter.com/babynome/>, which allows you to search by the number of syllables in the name. For other languages and countries, it may be possible in some cases to find similar resources.)

Difference. Needless to say, one should not choose a pseudonym which so closely approximates the original name in its sound and social meaning that it gives away the real name. It is better to sacrifice some social meaning and/or phonetic similarity to maintain anonymity. Avoid giving a "coded" name that is designed to be solved by a clever person based on clues or inference.

Why bleep?

If the audio data are published along with a discourse transcription (e.g. via optical media or on the Internet), it is often important to preserve the anonymity of individuals who participate in the recording, if their name happens to be spoken aloud by a participant. Changing the name in the transcription is a start, but it will also be necessary to modify the sound so that the name cannot be heard. We can call this modification of the sound "bleeping".

How to bleep

There are various ways to bleep the audio for a recorded discourse. One of them is to insert a sound (a "bleep") in place of the original sound, but this can often be quite distracting to users who are trying to listen to what is happening in the event. Another is to insert a silence, but this can be confusing as well. A better way is to filter the sound in a way that preserves some information about the original utterance—how loud it was, what the pitch was—while obscuring the segmental information that would allow a listener to understand what name was spoken.

It is beyond the scope of this section to discuss the technical means for bleeping. The role of this and the following sections is to discuss when to bleep, and why.

Transcribe first, bleep later

It is important that the transcriber should be the one in control of the process of bleeping names. The transcriber controls the process by identifying the specific words in the recording that are to be bleeped. The actual bleeping itself should be done only after the transcription has been completed, or at least brought to a high level of completion, including appropriate checking by other transcribers. This sequence of

transcribing first, and bleeping later, is necessary for several reasons, detailed below. Note that the reasons for transcribing before bleeping hold regardless of whether the bleeping is to be done by the transcriber herself, or by a separate individual such as a bleeping technician (who by preference would normally be someone who is expert in working with speech sound on computers).

Overlap transcription. The following example shows the danger of bleeping names before you transcribe. Because the bleep will obscure all sound produced during the utterance of the name, you won't be able to tell what other words were spoken during the overlap!

- ((*A Book About Death* SBC005 442.12-448.25))
- | | | |
|---|---------|--|
| 1 | PAMELA; | (H) I'm thinking one thing my mother always used to say; |
| 2 | | when I wouldn't go bicycling with my [father], |
| 3 | DARRYL; | [@~Pamela], |
| 4 | | @you @are [2@@@], |

Here the transcriber would have no way to hear the word *father*, and would simply have to leave this information unrepresented. In the following example from the same conversation, the presence of laughter and the word *she* would go undetected:

- ((*A Book About Death* SBC005 767.67-775.81))
- | | | |
|----|---------|---------------------------|
| 1 | PAMELA; | I said, |
| 2 | | ... |
| 3 | | love fills the stockings. |
| 4 | DARRYL; | Oh Go:d, |
| 5 | | [~Pamela]. |
| 6 | PAMELA; | [@@@ She] said, |
| 7 | | (H) she said, |
| 8 | | oh you mean, |
| 9 | | ... |
| 10 | | adults:? |

Familiarity. A “bleeper” technician can easily let something slip through that would compromise anonymity: a quietly uttered name, a name uttered in overlap that is not heard on the first listening, a phone number or password or other item that might not be immediately recognized as compromising anonymity. Transcribers, on the other hand, spend a huge amount of time listening to the recording, and are much less likely to miss such a thing.

Confidentiality. Since bleeper technicians and transcribers are both likely to be paid research assistants, the same issues of swearing them to uphold confidentiality arise. Nothing worthwhile is gained in confidentiality by having one set of research assistants (bleepers) hide the names from another set of research assistants.

In all, the most efficient and secure way to ensure anonymity is to have the transcribers and checkers mark the transcript for items that need bleeping (using the tilde sign), after which they or another technician bleeps it.

Repercussions for transcribing process

It is a demanding task to locate all the identifying words in a recording of spoken interaction in order to anonymize them. Basically, identifying such items reliably requires much of the same kind of attention as is involved in actually transcribing the recording. Thus it is highly advisable that the task of recognizing identifiers should be carried out by the transcribers themselves. Normally this would be done during the regular process of transcription (beginning as early as Level 1 or 2). It is also a good idea to make a special pass through the transcription after it is completed in order to make sure no identifiers have slipped by unnoticed. This effectively means that transcribers have to have “clearance” to hear all the private information on the recording. This should be thought about by the researchers well in advance of beginning the project, since it means that the consent agreement between researchers and those recorded must reflect this reality, recognizing that transcribers as research team members will necessarily have full access to the unedited and unbleeped recording, including any names or other identifiers that may be spoken on the recording. Only after the recording is fully transcribed, and the transcription is carefully scanned to locate and mark any and all identifiers (e.g. by prefixing each identifying word with a tilde symbol), can the transcription be passed along to the bleepers for processing. The transcriber should provide the bleeper with a full transcription (the transcription level may be fairly low, e.g. even Level 1 is adequate for the task; but it should be complete) which distinctively marks all the items that need to be bleeped. This allows the bleeper (who may or may not be the same person as the transcriber) to go then through and modify the sound for each individual identifying word, with reasonable confidence that no identifying words will slip through and end up getting published with the original sound.

Another reason that it is imperative that the transcriber have full access to the unedited and unmodified recording before it is bleeped is that identifying names may overlap with ordinary words spoken by another speaker (see Section X). The other speaker’s words will be needlessly obscured if the bleeping is done before the transcription is complete.

[revised 22-Nov-2006]

Chapter **: Participation

Categories and Representations in Discourse-Functional Analysis

What matter who's speaking?, someone said, what matter who's speaking?
—Samuel Beckett

Introduction

Participation is a matter of the forms of engagement between actors in a discourse event. Those who would engage with each other in conversation must define themselves as co-participants within a shared context for interaction, which is to say that in one way or another they must articulate a participation framework that will shape the interpretation of their actions, and their responsibility for and role in accomplishing these actions. In many cases the participation framework will be a simple and straightforward one, without significant complications. Co-present individuals take turns at talk in succession, with each participant acting individually and on their own behalf, and speaking in their own voice. The transcription of such simple forms of participation, where there is little to do but attribute each utterance to the individual who spoke the words, can be accomplished through transparent and widely used transcription conventions. For example, we can represent the attribution of words, laughter, and so on to the person who produced them by placing a speaker label at the beginning of each turn:

- (*Lambda* SBC002: 154.38-160.42)
156 JAMIE; Aren't you guys gonna stick up for me?
157 and beat up on him or something?
158 MILES; He's bigger than [I am].
159 PETE; [##],
160 .. @@@
161 MILES; He's not bigger than you.
162 PETE; .. No.

Even when we can't recognize what words were spoken (as in line 159), we can still typically identify who spoke them. This information is registered by supplying even unintelligible words with a speaker attribution label of their own. And it's not just words that are attributed in this way. Laughter and other vocal productions are recognized as attributable actions in exactly the same way (line 160).

While the need for speaker attribution may seem all too obvious, we become more conscious of its importance when it is absent. Consider in the following (reduced) transcription:

- (*Lambda* SBC002: 167.49-171.41)
I stuck up for you today at that store.
That's true.
(0.3)
You did.

The distribution of pronouns and pro-verbs here (e.g. *you did*), among other things, leads us to suppose that the words represented in this excerpt were likely uttered by two different speakers. But we really

need to know for sure just who said what. Once we incorporate explicit indicators of speaker identity, we notice that there are actually three speakers involved here:

- (*Lambda* SBC002: 167.49-171.41)
 - 1 PETE; I stuck up for you today at that store.
 - 2 HAROLD; That's true.
 - 3 (0.3)
 - 4 JAMIE; You did.

When Pete says *you* in line 1, he is addressing Jamie. But it isn't Jamie that responds first, it's Harold. This obviously makes a difference to the participants in the interaction, and so must become an integral component of any adequate transcription.

One reason that speaker attribution is almost universally represented in diverse transcription practices is that the participants themselves care a great deal about who said what. In fact, participants may feel they can't even interpret an utterance until they know who said it. This may lead a participant to explicitly raise the question of who is speaking, as in the following examples (presented in simplified transcription):

- ((LSAC 1321-01))
 - A; **Who said that?**
 - B; Kevin.
- ((LSAC 1682-02))
 - A; **Who said that?**
 - B; I did.
- ((LSAC 1609-01))
 - A; I had talked to a lady
 and she said it was up to you and um,
 - B; **Who said that?**
 - A; I ain't --
 I ain't saying no names.
 - B; @ Oh listen.

These examples mostly involve people telling stories about prior conversations, in which the source of the utterance in question is a third party, rather than someone participating in the current conversation. But we care just as much about who said what to whom in a live conversation in which we are presently involved. It's just that here it's not necessary to state explicitly who said what, because the information about who's responsible for a live utterance is carried more directly by the voice. It's by their distinctive voice quality that we are able to reliably distinguish one participant from another, even if we lack visual access to the speaker (whether due to darkness, distance, blindness, telephone, etc.).

It might seem that no serious researcher would ever omit something so basic as the identity of the speaker in a conversation. But in fact this has been done on a number of occasions, sometimes with problematic consequences.

[EXAMPLE TO BE INSERTED: transcription published without speaker labels]

While readers of a simple dyadic conversation may sometimes be able to guess where the speaker change occurs, and so “reconstitute” the missing speaker information, this is likely to fail once we consider conversations with 3 or more parties.

Another indicator of the importance of attributing vocal acts to participants is to consider cases where there is a real issue as to whose voice it is. One occasion where uncertainty may arise is when there is laughter by several in a group. Laughter may distort voice quality in ways that make it harder to recognize. (This is especially true for a transcriber who is unfamiliar with the voices in question.) In the example below, if only lines 8 and 9 are heard in isolation, the extended laughter of Pamela and Darryl sufficiently distorts their voices that it is not immediately clear whose voice is heard, nor when the laughter shifts from Darryl’s voice to Pamela’s. (This ambiguity is of course obscured in the transcription, which makes it clear who is laughing when—perhaps too clear. Yet again, in the original face-to-face interaction, it would have been clear to the participants who is laughing.)

- ((*A Book About Death* SBC005 1140.52- 1159.74))
- 1 PAMELA; [(H) Maybe] he’s a <MARCATO> very old soul </MARCATO>.
- 2 (Hx)
- 3 DARRYL; .. (Hx)
- 4 ...
- 5 PAMELA; Plea:[se].
- 6 DARRYL; [@Like @Old] [₂@King @Cole]
- 7 PAMELA; [₂@@@]
- 8 DARRYL; (LAUGHTER)[₃.]
- 9 PAMELA; [₃(LAUGHTER)]
- 10 Hm:.

So providing speaker labels is a basic requirement for almost any transcription (unless it is a pure monologue). Normally it’s an understood part of what we know about participation in any conversation that we can attribute every utterance to a particular speaker. Without an explicit representation of such attributions of responsibility in a transcription, the analyst is missing out on information that participants have, which moreover is highly salient to them.

While the practice of using speaker labels is more or less universal among serious approaches to transcription, it is easy to overlook what we really mean by it. Each speaker label represents the attribution of responsibility for all of the following words and actions to a particular named individual, whose identity remains constant throughout the interaction. An attribution label inserted at a given location remains in force until the next speaker attribution label is encountered within the linear stream of the discourse representation. The convention of speaker attribution is somewhat unusual among transcription categories, in that it doesn’t seem to represent an actual sound corresponding to what is written. We do not imagine that each time we see a speaker’s name in the left margin, the name was spoken aloud. Consider the following example:

- ((*A Book About Death* SBC005 442.12-448.25))
- 1 PAMELA; (H) I’m thinking one thing my mother always used to say:,
- 2 when I wouldn’t go bicycling with my [father],
- 3 DARRYL; [@~Pamela],
- 4 @you @are [₂@@@,

There is a difference between *Pamela* and *PAMELA* in this transcription. The former (as written in line 3) was spoken aloud by Darryl. The latter (as written at the beginning of line 1) was spoken aloud by no-one. Nor does the speaker label *PAMELA* correspond to an event specifically localized at the beginning of line 1. In acknowledging that this is a departure from representing the actual sound of names spoken, we might ask whether the speaker label is in fact a true transcription category at all. Should it rather be considered a notation for coding, or analysis—a convenient fiction to help the analyst keep track of who said what for research purposes? Is it a distortion, an imposition on the reality of the original interaction and the original words as uttered? One way to approach this question is to ask whether there is in fact any audible correlate of the speaker label available to participants in the interaction. Perhaps surprisingly, the answer is yes. While the speaker's name as such is not audible, a little reflection will show that the speaker's identity is, in fact, quite distinctly audible. In the most intimate of ways, the unique character of each individual's voice imparts an indelible stamp on every word they utter. The voice quality is audible to participants continuously, during the time that a given speaker is talking. If we wanted to present a visual analogue of the speaker's continuous display of their identity while speaking, we could try to make it more obvious in a transcription just what it means to attribute a sequence of words to a given speaker. For this we would need a notation that "colors" every word uttered in a turn, just as the speaker's voice colors every word. (One can imagine iconic conventions for representing each speaker's voice using a different color on a computer screen, or using a different font on paper.) So the speaker attribution label is after all a legitimate transcription category, even if its way of representing the audible voice is not particularly iconic.

The above cases of participation are rather straightforward, and are easily represented through standard notational conventions. In other cases, the question of how to analyze the participation framework is not so simple. There may be shifts in the participation framework within a conversation, and even rather complex negotiation among participants as to just what the current framework is. But given the manifest salience of the attribution of responsibility for talk and other participant actions, it seems that participants attend closely to the local organization of the participation framework. This makes it imperative to have a set of transcriptional devices that is sufficiently flexible to represent the necessary distinctions, while remaining reasonably accessible and transparent to the reader of transcriptions.

Recipient

In seeking to represent the participation framework that speakers in a conversation are orienting to, it could be argued that just as important as the role of speaker is the role of the recipient, or addressee. This is suggested by the following exchange (in reduced transcription):

- ((LSAC 1398-01))
- | | |
|--------|--------------------------------------|
| ALEPH; | Wait wait wait |
| | who who who said this to who? |
| BETA; | Chris did to Matthew. |

But usually the recipient is not actually specified in a conversational transcription. This may be for several reasons. First, it may seem too obvious. While this is doubtless valid when there are only two speakers, it becomes less obvious as soon as there are three participants in a conversation, and hence two distinct candidates for the recipient of any utterance. One could take as the default that all the co-present participants will be included as the participants, but this is not always the case. The following example is taken from a meeting of board members of a banking institution, where the topic is whether to approve a loan that has been applied for:

Version 1: no recipient marking

- (Bank Products SBC014: 721.95-723.83)

JIM; [since you seconded it,
 I assume so are [2you].
FRED; [2Yeah].
KURT; Yeah.

In a conversation or small face-to-face meeting like this one, it is normally clear who the speaker is at any given moment. But it is sometimes less clear who the intended recipient is, when there are more than two people present. When Jim says *Since you seconded it, I assume so are you*, two of the other three participants in the meeting respond (Fred, Kurt). In fact both say the same thing (*yeah*). But Jim's utterance is apparently only *addressed* to one of them. Who is "you"? The above transcription format doesn't make this clear. But if the participants know who the recipient is, this information is useful and relevant and should be made available to the analyst as well. One solution is to specify the addressee or recipient in the transcription—assuming it can be determined from the recording (whether by gaze direction, participant response, or other forms of discourse-sequential evidence). This is done by writing the name of the recipient following the name of the speaker, with a backslash between them, along the pattern of: **SPEAKER\RECIPIENT**;. Retranscribing the above example in this format clarifies the nature of the interaction:

Version 2: recipient marking

- (Bank Products SBC014: 721.95-723.83)

JIM\FRED; [since you seconded it,
 I assume so are [2you].
FRED; [2Yeah].
KURT; Yeah.

The next example, taken from an earlier stretch of the same meeting, provides an interesting contrast. While the same set of potential addressees/recipients is present, in this case the speaker specifies explicitly who among them is his intended recipient:

Version 1: no recipient marking

- (Bank Products SBC014: 413.95-421.02)

FRED; .. I'll second] then ~Joe.
KURT; (H)]
JOE; Okay.
...
JIM; So it's moved and seconded .. to uh,

One response to this explicit specification would be to say there is no need to indicate the addressee, since it can be derived from the speaker's words directly. The opposite conclusion can be drawn as well: since it is can be reliably derived from the speaker's words, the transcript should represent it, as follows:

Version 2: recipient marking

- (*Bank Products* SBC014: 413.95-421.02)

FRED\JOE; .. I'll second] then ~Joe.

KURT; (H)]

JOE; Okay.

...

JIM; So it's moved and seconded .. to uh,

The challenge is that it is not always clear if one among a set of potential addressees is being singled out as THE addressee, and if so, it is not always clear who the intended recipient is (regardless of whether the recording is audio or video).

Consider the following example:

- (*Lambada* SBC002: 154.38-159.02)

156 JAMIE; Aren't you guys gonna stick up for me?

157 and beat up on him or something?

158 MILES; He's bigger than [I am].

159 PETE; [##],

160 .. @@@

Here Jamie's *you guys* might seem to be transparently addressed to all the other participants present. But in addition to Miles and Pete, Harold is also present. Since Jamie is actually asking Miles and Pete to beat up on Harold, he can't be an addressee. This example suggests a more fundamental reason for leaving out information about recipients. It's quite a complex matter, with a lot of shifting going on within the unfolding development of any lively multi-party conversation, which may lead to a significant amount of ambiguity as to who the recipient is supposed to be. While video can certainly help in some cases, the ambiguity of reciprocity often holds for the participants as well as the analyst, so it is not necessarily resolved by access to video. The reality is that even in the most labor-intensive transcripts done by conversation analysts, there is seldom a consistent indication of the recipient throughout a transcription. It simply requires too much high-level analysis and interpretation.

That said, there are some cases where the recipient can be clearly identified, and where it makes a real difference in how the interaction is interpreted (by both conversational participants and the analyst). In such cases it is useful to indicate identity of the recipient.

- recipient-addressee JIM\FRED; post-backslash field marks recipient

- (*Hey Cutie-Pie* SBC028: 480.2-484.5)

JILL\JEFF; <TEL> Oh wait,

.. Jeff,

can you hang on just a sec </TEL> ?

JEFF\JILL; <TEL> Ye:s </TEL> .

JILL\KATE; What did you say?

KATE\JILL; #I'm .. #gonna #go #over #there.
 JILL\KATE; (H) Oh you are?

One can also place all of the information about participation framework in the fielded format, using the slash character to separate fields. In the end, this may be the cleanest looking transcription, and the one that is easiest to read and understand:

- (*Hey Cutie-Pie* SBC028: 480.2-484.5)
 JILL\TEL\JEFF; Oh wait,
 .. Jeff,
 can you hang on just a sec?
 JEFF\TEL\JILL; Ye:s.
 JILL\KATE; What did you say?
 KATE\JILL; #I'm .. #gonna #go #over #there.
 JILL\KATE; (H) Oh you are?

Voicestream

One of the things that makes a participation framework consequential to participants is that individuals who are thus engaged with each other orient to a single shared timestream. They must manage their own contributions, verbal and otherwise, in relation to the shared timestream in ways which respond interactively to on-going, and potentially unpredictable, developments in the current interaction. This is what it means for persons to be engaged as social actors within a participation framework: their actions are interpreted within the shared interactional space articulated by the framework.

Agency attribution

One important aspect of participation concerns which agents are recognized as co-participants within a given context of interaction. As already mentioned, this is important because of the co-participants' orientation to a shared timestream, which is a powerful determinant of the organization of a given interaction.

There are a number of different categories of participation framework, and a given participant can participate in different ways on different occasions. Some relevant symbols are as follows:

- speaker/agent attribution JILL; semicolon following name in CAPS

Identifying participants

It's not always a trivial task to identify who is speaking. For cases when it is simply not possible to identify the speaker, we can so indicate, using the # symbol for "unintelligible," but placing the symbol in the location for representing the speaker. Similarly, if we are uncertain about the (tentative) identification of a speaker, this can be indicated by prefixing the symbol for uncertain transcription to the speaker label. And finally, if the utterance can be attributed to either one of two alternative speakers, we can represent this by writing each of the alternative names in succession, each preceded by the uncertain hearing sign.

- unidentified speaker #; cannot determine whose voice

- uncertain speaker #JILL; not sure that speaker attribution is correct
- alternative speakers #JILL_#JOAN; not sure which of 2 speakers is speaking

Joint action

Participants sometimes perform an action jointly, so it is useful to have a notation for assigning responsibility for the action to more than one individual. This can be represented by listing each name, with a plus sign between them, as follows:

- multiple speakers JILL+JIM; for simultaneous joint action of individuals

In some circumstances, this could be used for the same words uttered simultaneously by two or more different people:

- (*Lambada* SBC002: 304.50-308.16)

1	MILES;	Does she have a what?
2	JAMIE+HAROLD;	A ma:n.
3	PETE;	@@@

However, for a variety of reasons it is probably better in most such cases of simultaneous speech to use the speech overlap notation, which has the advantage of giving each speaker their own separate line:

- (*Lambada* SBC002: 304.50-308.16)

1	MILES;	Does she have a what?
2	JAMIE;	[A ma:n].
3	HAROLD;	[A ma:n].
4	PETE;	@@@

This allows for differences between the speakers, such as in the presence of distinctive prosodic features.

In some cases, there are too many participants to reliably distinguish who are responsible for a given action. For this we can use notations like *AUDIENCE*, followed (if necessary) by two plus signs to explicitly indicate that some (indefinite) plurality of participants is involved.

- group of speakers AUDIENCE++; for simultaneous joint action of crowd

Voice

Sometimes it is useful to indicate that a speaker is mimicking the voice of another. While rarely indicated in an actual transcription, it may be useful in some specialized cases, as when a speaker is performing more than one “part” when telling a story, reading the dialogue in a play, and so on.

- voicing ALINA=DOG; speaker uses voice of another
- (*Cuz* SBC006: 699.00-710.32)

ALINA;	If we went to parties,
	and there was somebody who came up to try: to talk to us,
	(H) he'd keep em away.

```

.. # Just like a little t-,
um,
Australian sheep dog.
...
Running around.
ALINA=DOG; .. <VOX> rar rar rar rar rar [rar </VOX>].
LENORE;                                     [@@]@@ @@[2@@2]

```

In cases where the speaker is using vox or the “voice of another,” an alternative is to use the vox notation as the place to specify the nature of the voice being used:

- (Cuz SBC006: 699.00-710.32)
ALINA; .. <VOX=DOG> rar rar rar rar rar [rar </VOX>].

Non-interactive agents: Non-human

More and more in the modern world, we encounter entities that intervene in our conversations without being fully human. Toys, electronic games, computers, and so on may produce a noise or an action that simulates a human one, or at least becomes relevant to the humans who interact with it. The fact that participants orient to such sounds, and care to identify their sources, means that an adequate representation of discourse must come to terms with the issue of mechanical agency. One way to acknowledge the role of non-human entities that are capable of intervening in a conversation in a consequential way is attribute actions to identifiable mechanical entities. While this can be done using the existing mechanisms for attribution of actions to humans, it is important not to overestimate the role of a mechanical agent in an interaction. It would be misleading to give a talking doll or a squawking computer the same status as a person who is contingently responsive to the immediate situation. We address the two demands by recognizing a category of *non-interactive agent*, which applies to machines and other entities which can produce sounds or other actions that become relevant to participants in a conversation, but which cannot “answer back” interactively if spoken to.

Non-participant: When the agent or source of the sound or action is not a full participant in the targeted interaction, this is indicated by prefixing the agent’s name with a hyphen/minus (-). (The agent so indicated may not be capable of full interaction, or may be a full participant in some other interaction, but not the present one.)

- non-interactive agent ±COMPUTER; marks source of sound or event (hyphen)

In the midst of the dynamic interaction between participants that constitutes conversation, there may appear contributions by certain kinds of agents who are not full participants. Sources of actions that may impact a conversation include entities which cannot be considered full-scale interaction. If a clap of thunder is heard and responded to by participants, the non-human event must be part of a representation of that discourse, if we wish to interpret the human response. The same is true of a dog’s barking, if the conversational participants respond to this.

Examples:

```

±DOG;          ((BARK)) ((BARK))
±COMPUTER;    ((BEEP)) ((BEEP))
±AMBIENT;     ((THUNDER))

```

±DOLL; words
 ±TOY; words

Non-interactive agents: Human

On another level, if a conversation is taking place with a radio or television going on in the background, the question arises of how to deal with the spoken words of the radio announcer or television personality. In a sense, the radio announcer may play a role that is not altogether different from that of an answering machine or a talking computer, or even a barking dog. Neither the dog nor the radio announcer can answer back when spoken to, at least not in the manner of a full participant. Nevertheless, because participants may attend to what non-interactive agents do, and even respond to their actions, the non-interactive contribution should be represented in some way (assuming it is demonstrably relevant).

Still, it remains important to distinguish the contributions of non-participants from those of full participants. And so we introduce a way of representing the difference between a full participant in the present (targeted) interaction, and an agent which is not a full participant. The latter sort of agent is marked with a hyphen (-) prefixed to their name, while the full participant is left unmarked.

±ANNOUNCER; -That's the news.
 ±HOST; -Our next guest is,

It's not just mass media that lead to humans becoming non-interactive agents in an interaction. If someone plays back an answering machine message from their friend, they will hear a recording of a real person speaking, and that may be of interactional significance in the unfolding of events in the present context. Yet the recording will not be able to answer back, nor to take account of how its words are being received by the listeners. For these reasons, it is useful to mark the recorded voice as distinct from a live voice, even if the voice is that of the same individual.

Sometimes we are uncertain about whether or not a particular sound comes from a human interactant. For this we can combine the symbol for non-interactive agent with the symbol for uncertain transcription, yielding the following notation:

- uncertain interactiveness #±; unsure if sound source is a human interactant

Framework

The participation framework category specifies what the relevant group of participants is for a particular interaction.

If every recording of interaction contained only one participation framework, it would be a simple matter to specify it at the outset, in a header which would apply to the whole transcription. Often enough this is indeed the case, and no further specification of the participation framework is needed beyond the opening statement. But in many conversations it happens that the participation framework will shift in various ways during the course of the interaction. The shifts may be subtle or obvious, brief or lengthy. Given such variability, it is useful to have a more flexible means of representing local changes in the participation framework. As usual in transcription practice, it is important to combine flexibility with systematicity.

- participation framework ±HOST/TV; 1st post-slash field marks framework (slash)

This notation is often used for cases where a speaker switches between two contexts, when it is useful to be able to distinguish the contexts. One common example of this use of the participation framework notation is when a speaker switches between talking to someone over the telephone and talking to people who are immediately co-present at the speaker's home. Consider the following example:

- (*Hey Cutie-Pie* SBC028: 480.2-490.16)

JILL/TEL; Oh wait,
 .. Jeff,
 can you hang on just a sec?
 JEFF/TEL; Ye:s.
 JILL; What did you say?
 KATE; #I'm .. #gonna #go #over #there.
 JILL; (H) Oh you are?
 O[kay].
 KATE; [Yeah].
 ...
 JILL; Okay,
 I'll come over.
 ...
 JILL/TEL; [Um],
 JEFF/TEL; [Where]'s ~Kate going.
 JILL/TEL; .. She's going next door?

Alternatively, because this is part of a long phone call, with only a short exchange taking place in the local environment, one could reverse the markedness, and indicate the face-to-face exchange as the marked case, as FACE:

- (*Hey Cutie-Pie* SBC028: 480.2-490.16)

JILL; Oh wait,
 .. Jeff,
 can you hang on just a sec?
 JEFF; Ye:s.
 JILL/FACE; What did you say?
 KATE/FACE; #I'm .. #gonna #go #over #there.
 JILL/FACE; (H) Oh you are?
 O[kay].
 KATE/FACE; [Yeah].
 ...
 JILL/FACE; Okay,
 I'll come over.
 ...
 JILL; [Um],
 JEFF; [Where]'s ~Kate going.
 JILL; .. She's going next door?

An alternative notation would mark the start and end of the face-to-face exchange, using angle brackets:

- (*Hey Cutie-Pie* SBC028: 480.2-490.16)
 JILL; Oh wait,
 .. Jeff,
 can you hang on just a sec?
 JEFF; Ye:s.
 JILL; <FACE> What did you say?
 KATE; <FACE> #I'm .. #gonna #go #over #there.
 JILL; (H) Oh you are?
 O[kay].
 KATE; [Yeah] </FACE>.
 ...
 JILL; Okay,
 I'll come over </FACE>.
 ...
 JILL; [Um],
 JEFF; [Where]'s ~Kate going.
 JILL; .. She's going next door?

One could even go so far as to mark both contexts explicitly:

- (*Hey Cutie-Pie* SBC028: 480.2-484.5)
 JILL/TEL; Oh wait,
 .. Jeff,
 can you hang on just a sec?
 JEFF/TEL; Ye:s.
 JILL/FACE; What did you say?
 KATE/FACE; #I'm .. #gonna #go #over #there.
 JILL/FACE; (H) Oh you are?

However, this would quickly get rather tedious to read. Better to pick the prevailing context for a given interaction as the (locally) unmarked case, and only write in the framework when there is a departure from the local norm.

Note that an alternate notation for indicating a context like a phone conversation is to use the manner/quality notation, signaled by angle brackets. The decision as to which convention to use may depend in some cases on which is more readable, or otherwise more convenient.

- (*Hey Cutie-Pie* SBC028: 480.2-490.16)
 JILL; <TEL> Oh wait,
 .. Jeff,
 can you hang on just a sec </TEL> ?

JEFF; <TEL> Ye:s </TEL> .
 JILL; What did you say?
 KATE; #I'm .. #gonna #go #over #there.
 JILL; (H) Oh you are?
 O[kay].
 KATE; [Yeah].
 ...
 JILL; Okay,
 I'll come over.
 ...
 JILL; <TEL> [Um],
 JEFF; <TEL> [Where]'s ~Kate going </TEL> .
 JILL; .. She's going next door </TEL> ?

Examples of additional uses are as follows:

±ANNOUNCER/TV; ¬That's the news.
 ±HOST/RADIO; ¬Our next guest is
 ±SINGER/MUSIC; Tell me what'd I say,
 ±COMPUTER/GAME; words
 ±TOY/GAME; words

In fact, it is useful make clear the special status of certain events and actions within the larger conversation, as constituting a separate framework for (possibly limited) participation, distinct from that which is in force for the current targeted interaction. For example, participants in a television talk show respond to each other, but not to the people in the living-room audience who are talking about them—or even talking to them. Suppose we are recording in a particular living-room, in which participants from time to time attend to and react to what is being said on the television. Here the television represents one participation framework, while the living room constitutes another participation framework. On a recording of the living-room interaction, one may find that the first is embedded in the second. Talk from the pre-recorded broadcast may at times intrude on, and become consequential for, developments in the living-room context.

This and other related cases raises the need to specify the participation framework, in some way which is sensitive to the possibility of change in the participation framework, within the larger context of the overall ongoing interaction. We allow for shifting participation frameworks by writing the special context in a distinct field. Some relevant participation frameworks are as follows (all of which are options to be written following the slash, as above):

/TEL Participation framework is mediated by a full-duplex (real-time 2-way) interactive channel (e.g. telephone)
 /CELL Marks use of cellular telephone (only where this distinction may be relevant)
 /FACE Participation framework is face-to-face interaction (usually unmarked)

/DOOR	Participation is mediated by intervening door (e.g. door knocking, doorbell ringing)
/REC	Previously recorded voice (e.g. when answering machine message is played back)
/1	Split conversation: utterance oriented to conversation #1
/2	Split conversation: utterance oriented to conversation #2
/GAME	Participation framework is playing a game (e.g. with other participants or a computer)
/PET	Talking to a pet (possibly, also: communication from a pet)
/BABY	Talking to a baby, and communication from a baby
/TV	Television program talk
/RADIO	Radio talk
/MUSIC	Pre-recorded music (or video, etc.)
/BROADCAST	Broadcast talk
/AD	Pre-recorded advertisement, e.g. interrupting live radio talk show
/PA	Public address system (amplified audio addressed to crowd)
/OFF-MIC	Off-microphone talk (asides) during public address or broadcast context
/FUTURE	Self-consciously addressing words to the future listening audience of the present recording
/NATURE	Sounds of nature
/ENVIRONS	Ambient sounds from environment

Participation Framework: Miscellaneous Examples

The following example illustrates several of the relevant categories and notations for representing various aspects of participation:

- (*Hey Cutie-Pie* SBC028)
- | | | |
|-----------|---------------|---|
| 00.000000 | ±JILL/TEL; | ((DIAL_TONE)) |
| 00.272834 | | ((SPEED_DIAL)) |
| 02.377143 | | (2.5) |
| 04.887800 | | ((RING)) |
| 05.305760 | | (3.4) |
| 08.717642 | ±JEFF/REC; | ((PICK-UP)) |
| | | (H) Hello:. |
| | | You have reached Jeff's answering machi:ne. |
| | | (H) So plea:se leave a messa:ge, |
| | | a:nd I'll get back to you, |
| | | #as soon as I ca:n. |
| | | Thank you. |
| | | (0.2) |
| | | Bye. |
| 17.459955 | ±MACHINE/REC; | ((BEEP)): |
| | | (0.3) |
| 17.939592 | JILL/TEL; | (TSK) He:y cutie pie:, |
| | | are you the:re? |
| | | (0.3) |
| 21.269478 | | (H) Or are you sleeping:. |
| 23.307029 | | (0.9) |
| 24.311293 | | (TSK) (H) Ah: I guess you're sleeping. |

26.203719

Or you're not there.

[NB: For additional information, see also "Non-Vocal Events"]

[rev. 22-Nov-2006]

Field Structure in Discourse Representations

Introduction

The representation of discourse that is implicit in the typical discourse transcription treats the data as a simple text file. Such representations are relatively free-form, with few built-in assumptions about the internal structure of the representation, beyond the linear sequence of represented events (e.g. words, pauses, turns, prosodic events, and so on) which make up the ongoing stream of spoken discourse. For the most part, the discourse representation treats the linear sequence of represented events as if they were beads on a string—discrete events, arranged in strict linear sequence. For many purposes, this unstructured treatment is adequate, and may even be preferable to a more structured approach. The lack of an imposed structure may be thought to reflect the freedom and unpredictability of discourse. This seems reasonable from a perspective which assumes that discourse (in contrast to grammar, with its constraints and structures) is limited only by the creative ways that speakers find to use language to engage with each other.

But at another level, discourse reveals a great deal of structure, even in the free-wheeling realm of spoken interaction. Certain aspects of the structure of discourse need to be acknowledged in the discourse representation, through a systematic accounting of the different kinds of information that are entailed by the phenomena of discourse. Among the types of information about an utterance that we may find it necessary to distinguish, for example, is the contrast between the words spoken and the identity of the person who spoke them. Because these are two very different kinds of information, the difference between them is ordinarily reflected overtly in the way they are handled in the transcription. Indeed, virtually all transcriptions mark a careful distinction between the words spoken and the speaker identity label (assuming speaker labels are included at all), for example by placing the speaker labels in a column at the left margin of the page, and perhaps marking them with a special symbol such as a colon or semi-colon placed at the end of the label. But many discourse representations more or less stop there, formally distinguishing only these two types of transcription information. There may be other kinds of information implicit in the transcription, but these are often obscured by a failure to recognize what they are, and to structure the representations in ways that allow transcribers and transcription users to recognize the distinctive information they carry. The present discussion seeks to go further, building on the understanding that discourse representations have inherent structure, like that which is implicit in the contrast between speaker label and utterance text. We ask what additional distinctions should be made recognized regarding the domains of structured information that may be usefully and appropriately represented within a discourse transcription.

Here it is useful to apply the concept of *field* to our representations. This concept, originating in the context of database design, implies the recognition of distinct categories of information, each of which may be pertinent to a given object of interest. In this case, we can think of a single line of a transcription (corresponding to single unit of discourse, such as an intonation unit) as being the object of our interest. Can we recognize distinct fields that should be distinguished within the lines of any transcription? We take up this question in the rest of this chapter, as we build toward the recognition of a framework for representing and managing the information in discourse transcriptions in terms of a kind of field structure. Field structure seems to be implicit in certain practices for transcribing spoken discourse, but it often goes unrecognized, and is treated in haphazard ways. There is a lot to gain by becoming aware of the inherent structure of discourse representations, and making certain aspects of it explicit.

Text & Pre-Text

The first candidate for field structure in discourse transcription is implied by the contrast between the participant's actions (words, gestures, movements) and everything else—including, in particular, the identities of the participants themselves, as agents of the actions in question. Put simply, we want to distinguish at the outset between aspects of the transcription which represent the words spoken and those which represent the identity of the person who spoke them. At the most basic level, these can be seen as two distinct fields, each containing a distinct type of information. More broadly, if we look at a transcription line-by-line, we see that it is useful to make a broad distinction between what we may call the *text* and the *pre-text*. The *text* consists of the specific events that take place one after another, as represented in the main portion of the transcription, such as the words uttered, gestures produced, and other actions that may be interspersed among the words spoken. (Note that the word *text* here is not meant to connote writing, nor to suggest that speaking is equivalent to writing; it is used simply as a cover term for the written representation of the participants' actions.) The *pre-text* is literally everything that comes before the text in a line of transcription. What is included in the pre-text will vary considerably depending on the transcription practice, but typical examples of pre-text categories would be speaker labels and line numbers. (Other information types will be discussed below.) In general, information in the text pertains to the actual overt actions that can be temporally located within the stream of discourse, while information in the pre-text presents information about the context of these utterances. The former includes rapidly shifting events such as utterances, gestures, and so on, while the latter includes relatively less changeable or more timeless information, such as the identity of the agent of an action. While the text presents a moment-by-moment representation of reality, the pre-text instead presents a broader contextualization for the text events. In addition, the pre-text may contain meta-transcriptional information of various kinds, such as line numbers for reference purposes.

Because the difference between text and pre-text is so important, it is crucial to mark it consistently. Many transcription systems use a specific symbol to mark the end of the speaker label, which in effect can be considered as a dividing line between pre-text and text. In the present transcription system, the semi-colon can be taken as marking not only the end of speaker attribution, but also the beginning of the text itself, which immediately follows. (While it is common to use a colon for marking speaker labels, this can create a problem if the colon is also used to mark lag or lengthening. Such a fundamental ambiguity can create serious problems for researchers. Although it may be possible to use transcription-parsing strategies to discriminate between a colon which marks length and a colon which marks a speaker label, this can be risky. The division between text and pre-text is such a fundamental aspect of the architecture of any discourse transcription that it is better not to create the ambiguity in the first place.)

Tabs and Fields

Tabs may be used in transcribing to help organize the presentation of transcription information, displaying it in a usefully iconic way. For the most consistent transcription practice, tabs should be used for one purpose only; for example, to distinguish between separate “fields” of data in each line. For a basic transcription, there are two fields. The first field in each line indicates who the speaker is (via a speaker label written in capital letters, followed by a semi-colon). The second field represents the actual utterance, or transcribed speech. Representing the two fields requires one tab, which in effect separates the two types of information.

The fields for speaker label and text should be separated by a single tab character. Consequently, in a basic transcription with just these two fields, every line should contain exactly one tab. This holds true

whether the speaker label is overtly specified or not in a given line. If there is a speaker label in the line, the tab comes immediately after it; if there is no speaker label in the line, the tab is the first character in the line. Either way, the tab immediately precedes the text. (Note that an alternative to employing tabs in this way is to use tables, as provided by word processor, spreadsheet, or database software. In using tables, each transcription field would correspond to a different column in the table. Most of what is said here about strategies for using tabs can be translated more or less directly into strategies for using columns in tables.)

More generally speaking, in a more sophisticated transcription there can be more than two fields. Each field will represent a different type of data, presented in a different portion of the line, e.g. with a different location on the paper or screen. The general rule is that for a transcription with a certain number of fields (call the number *N*), the number of tabs in each line should be one fewer (*N*-1). (If using tables, the number of columns will be the same as the number of fields.)

Parsing Transcriptions

The content and order of the fields used should be specified (e.g. in a header). In general, it can be parsed without too much difficulty. For example:

23 87.352 88.917 ±JOAN=CHILD/REC\JIM; Hey baby,

Given the conventions for marking implicit fields in Discourse Transcription, this can be unambiguously parsed as:

Index	StartTime	EndTime	Interactive	Agent	Role	Framework	Recipient	Text
23	87.352	88.917	±	JOAN	CHILD	REC	JIM	Hey baby,

There are a variety of transcription categories found in the pre-text that have some bearing on the parsing of transcriptions into fields in this way. A listing of relevant categories is given in Table 1.

Table 1. Transcription Categories Relevant to Field Identification

Participation

31. speaker attribution	JILL;	semicolon following name in CAPS
32. role/voice	JILL=CHILD;	speaker uses voice of another
33. participation framework	±HOST/TV;	post-slash field marks framework
34. recipient	JILL\DOG;	post-backslash field marks recipient

Time

35. timestamp	<TIME=154.73>	time in seconds from start of recording
36. index/line numbers	<LINE=1>	each line numbered (and/or timestamped)

Meta-transcription

37. comment line	\$	for demographic or database info
------------------	----	----------------------------------

In addition, there are some cases that go beyond what can easily be represented in a simple field structure. This includes especially cases where there is more than one value for the same field, as when two or more speakers are represented in the speaker field. (There are more sophisticated structures which can handle this kind of information, such as a relational database which allows a many-to-one relation

between agent and action.)

Table 2. Additional Transcription Categories

Multiple Participation

- | | | |
|-----------------------|-------------|--|
| 38. multiple speakers | JILL+JIM; | for simultaneous joint action of individuals |
| 39. group of speakers | AUDIENCE++; | for simultaneous joint action of crowd |

Field Structure and the Transcription Delicacy Hierarchy

Tables 3-5 below present several alternative ways of thinking about the organization of transcriptions into structure fields, in relation to the various ways of prioritizing information in a hierarchy of transcription delicacy. In Table 3, the columns represent different field types in a discourse transcription: text or utterance, agent or speaker, index or line numbers, start time, end time, participation framework, recipient or addressee, role or voice, and interactiveness. The rows represent different types of transcription, ranging from a minimal or basic transcription that includes only one or two fields, up to a maximal transcription that distinguishes all of these fields, representing the highest level of delicacy. In between the minimum and the maximum, the decision about what fields a transcription should contain will depend on what kinds of information are most relevant to a particular researcher. For example, a researcher whose first priority is to understand the exact temporal organization of discourse might prioritize the fields for start time and end time more highly, as represented in the delicacy hierarchy which is implicit in Table 3 below. In each transcription type (each row) listed in Table 3, the fields which are to be included in that transcription are marked with a bullet.

Table 3. Fields in Discourse Transcription, by delicacy (temporal focus)

		Granularity								
		←Lower					Higher→			
	Transcription Type	Text	Agent	Index	Start Time	End Time	Framework	Recipient	Role	Interactive
•	Minimal	•								
•	Basic	•	•							
•	Indexed	•	•	•						
•	Timestamped	•	•	•	•					
•	Aligned	•	•	•	•	•				
•	Framework	•	•	•	•	•	•			
•	Participation	•	•	•	•	•	•	•		
•	Participation-Role	•	•	•	•	•	•	•	•	
•	Maximal	•	•	•	•	•	•	•	•	•

On the other hand, a researcher whose first priority is to understand the interactional dimensions of the participation framework might prioritize the information differently, as indicated in Table 4.

Table 4. Fields in Discourse Transcription, alternative hierarchy (interactional focus)

		<i>Granularity</i>								
		← <i>Lower</i>				<i>Higher</i> →				
	Transcription Type	Text	Agent	Framework	Recipient	Role	Interactive	Index	Start Time	End Time
•	Minimal	•								
•	Basic	•	•							
•	Framework	•	•	•						
•	Participation	•	•	•	•					
•	Participation-Role	•	•	•	•	•				
•	Participation-Full	•	•	•	•	•	•			
•	Participation-Index	•	•	•	•	•	•	•		
•	Participation-Time	•	•	•	•	•	•	•	•	
•	Maximal	•	•	•	•	•	•	•	•	•

Table 5 presents the same list of transcription types as Table 3, but arranges the columns according to the linear sequence of each field as it appears in a typical transcription. In this way of looking at the organization of a transcription, the descending rows represent successive transcription types, each with a slightly higher granularity/delicacy than the previous row. The fields (columns) appear in the same left-to-right sequence that they would in a line of a typical transcription, if it includes such information.

Table 5. Fields in Discourse Transcription, by linear sequence

		Sequence								
		Left Right								
	Transcription Type	Index	Start Time	End Time	Interactive	Agent	Role	Framework	Recipient	Text
17.	Minimal									•
18.	Basic					•				•
19.	Indexed	•				•				•
20.	Timestamped		•			•				•
21.	Aligned		•	•		•				•
22.	Aligned-Indexed	•	•	•		•				•
23.	Framework					•		•		•
24.	Participation					•		•	•	•
25.	Participation-Role					•	•	•	•	•
26.	Participation-Full				•	•	•	•	•	•
27.	Maximal	•	•	•	•	•	•	•	•	•

[rev. 22-Nov-2006]

Appendix 1a

Transcription Delicacy Hierarchy

[Version 2.3]

Table 1. *Transcription Delicacy Hierarchy for Discourse Transcription*

	Level	1	2	3	4	5	6	7	
	Type	Foundation	Basic	Boundary	Interaction	Intonation	Prosody	Multimodal	
1	Core	words speakers intonation units sequence	closure/finality appeal/question pause laughter marginal words	overlap truncation/cut-off breath pause duration lag/lengthening hold/micropause utterance start latching linking discontinuous IU	vocalism manner/quality vox affect non-vocal event co-action gaze gesture body event duration	accent appeal onset onset/head/tail tone tune/contour chant/stylization terminal pitch key/baseline pitch reset booster	declination unit prosodic sentence paratone isotony/tone concord boundary attenuation rush/anacrusis stress accent unit isochrony register/range	audio waveform timeline/score pitch curve amplitude curve spectrogram energy curve duration curve intonation stylization indexed semiosis indexed context	1
2									2
3									3
4									4
5									5
6									6
7									7
8									8
9									9
10									10
11	Meta		comments unintelligible uncertain pseudograph header	index recording alignment timestamp	non-interactive source participation frame recipient/addressee participant gloss alternative hearing	code-switch disfluency/repair pause attribution turn/backchannel	phonetic variation phonetic segments segment release pre-phonation	transcription tool link annotation graph coding links participant database event database	11
12									12
13									13
14									14
15									15

[revised 26 September 2010]

Appendix 1b

Transcription Delicacy Hierarchy (Intermediate)

[Version 2.3]

Table 1. *Transcription Delicacy Hierarchy (Intermediate)*

		1	2	3	3+
		Foundation	Basic	Intermediate	Intermediate+
1	Core	words speakers intonation units sequence	closure/finality appeal/question pause laughter marginal words	overlap truncation/cut-off breath pause duration lag/lengthening hold/micropause utterance start latching vocalism	manner/quality vox accent non-vocal event co-action gaze gesture body
2					
3					
4					
5					
6					
7					
8					
9					
11	Meta		comments unintelligible uncertain pseudograph header	index recording alignment timestamp	
12					
13					
14					
15					

[revised 26 September 2010]

Appendix 2a

Basic Transcription Symbols (by Level)*(Level 1 to Level 4.5)*

MEANING	SYMBOL	COMMENTS
Level 1: Foundation		
words	word word word	space before & after delimits words (standard orthogr.)
speakers	JILL;	attribution to speaker: semicolon follows name in CAPS
intonation units		each intonation unit is written on its own line
sequence: words	word1 word2 word3	words written in conventional order, left-right (standard)
sequence: IU's		lines arranged vertically on page, top-down (standard)
sequence: turns		speaker change sequence is (approximate) turn sequence
Level 2: Basic		
pause, untimed	...	pause lasting more than 0.2 seconds, approximately
laughter	@	one @ per pulse or particle of laughter
marginal words		<i>uh, um, mm, unh-unh</i> , etc. (quasi-standard orthography)
closure: final intonation	.	intonation signals current action as complete (period)
closure: continuing inton.	,	intonation signals current action as incomplete (comma)
appeal intonation	?	intonation (high rise) signals appeal for response
Metatranscription		
header		general information: recording source, location, etc.
conventions		cite transcription conventions used (e.g. via comment)
comments	((WORDS))	transcriber/analyst comment on any topic
unintelligible words	###	one # symbol per syllable
uncertain transcription	#you're #kidding	transcribed words are uncertain
Level 3: Boundary		
overlap (first set)	[]	simultaneous speech; align left square brackets vertically
overlap (2nd set)	[₂]	align left brackets, co-indexed with subscript numeral
truncated intonation unit	—	aborting projected IU (em dash; <i>or</i> two hyphens)
truncated/cut-off word	wor—	aborting projected word (en dash; <i>or</i> word-final hyphen)

MEANING	SYMBOL	COMMENTS
in-breath	(H)	audible inhalation
exhalation	(Hx)	audible exhalation
pause, timed	(1.2)	pause duration in seconds and tenths of seconds
lag/prosodic lengthening	:	colon marks slowing of local tempo, lengthened segment
hold/micropause	..	< 180 milliseconds; brief silence, break in phonation
utterance start	<i>Capital initial</i>	Capitalize for beginning of new sentence/utterance
latching	=	no pause between speaker turns (used in pairs)
linking (absence of break)	_	reduction/merging of words in rapid speech (underscore)
discontinuous IU	&	written only when another speaker intervenes

Metatranscription

pseudograph	¬Jill	name change (etc.) to preserve anonymity (negation sign)
index/line numbers	1	number each line for referencing
recording alignment		link transcription w/corresponding audio/video (optional)
timestamp	<T=154.762>	indicate time in seconds from start of recording

Level 4: Interaction (4.5)

vocalism	(COUGH)	various notations: (SNIFF), (AHEM), (YAWN), etc.
click	(TSK)	alveolar click
glottal stop, creak	(%)	creak vocalism produced as a separate “word”
manner/quality	<MISC> </MISC>	various notations for manner of speaking
laughing words	@you’re @kidding	preposed laugh symbol marks laughter during each word
laughing words	<@> words </@>	marks laughter over several words (for 1 full IU or more)
vox	<VOX> </VOX>	speaking in the voice of another
affect	<AFFECT> </>	voice indexes emotion/affect (specified or not)
non-vocal event	{KNOCK}	describes action by participant, ambient noise, etc.

NB: For additional Level 4 categories and conventions (i.e. beyond level 4.5), consult the text and/or the Reference List of Transcription Symbols.

[rev. 9-Jul-2010]

Appendix 2b

Basic Transcription Symbols (by Topic)

Transcription Level 3+: Intermediate

MEANING	SYMBOL	COMMENTS
Unit		
word	word word word	space before & after delimits words (standard orthogr.)
intonation unit	<u>LINE</u>	each intonation unit is written on its own line
Participation		
speaker attribution	JILL;	attribution to speaker: semicolon follows name in CAPS
Pause/tempo/sequence		
pause, timed	(1.2)	pause duration in seconds and tenths of seconds
hold/micropause	..	< 180 milliseconds; brief silence, break in phonation
pause, untimed	...	pause of 0.2 seconds or more (timed pause is preferred)
lag/prosodic lengthening	:	colon marks slowing of local tempo, lengthened segment
latching	=	no pause between speaker turns (used in pairs)
linking (absence of break)	—	reduction/merging of words in rapid speech (underscore)
utterance start	<i>Capital initial</i>	Capitalize for beginning of new sentence/utterance
Intonation		
final intonation	.	intonation signals current action as complete (period)
continuing intonation	,	intonation signals current action as incomplete (comma)
appeal/question intonation	?	intonation (high rise) signals appeal for response
truncated intonation unit	—	aborting projected IU (em dash; OR: two hyphens)
Disfluency		
truncated/cut-off word	wor—	aborting projected word (en dash; OR: single hyphen)
discontinuous IU	&	written only when another speaker intervenes
Overlap		
speech overlap (first pair)	[]	simultaneous speech; align left square brackets vertically
speech overlap (2nd pair)	[₂]	align left brackets, co-indexed with subscript numeral
Vocalisms		
in-breath	(H)	audible inhalation
exhalation	(Hx)	audible exhalation

MEANING	SYMBOL	COMMENTS
laugh	@	one @ per pulse or particle of laughter
laughing words	@you're @kidding	preposed laugh symbol marks laughter during word
vocalism	(COUGH)	various notations: (SNIFF), (AHEM), (YAWN), etc.
click	(TSK)	alveolar click
glottal stop, creak	(%)	creak vocalism produced as a separate "word"
Speech		
marginal words		<i>uh, um, mm, unh-unh</i> , etc. (quasi-standard orthography)
Manner		
manner/quality	<MISC> </MISC>	various notations for manner of speaking
laughing words	<@> words </@>	marks laughter over several words (for 1 full IU or more)
vox	<VOX> </VOX>	speaking in the voice of another
affect	<AFFECT> </>	voice indexes emotion/affect (specified or not)
Action		
non-vocal event	{KNOCK}	describes action by participant, ambient noise, etc.
Metatranscription		
header		general information: recording source, location, etc.
conventions		cite transcription conventions used (e.g. via comment)
comments	((WORDS))	transcriber/analyst comment on any topic
unintelligible words	###	one # symbol per syllable
uncertain transcription	#you're #kidding	transcribed words are uncertain
uncertain transcription	<#> words </#>	marks uncertainty over several words (1 full IU or more)
pseudograph	-Jill	name change (etc) to preserve anonymity (negation sign)
Time		
index/line numbers	1, 2, 3...	number each line for referencing
recording alignment		link transcription w/corresponding audio/video (optional)
timestamp	<TIME=154.762>	indicate time in seconds from start of recording

[rev. 9-Jul-2010]

Appendix 3

Reference List of Discourse Transcription Symbols

“DT2” Version 2.2.1

<u>MEANING</u>	<u>SYMBOL</u>	<u>COMMENTS</u>
General		
Sequence		
1. word sequence		left to right order marks temporal sequence
2. IU sequence		top to bottom order of lines marks IU sequence
3. turn sequence		top to bottom order of speaker labels suggests turns
Pause		
4. pause, timed	(1.2)	pause duration in seconds (preferred)
5. hold/micropause	..	< 180 milliseconds; brief silence, break in phonation
6. pause, untimed	...	180 milliseconds or more (timed is preferred)
Tempo/Flow		
7. latching	=	no pause between speaker turns (used in pairs)
8. lag/prosodic length	:	lengthening of segment, slowing of local tempo
9. truncated/cut-off word	wor–	word not completed as projected (en dash)
Overlap		
10. overlap (first set)	[]	align left square brackets vertically
11. overlap (2nd set)	[₂]	align left bracket and index w/ subscript number
Vocalism (basic)		
12. in-breath	(H)	audible inhalation
13. out-breath	(Hx)	audible exhalation
14. laugh	@	one symbol per pulse of laughter
15. vocalism	(DESCRIPTION)	various notations: (COUGH), (SNIFF), etc.
16. click	(TSK)	alveolar click

	<u>MEANING</u>	<u>SYMBOL</u>	<u>COMMENTS</u>
17.	glottal sound	(%)	creak produced as separate “word”
18.	glottalized manner	%word	creaky, modified production of sounds in a word
<i>Manner (basic)</i>			
19.	laugh voice	@you’re @kidding	preposed laugh symbol marks laughter during word
20.	laugh voice	<@> </@>	laughing while speaking (for seven+ words)**
21.	vox: voice of another	<VOX> </VOX>	alternate notation
22.	vox: voice of another	« »	alternate (double angle quotation mark)
23.	piano, attenuated	<P> </P>	quiet speech
24.	piano, attenuated	°word word°	alternate notation (degree sign)
25.	whisper voice	¤word word¤	alternate notation (currency sign)
<i>Meta-transcription</i>			
26.	comment	((WORDS))	transcriber/analyst perspective inserted in transcript
27.	unintelligible	###	one # per indecipherable syllable
28.	uncertain hearing	#you’re #kidding	marks transcribed words as uncertain
29.	uncertain (long)	<#> words </#>	for seven or more words
30.	alternative hearings	#can_((#can't))	uncertain which word is heard (no space)
31.	alternative hearings (complex)	<#1> words </#1>_ <#2> words </#2>	uncertain which of 2 word sequences is heard (no space)
32.	pseudograph	¬Jill	name change (etc.) for anonymity (negation sign)
33.	speaker's error	<SIC> </SIC>	transcription is accurate, error is speaker's
<i>Time/Index</i>			
34.	index/line numbers	<LINE=1>	each line numbered in sequence
35.	timestamp	<TIME=154.73>	time in seconds from start of recording
36.	time landmark	‡	time-stamp location (alternate)
37.	duration	<DUR=1.2> um </DUR>	duration in seconds, bracketing event
<i>Participation</i>			
38.	speaker/agent attribution	JILL;	semicolon following name in CAPS
39.	multiple speakers	JILL+JIM;	for simultaneous joint action of individuals
40.	group of speakers	AUDIENCE++;	for simultaneous joint action of crowd
41.	role voicing	JILL=JIM;	speaker (1st) speaks in role of another (2nd)
42.	noninteractive source	±COMPUTER;	marks source or agent of sound or event

	<u>MEANING</u>	<u>SYMBOL</u>	<u>COMMENTS</u>
43.	participation frame	JILL/TEL;	slash symbol precedes interactional context
44.	recipient	JILL\DOG;	backslash symbol precedes addressee
45.	unidentified speaker	#X;	cannot determine whose voice
46.	uncertain speaker	#JILL;	unsure whether speaker attribution is correct
47.	alternative speakers	#JILL_((#JOAN));	unsure which of two speakers is speaking
48.	uncertain interactive	#±;	unsure if sound source is a human interactant
49.	prior's pause	△	boundary pause belongs to prior speaker
50.	next's pause	▽	boundary pause belongs to next speaker
51.	open pause	▷	boundary pause not attributed to any speaker
52.	participant gloss	_((μ=GLOSS))	what the speaker means (e.g. for child speech)

Intonation

Note: In the sections which follow, certain categories and symbols remain experimental, and are subject to change (e.g. those for intonation contour, tone, key/baseline, gaze, gesture, co-action, etc.).

Unit Boundary

53.	word	<u>SPACE</u>	space before and after delimits word
54.	intonation unit	<u>LINE</u>	one new line for each Intonation Unit
55.	intonation unit		alternate, without line breaks (double vertical line)
56.	no IU boundary		for editing & correcting transcriptions only
57.	accent group		rhythmic subunit of intonation unit (vertical line)
58.	{unit boundary TBD}		IU-internal unit (broken bar)
59.	{unit boundary TBD}	}	(wavy line)
60.	syllable boundary	.	raised dot (hyphenation point)
61.	reduction/cliticization	'	follows standard orthography (apostrophe)
62.	linking (no break)	—	merging of words in rapid speech (underscore)
63.	rush start	„	anacrusis begins
64.	rush end	^	anacrusis ends (may be omitted if redundant)
65.	attenuated IU boundary	⌒	intonation unit boundary cues largely suppressed
66.	discontinuous IU	&	used only when 2nd speaker intervenes
67.	utterance start	<u>Capital initial</u>	Capitalize for start of new “utterance”
68.	prosodic sentence	§	(section sign)
69.	paratone	¶	(paragraph sign)

	<u>MEANING</u>	<u>SYMBOL</u>	<u>COMMENTS</u>
	<i>Boundary Intonation</i>		
70.	final contour	.	intonation marks current action as complete (period)
71.	continuing contour	,	intonation marks current action as incomplete
72.	appeal contour	?	intonation marks appeal for recipient response
73.	final appeal	?.	combines with following period (option, unmarked)
74.	continuing appeal	?,	combines with following comma
75.	truncated IU	—	aborting projected IU (em dash; or 2 hyphens)
	<i>Accent/Stress</i>		
76.	primary (pitch) accent	^	(circumflex accent)
77.	secondary accent	`	(grave accent)
78.	primary stress	'a	(vertical line, modifier letter)
79.	secondary stress	,a	(low vertical line, modifier letter)
80.	booster	!	extra prominence of movement (pre-posed)
81.	low booster	!	pre-posed (inverted exclamation mark)
82.	beat/stress	*	prominence/accent/beat timing (asterisk)
83.	delayed peak	^*	
84.	early peak	*^	
85.	chant/stylization	÷word÷	stylized contour, e.g. sung level pitches, bracketed
	<i>Tune/Contour</i>		
86.	rise	↗	global rising contour [IPA]
87.	fall	↘	global falling contour [IPA]
88.	level	→	sustained tone
89.	rise-fall	↗↘	unitary rising-falling contour
90.	fall-rise	↘↗	unitary falling-rising contour
91.	rise-fall	↗↘	(alternate notation)
92.	fall-rise	↘↗	(alternate notation)
	<i>Terminal pitch</i>		
93.	top	↑	top of speaker range
94.	high	↑	high in speaker range
95.	level	→	level; movement to mid in speaker range
96.	low	↓	low in speaker range

	<u>MEANING</u>	<u>SYMBOL</u>	<u>COMMENTS</u>
97.	bottom	↓	bottom of speaker range
	<i>ToBlcon</i>		
98.	high	↑	relatively high in speaker range
99.	low	↓	relatively low in speaker range
100.	downstepped high	↑̣	lowered high relative to previous high tone
101.	[upstepped low]	↓̣	raised relative to previous tone
102.	[sustained/level]	→	usually used for boundary tone only
	<i>Tonetic</i>		
103.	rise	á	(acute accent, modifier letter)
104.	fall	à	(grave accent, modifier letter)
105.	level	ā	(macron, modifier letter)
106.	rise-fall	â	(circumflex accent, modifier letter)
107.	fall-rise	ǎ	(caron, modifier letter)
	<i>Tone</i>		
108.	extra-high	↑̥	extra-high in speaker range
109.	high	↑	high in speaker range
110.	mid	↑̤	middle of speaker range
111.	low	↓̤	low in speaker range
112.	extra-low	↓̥	extra-low in speaker range
113.	rising	↑̤̚	(IPA)
114.	falling	↓̤̚	(IPA)
115.	high rising	↑̥̚	(IPA)
116.	low rising	↓̥̚	(IPA)
117.	rising-falling	↑̤̚̚	(IPA)
	<i>Key/Baseline</i>		
118.	appeal start	¿	onset of appeal plateau (inverted question mark)
119.	head end	/	precedes first primary accented syllable of IU
120.	tail start	\	precedes IU-final (trailing) unaccented syllables
121.	up-key (plateau)	↗	onset of incrementally raised baseline
122.	down-key	↘	low top: onset of incrementally lowered topline
123.	mid-key	⇒	reset gridlines (usually redundant)
124.	high-key	↗̥	(alternate) large rise in declination gridlines

	<u>MEANING</u>	<u>SYMBOL</u>	<u>COMMENTS</u>
125.	low-key	⇓	(alternate) large fall in declination gridlines
126.	pitch reset (minor)	┌	small shift in pitch of unstressed syllables
127.	pitch reset (major)	┌─	major pitch reset (cf. high-key)
128.	amplitude reset (minor)	┘	small shift in overall amplitude of intonation unit
129.	amplitude reset (major)	┘	large shift in overall amplitude of intonation unit
130.	reset up	^	combines with pitch or amplitude reset (optional)
131.	reset down	∨	combines with pitch or amplitude reset (optional)

Resonance

132.	isotony start	≈	resonance between intonation contours of IU's
133.	isotony end	≠	intonational resonance ends
134.	isochrony start	≡	beginning of steady tempo of accent beats
135.	isochrony end	≡	end of steady tempo of accent beats

Vocalism

136.	cough	(COUGH)
137.	throat-clear	(AHM)
138.	gulp, swallow	(GULP)
139.	yawn	(YAWN)
140.	sigh, expressive exhale	(SIGH)
141.	gasp, expressive inhale	(GASP)
142.	blowing air	(BLOW)
143.	creak	(%)
144.	creak (alternate)	(CREAK)
145.	groan, moan	(GROAN)
146.	choke	(CHOK)
147.	sob	(SOB)
148.	grunt	(GRUNT)
149.	burp	(BURP)
150.	belch, loud burp	(BELCH)
151.	alveolar click	(TSK)
152.	lateral click	(TCHK)
153.	kiss	(KISS)
154.	sniff	(SNIFF)

	<u>MEANING</u>	<u>SYMBOL</u>	<u>COMMENTS</u>
155.	snort	(SNORT)	
156.	sneeze	(SNEEZE)	
157.	snore	(SNORE)	
158.	snuffle, nose-blowing	(NOSE-BLOW)	
159.	snuffle, nose-blowing	(SNUFFLE)	
160.	scream	(SCREAM)	
161.	yell	(YELL)	
162.	howl	(HOWL)	
163.	growl	(GROWL)	
164.	cry	(CRY)	
165.	whistle	(WHISTLE)	
166.	pow	(POW)	onomatopoetic
167.	bronx cheer	(BBBB)	onomatopoetic
168.	buzz	(BUZZ)	onomatopoetic
169.	laugh with vowel quality	@ ^a @ ^e @ ⁱ @ ^o @ ^u	superscript lower-case vowel indicates quality
170.	laugh, voiceless release	@ ^h	superscript lower-case h
171.	laugh, nasal release	@ ⁿ	superscript lower-case n
172.	laugh-gasp	(H) [@]	in-breath when out of breath from laughing
173.	laugh-gasp (altern.)	(GASP) [@]	in-breath when out of breath from laughing
174.	nasal in-breath	(H) ⁿ	in-breath with mouth closed
175.	breath-hold	(<u> </u>)	silence between in-breath and speech

Manner

Voice

176.	laugh voice	<@> </@>	
177.	smile voice	£you're kidding£	marks smiling during word
178.	smile voice	<SMILE> </SMILE>	marks smiling during word (alternate)
179.	smile voice	☺ you're ☺ kidding	marks smiling during word (alternate)
180.	smile voice	<☺> </☺>	marks smiling during word (for seven+ words)
181.	voice of another	<VOX> </VOX>	speaking in the voice of another
182.	voice of another	« »	(alternate notation)
183.	voice of another	<VOX=DUMB> </>	specify voice quality, e.g. stereotype voice
184.	quotation quality	<QUOTE> </>	only when voice quality indexes quote (rare)

	<u>MEANING</u>	<u>SYMBOL</u>	<u>COMMENTS</u>
185.	onomatopeia	<ONOM> </>	word or sound is onomatopoetic
186.	onomatopeia	<ONOM=DOG> </>	specifies what the onomatopoetic sound represents
187.	whisper	<α> </>	alternate
188.	whisper	<WHISPER> </>	alternate
189.	breathy	<BREATHY> </>	
190.	creaky	<%> </>	
191.	yawning	<YAWN> </>	
192.	reading	<READ> </>	
193.	singing	<SING> </>	
194.	shouting	<SHOUT> </>	
195.	crying	<CRY> </>	
196.	humming	<HUM> </>	
197.	slurred speech	<SLUR> </>	
<i>Prosody</i>			
198.	forte, loud	<F> </>	
199.	piano, attenuated	<P> </>	
200.	piano, attenuated	<°> </>	(alternate notation)
201.	high pitch (region)	<HI> </>	
202.	low pitch (region)	<LO> </>	
203.	wide pitch	<WIDE> </>	
204.	narrow pitch	<NARROW> </>	
205.	marcato speech	<MARCATO> </>	
206.	fast speech	<ALLEGRO> </>	
207.	slow speech	<LENTO> </>	
208.	increasingly fast	<ACCELERANDO>	speech gradually speeds up
209.	increasingly slow	<RALENTANDO>	speech gradually slows down
210.	parenthetical intonation	<PAREN> </>	
<i>Affect</i>			
211.	affect	<AFFECT> </>	for overt bodily expression of marked emotion
212.	affect, specified	<AFFECT=JOY> </>	sadness, anger, disgust, fear, happiness, etc.
<i>Other manner</i>			
213.	talk w/food in mouth	<FOOD> </FOOD>	speech is modified by food in mouth

	<u>MEANING</u>	<u>SYMBOL</u>	<u>COMMENTS</u>
214.	talk while smoking	<SMOKE> </>	speech is modified by smoking while talking
215.	talking on telephone	<TEL> </TEL>	(alternate notation)
216.	media speech	<MEDIA> </>	(alternate notation)

Group Talk

NB: This set of categories are neither vocalisms, non-vocal event, nor lexically transcribed speech. They fall into an intermediate category, here represented by the double-parenthesis notation used for ad hoc comments.

217.	murmur, extended	((MURMUR))	for mass talk by large audiences, etc. (group)
218.	laughter, extended	((LAUGHTER))	involves multiple laugh pulses, usually many
219.	cheering, extended	((CHEER))	for mass actions of large audiences, etc. (group)
220.	talk, extended	((TALK))	for background talk that is not transcribed
221.	background talk start	((TALK))¢	talk continues in background
222.	background talk end	¥((TALK))	talk in background ends

Non-Vocal Event

223.	non-vocal event	{EVENT}	general pattern for acts, events (open-ended)
224.	closes door	{CLOSE-DOOR}	action done by agent, or door moves by itself
225.	opens door	{OPEN-DOOR}	action done by agent, or door moves by itself
226.	enters room	{ENTER}	
227.	leaves room	{EXIT}	
228.	knocks on door	{KNOCK}	
229.	beat	{BEAT}	pound, tap, thump or beat table, etc.
230.	snap fingers	{SNAP}	
231.	slap	{SLAP}	
232.	clap hands	{CLAP}	distinct punctual acts (e.g. by one individual)
233.	applaud	{APPLAUSE}	continuous mass activity (e.g. by crowd)
234.	eat	{EAT}	
235.	drink	{DRINK}	
236.	smoke	{SMOKE}	
237.	fall	{FALL}	e.g. fall on floor
238.	write	{WRITE}	
239.	erase	{ERASE}	
240.	running water	{RUN-WATER}	
241.	washing dishes	{WASH-DISHES}	

	<u>MEANING</u>	<u>SYMBOL</u>	<u>COMMENTS</u>
242.	machine noise	{MACHINE}	
243.	rumbling noise	{RUMBLE}	
244.	thunder	{THUNDER}	
245.	rain	{RAIN}	
246.	hail	{HAIL}	
247.	telephone rings	{RING}	
248.	computer beeps	{BEEP}	
249.	television plays	{TELEVISION}	
250.	media plays	{MEDIA}	
251.	microphone noise	{MIC}	
252.	move microphone	{MOVE-MIC}	
253.	move papers	{MOVE-PAPER}	
254.	footsteps	{FOOTSTEPS}	
255.	dog barks	{BARK}	
256.	cat mews	{MEW}	
257.	continuous event start	{MUSIC}ϕ	non-vocal event continues in background
258.	continuous event end	¥{MUSIC}	non-vocal event in background ends
Co-Action			
259.	co-action (general)	©	action while speaking
260.	co-action: beat	⊗	strike surface (e.g. pound or tap hand on table)
261.	co-action: snap	Ⓢ	snap fingers
262.	co-action: gesture	ⓖ	make gesture stroke while speaking
263.	co-action, ad hoc	<©=SLAP> w©ord </©>	= describes what action, © specifies where
Gaze			
264.	gaze approach start	<	experimental notation
265.	gaze depart end	>	experimental notation
266.	mutual gaze start	≲	experimental notation
267.	mutual gaze end	≳	experimental notation
268.	no mutual gaze start	⋈	experimental notation
269.	no mutual gaze end	⋈	experimental notation

<u>MEANING</u>	<u>SYMBOL</u>	<u>COMMENTS</u>
Gesture		
270. gesture stroke start	≡	experimental notation
271. gesture stroke end	≡	experimental notation
272. gesture approach start	≡	experimental notation
273. gesture depart end	≡	experimental notation
274. no gesture	≡	experimental notation
275. no gesture	≡	experimental notation
Phonetics		
276. heavy aspirated release	but_h	suffixed, for strongly aspirated stop consonant release; not = exhalation
277. pre-phonation	m_word	prefixed single consonant (rare), e.g. m_, n_, p_
278. glottal stop, prosodic	%o	option for high delicacy transcription (<i>cuto%-</i>)
279. glottal stop, phonemic	?	for languages with distinctive/phonemic glottal stop
280. long phoneme	a'	lexically long (= IPA 'half-length' sign)
281. elision	wo<r>d	for e.g. sociolinguistic phonetic variation
282. phonetic transcription	[]	IPA transcription (follows orthographic)
283. unidentified phonetic segment	○	for transcriber uncertainty in IPA transcription
Other		
284. code-switch	<L2> words </L2>	switch to a different language from main one
285. code-switch	<L2=DUTCH> </L2>	(alternate notation, language specified)
286. code-switch	<LANGUAGE=DUTCH>	(alternate notation, language specified)
287. false start/ reparandum	×word	rarely marked (multiplication sign)
Tiers		
288. tier (open-ended)	\$tier;	line with categories linked to utterance above
289. gaze tier	\$gaze;	gaze events
290. face tier	\$face;	face events
291. gesture tier	\$gesture;	gesture events
292. body tier	\$body;	body events
293. act tier	\$act;	actions
294. voice tier	\$voice;	voice quality/manner

	<u>MEANING</u>	<u>SYMBOL</u>	<u>COMMENTS</u>
295.	intonation tier	\$intonation;	intonation contours/tones
296.	prosodic tier	\$prosody;	other prosodic features
297.	phonetic tier	\$phonetic;	segmental phonetic transcription
298.	variation tier	\$variation;	sociolinguistic variation
299.	morphological tier	\$morph;	morphological analysis
300.	gloss tier	\$gloss;	morpheme or word gloss
301.	syntax tier	\$syntax;	syntactic analysis
302.	free translation tier	\$free;	free translation
303.	comment tier	\$comment;	comments (open-ended)
304.	tier event type 1 start	┐	link tier to speech
305.	tier event type 1 end	┑	link tier to speech
306.	tier event type 2 start	└	link tier to speech
307.	tier event type 2 end	┘	link tier to speech
Grammar			
308.	morpheme boundary	-	marks morphemes & orthography (hyphen)
309.	clitic boundary	+	used for grammatical coding of morphology
310.	syntactic phrase	()	used for syntactic coding of constituent structure
311.	root (grammar)	√	for grammatical analysis only
Publishing			
312.	constructed example	~	for publishing etc., not transcription per se (tilde)
313.	omission of lines	⋮	for excerpt presentation, not transcription per se
Technical & Other			
314.	audio filter start	⌘	filtered audio used to make names unintelligible
315.	audio filter end	⌞	filtered audio used to make names unintelligible
316.	real name	⌘Jill	name that does not have to be pseudonymized
317.	interesting item	⌘	researcher's note for item of interest
318.	interesting item (alt.)	⌘*	researcher's note for item of interest

Comments

The symbol conventions presented in this reference list are those used in the transcription practice known as Discourse Transcription 2 (DT2), as updated to take account of recent developments, e.g. in Unicode, intonation transcription conventions such as INTSINT (Hirst and Di Cristo, 1998), and recent developments in Conversation Analysis (Jefferson, 2004). These symbols update and supercede the earlier conventions of the original Discourse Transcription (DT1) system (Du Bois 1991, Du Bois et al. 1992, Du Bois et al. 1993).

Symbols for representing aspects of the body other than the voice are minimally represented here. For further information on the relevant transcription issues, please consult the work of the relevant authorities (e.g. Pittenger, Hockett & Danehy, Kendon, Ekman, Goodwin, McNeill, Streeck, and others).

For details about the special characters used here, including Unicode standard character names, Unicode numbers, and so on, see Appendix A.6 “Unicode Characters for Discourse Transcription.”

Selected References

- Du Bois, John W. 1991. Transcription design principles for spoken discourse research. *Pragmatics* 1:71-106.
- Du Bois, John W., Cumming, Susanna, Schuetze-Coburn, Stephan, and Paolino, Danae. 1992. Discourse transcription. *Santa Barbara Papers in Linguistics* 4:1-225.
- Du Bois, John W., Schuetze-Coburn, Stephan, Cumming, Susanna, and Paolino, Danae. 1993. Outline of discourse transcription. In *Talking data: Transcription and coding in discourse research*, eds. Jane A. Edwards and Martin D. Lampert, 45-89. Hillsdale, NJ: Erlbaum.
- Hirst, Daniel, and Di Cristo, Albert eds. 1998. *Intonation systems: A survey of twenty languages*. Cambridge: Cambridge University Press.
- Jefferson, Gail. 2004. Glossary of transcript symbols with an introduction. In *Conversation Analysis: Studies from the first generation*, ed. Gene H. Lerner, 13-31. Amsterdam: Benjamins.
- Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press.
- Pittenger, Robert E., Hockett, Charles F., and Danehy, John J. 1960. *The first five minutes: A sample of microscopic interview analysis*. Ithaca: Paul Martineau.

[rev. 30-July-2009]

Appendix 4

Comparison of Transcription Symbols

MEANING	DT1	DT2	CA
Unit			
40. word	<u>SPACE</u>	<u>SPACE</u>	<u>SPACE</u>
41. intonation unit	<u>LINE</u>	<u>LINE</u>	
Pause			
42. pause, timed	...(1.2)	(1.2)	(1.2)
43. pause, short	(.)
44. pause, long (untimed)	
45. latching	(0)	=	=
46. lag (prosodic length)	=	:	:
Sequence			
47. overlap (1st set)	[]	[]	[]
48. overlap (2nd set)	[2 2]	[₂]	[]
Disfluency			
49. truncated/cut-off word	wor-	wor-	wor-
Vocalism			
50. inhale	(H)	(H)	.hhh
51. exhale	(Hx)	(Hx)	hhh
52. vocalism	(SNIFF)	(SNIFF)	(sniffle)
53. click	(TSK)	(TSK)	.t, t!
54. laugh pulse	@	@	heh
55. laughing word	wo@rd	wo@rd	wo(h)rd
56. glottalized/creaky vocalism	%	(%)	
57. glottalized sound	wo%rd	wo%rd	woghrd
Manner/Quality			
58. manner/quality	<MISC MISC>	<MISC> </MISC>	((description))
59. vox: voice of another	<Q Q>	<VOX> </VOX>	
60. piano, attenuated speech	<P P>	°words°	°words°
61. smile quality	<SM SM>	<☺ > </☺ >	£words£
Metatranscription			
62. unintelligible words	XXX	###	()
63. uncertain transcription	<X word X>	#word	(word)
64. comment	((WORDS))	((WORDS))	((words))
65. pseudograph	~Jill	~Jill	
Participation			
66. speaker/turn attribution	JILL:	JILL;	Jill:
67. unidentified speaker	X:	#;	():
68. uncertain speaker		#JILL;	(Jill):

Boundary tone/closure

69. final intonation	.	.	.
70. continuing intonation	,	,	,
71. truncated intonation unit	--	—	
72. appeal/question (final)	?	?.	?
73. appeal/question (continuing)	?	?,	?

Prosody/intonation

74. primary accent	^	^	<u>word</u>
75. secondary accent	`	`	<u>word</u>
76. forte	<F word F>	<F> word </F>	WORD
77. rising intonation contour	/	↗	
78. falling intonation contour	\	↘	
79. level intonation	—	→	
80. high pitch (top)		↑	↑
81. high pitch		↑	<u>word</u>
82. low pitch		↓	
83. low pitch (bottom)		↓	↓

Commentary

This Table presents a comparison between three sets of conventions for symbols used to represent basic transcription categories. The three sets are:

- ♦ **DT1** *Discourse Transcription* conventions as in the version of
Du Bois (1991), Du Bois et al. (1992), and Du Bois et al. (1993)
- ♦ **DT2** *Discourse Transcription* conventions as currently revised
- ♦ **CA** *Conversation Analysis* conventions as currently revised (Jefferson 2004)

A careful comparison of these transcription systems reveals both similarities and differences. A meaningful evaluation of such similarities and differences must penetrate more than skin deep, however. The comparison must move beyond mere surface equivalences or differences in symbols to consider the symbol's function and meaning in terms of transcription categories and transcription practices. Consider the Discourse Transcription systems, on the hand, as compared and contrasted with Conversation Analysis on the other. Here we find that a difference in symbols in some cases disguises an underlying similarity in the transcription categories, or in de facto transcriber practice. Conversely, the use of the same or a similar symbol does not guarantee an identical transcription category, nor equivalent analytical practice. Even similar names or glosses for the transcription categories in question should be taken only as a rough guide to the actual analytical practice. Where transcription practices differ, the effective meaning of the categories and symbols used is likely to differ as well. Thus, any full comparison of these transcription systems must ultimately make reference to a detailed consideration of the actual transcription practices as described in the relevant literature for each system, and as actually carried out in the body of published research.

Still, it can be stated that in many cases, the similarity suggested in a given row of compared symbols is real, even if the notations may differ. In the end, the choice of symbol matters less than the conceptualization of the category being represented, and the analytical practice it indexes.

Some observations on the comparison of individual transcription notations, categories, and practices is

given below.

Speaker label. DT2 places a semi-colon after the speaker label, where DT1 and CA write a colon. While the colon is admittedly more aesthetic, it can be confused with another important function, that of marking prosodic lengthening or “lag”. The use of semicolon in DT2 is designed to avoid this confusion. This is especially important for any computer processing of transcription files, as the boundary between the “text” and “pre-text” information in any transcribed line is demarcated precisely by the speaker label. There are few symbols which need to be as reliably distinguished from all others as this particular division of transcription information.

Truncated/Cut-Off Word. DT2 uses the en dash, where DT1 and CA use a simple hyphen. The reason for preferring en dash in DT2 is to avoid confusion with the use of hyphen in compound (hyphenated) words and other functions (e.g. morpheme boundaries, if these are marked). However, the plain hyphen can be used in DT2 as an interim measure, for convenience.

Truncated Intonation Unit. DT1 uses two hyphens, while DT2 uses an em dash. The use of em dash in DT2 is designed to distinguish intonation unit truncation from word truncation, as well as from compound (hyphenated) words and morpheme boundaries. However, it is possible to use two hyphens in DT2 as an interim measure, for convenience.

Piano. DT2 provides two alternate conventions for piano, attenuated, or quiet speech. The standard DT2 notation for manner can be used, in which the words spoken quietly are enclosed within angle brackets marked with a “P”: <P> words </P>. Alternatively, the Conversation Analysis convention of enclosing the words within degree signs can be used: °words°.

Pause. Standard practice in DT2 is to measure all pauses to the tenth of a second (except micropauses, i.e. pauses of less than 180 milliseconds in duration). Thus, the symbols for (unmeasured) medium and long pauses are to be used only in the broadest levels of transcription delicacy.

Pseudograph. This term refers to a notation in which a person’s name (or other sensitive indicator of identity, such as an address or cell phone number) has been changed to preserve anonymity. It is not used for speaker labels, but only where the word was actually uttered.

References

NB: For recent updates on transcription, see <http://www.linguistics.ucsb.edu/projects/transcription/>

- Du Bois, John W. 1991. Transcription design principles for spoken discourse research. *Pragmatics* 1:71-106.
- Du Bois, John W., Cumming, Susanna, Schuetze-Coburn, Stephan, and Paolino, Danae. 1992. Discourse transcription. *Santa Barbara Papers in Linguistics* 4:1-225.
- Du Bois, John W., Schuetze-Coburn, Stephan, Cumming, Susanna, and Paolino, Danae. 1993. Outline of discourse transcription. In *Talking data: Transcription and coding in discourse research*, eds. Jane A. Edwards and Martin D. Lampert, 45-89. Hillsdale, NJ: Erlbaum.
- Jefferson, Gail. 2004. Glossary of transcript symbols with an introduction. In *Conversation Analysis: Studies from the first generation*, ed. Gene H. Lerner, 13-31. Amsterdam: Benjamins.

[rev. 25-Sep-2009]

Appendix 5

Transcription Convention Updates (DT1 to DT2)

Introduction

The following are updated versions of symbol conventions used in the discourse transcription approach of the “Santa Barbara School,” sometimes referred to as the “Discourse Transcription” (DT) conventions (Du Bois et al., 1993). The new conventions supercede those described in earlier publications, and are recommended for use in their place. Some of these changes have already been implemented in the published edition of the Santa Barbara Corpus of Spoken American English (Du Bois et al. 2000-2005), while others will be implemented in future editions. All conventions which are not mentioned here remain as previously described in the cited publications.

Some of these revised conventions represent efforts to make the transcriptions simpler to the eye. For example, devoting a special symbol ‘#’ to the high-frequency notations for unintelligible words and uncertain hearings allows indecipherable words to be directly prefixed using a single-character notation, which eliminates the need for angle bracket notations in many cases. Similarly for laughing while speaking.

Other revised conventions are for greater consistency and effectiveness when using the computer as a research tool. For example, discourse features of manner of speaking, which have long or variable scope (potentially extending over several words), are indicated using angle brackets as before, but now in a style more closely corresponding to the conventions of XML (eXtensible Markup Language).

Comments

Some observations on specific features follow.

Overlap

Note that in the new DT2 convention for marking overlapping speech, there is no space between the brackets (or the bracket index number) and the words they enclose. Only the left overlap bracket needs to be indexed numerically, unless there is a chance for confusion, in which case both brackets should be numbered.

Utterance Start

The category of “utterance start” has a somewhat marginal standing as a transcription category, since it does not always correspond to an easily identifiable and constant audible phenomenon. At the least, however, it may be considered an aid to the reader. It helps make clear where a new utterance has been begun by a speaker (often corresponding to a new start, or a re-start, of a sentence). The category of utterance start is not to be used for imposing a normative notion of the sentence, borrowed from written language conventions, onto the conversational transcript. There is no requirement, for example, that initial capitalization should co-occur with a preceding period. There will be many false starts, hesitation words, etc. which are capitalized according to this transcription practice, in contrast to a normative sentence-based writing style, which would simply edit such features out of the transcript.

Extra-Long Intonation Units

Most intonation units fit easily on a single line of the paper. However, for extra-long IUs that contain more words than can fit on one line (due to limitations of page width, large font size, etc.), it may be felt that something should be done about the word-wrapping. In DT1, there was a special convention, but in DT2, the current practice is simply to do nothing. That is, insert no special symbol, nor any indenting, tabbing, or spacing. The result, in a standard word-processing program, will be that the words at the end of the extra-long line will automatically wrap around to the beginning of a second line on the screen.

Admittedly, the result is not as iconic or visually attractive as when one line on the page is wide enough to fit the whole (short) IU within it. But when an IU is simply too long to fit within the width of a single line, automatic wrapping turns out to be an acceptable solution. When the long IU is displayed in printed form or on a computer screen, the difference between an extra-long IU and a normal one-line IU becomes fairly easy to see, because when the program automatically wraps the long IU's final words onto the second line, the wrapped words will appear at the left margin, where normally only speaker labels appear. The advantage of this convention of doing nothing special is that it allows the transcriber to change margins, font size, and other aspects of display without creating formatting problems, confusion, or extra work. Moreover, it helps ensure that the transcription data remain consistent when they are accessed using different kinds of software, such as databases, concordance, or media-alignment programs. (For some kinds of paper-published presentations, where such wrapping of words may be stylistically unsuitable or proscribed, an alternate solution is needed. The first choice is to use a smaller font, or landscape mode, or both, in order to eliminate the word-wrapping. As a dispreferred alternative, in some cases transcribers may choose to break the extra-long line into two lines using a hard carriage return, and then use the "unit continued" symbol (&) to show that the two lines belong together in one intonation unit. This fix may be called for if interlinear glosses must be included. But it is still preferable for the database form of the transcription to be maintained without such line breaks.)

Quotation vs. Vox

For utterances which were formerly marked with the notation <Q Q>, one should in each instance listen to the recording in order to determine if vox (speaking in the voice of another) is truly justified. If so one should use the <VOX> notation, otherwise use nothing.

Discourse Transcription 2: Convention Updates

<u>Meaning</u>	<u>DT1 (Old)</u>	<u>DT2 (New)</u>
28. Unintelligible (syllables)	X	#
29. Uncertain hearing (words)	<X you're kidding X>	#you're #kidding
30. Pseudograph (fake name, address etc.)	~Jill	¬Jill
31. Real name, address, etc.	#Jill	Jill
32. Long-scope features (various)	<A two words A>	<A> two words
33. Laughter during speech (1-5 words)	<@ two words @>	@two @words
34. Laughter during speech (6+ words)	<@ six words @>	<@> six words </@>
35. Overlap, 3rd instance	[3 word word word 3]	[₃ word word word]
36. Overlap, 2nd instance	[[word word word]]	[₂ word word word]
37. Vox: voice of another	<Q words Q>	<VOX> words </VOX>
38. Word truncation/cut-off with no glottal	wor- (<i>hyphen</i>)	wor- (<i>en dash</i>)
39. Word truncation/cut-off with glottal	wor- (<i>hyphen</i>)	wor%- (<i>en dash</i>)
40. Intonation unit truncation	-- (<i>two hyphens</i>)	— (<i>em dash</i>)
41. Morpheme boundary		- (<i>hyphen</i>)
42. Extra-long IU	(<i>indent</i>)	(<i>word wrap</i>)
43. Pause, timed	...(1.2)	(1.2)
44. Pause, short (< 180 milliseconds)
45. Pause, untimed (> 180 milliseconds)
46. Pause location (if at IU boundary)	(<i>line-initial</i>)	(<i>on a separate line by itself</i>)
47. Latching	(0)	=
48. Speaker label	J:	JILL;
49. Utterance start		(<i>Capital letter</i>)
50. Reset	(<i>Capital letter</i>)	†
51. Repair/editable	<word>	×word

COMPUTER-ORIENTED SPECIALIZED CONVENTIONS:

<u>Meaning</u>	<u>DT1 (Old)</u>	<u>DT2 (New)</u>
52. Timestamp (start time)	00:01:06:21	<TIME=66.21>
53. Timestamp (start and end time)	00:01:06:21 00:01:07:39	<TIME=66.21_67.39>
54. Duration (of event, e.g. long laughter)	@(9.3)	@_<DUR=9.3>

References

NB: For recent updates on transcription, see <http://www.linguistics.ucsb.edu/projects/transcription/>

Du Bois, John W., Schuetze-Coburn, Stephan, Cumming, Susanna, and Paolino, Danae. 1993. Outline of discourse transcription. In *Talking data: Transcription and coding in discourse research*, eds. Jane A. Edwards and Martin D. Lampert, 45-89. Hillsdale, NJ: Erlbaum.

Du Bois, John W. ed. 2000. *Santa Barbara corpus of spoken American English, Part 1*. Philadelphia: Linguistic Data Consortium.

Du Bois, John W. ed. 2003. *Santa Barbara corpus of spoken American English, Part 2*. Philadelphia: Linguistic Data Consortium.

[rev. 25-Sep-2009]

Appendix 6

Unicode Characters for Discourse Transcription (Selected)

For transcribing spoken discourse at higher levels of delicacy in the Discourse Transcription system (DT2), it is preferable to use a Unicode font, such as Charis SIL, Doulos SIL, Gentium, or Arial Unicode MS. Unicode is an explicit system for representing characters which can help to avoid confusions and incompatibilities in using special characters (e.g. in sending or transferring transcription data from one user or computer format or platform to another). It is especially important for transcribers when they want to pursue high-delicacy transcription using notations for prosody, gesture, and especially detailed and iconic intonation notations, like those made available in DT2.

Tables 1 and 2 below present Unicode information for various characters which are either used in or are candidates for use in discourse transcription. Inclusion in these tables does not guarantee inclusion in the Discourse Transcription standard, but merely provides basic Unicode information for relevant symbols that may prove useful in a transcription system of this kind. The Unicode numbers are given in hexadecimal notation. In addition, for characters which are included in the ANSI or ASCII standard, the decimal equivalent is given. For characters which are used with the same or similar meaning in the International Phonetic Alphabet (IPA), the standard IPA number is given.

Note that many of the symbols listed here are not present in a standard computer keyboard layout. There are certain methods of typing such characters which, though somewhat cumbersome, are likely to be easily available to the transcriber. For example, one can access specialized Unicode characters in Microsoft Word by choosing the “Insert” function from the menu, then “Symbol”, and then select the character needed. Or, type the 4-character Unicode number directly into the transcript, and then type Alt-x (i.e. while holding down the Alt key, type the letter x). In OpenOffice.org, select “Insert/Special Character” from the menu, and select the relevant character. Any transcriber who intends to make regular use of the non-ANSI characters in their transcribing practice will likely find it worthwhile to seek a more convenient method. One way is to take advantage of software designed for creating specialized computer keyboard layouts for data entry (e.g. Tavultesoft Keyman). Another is to find a keyboard layout that has already been developed by someone else for this purpose.

Table 1 presents characters which create no special difficulties, since they are mostly in the ASCII and/or ANSI standards, widely available in most computers (though with some variations in the higher-numbered characters); still it may be useful to have the Unicode details ready to hand. Table 2 presents symbols which depend far more on the Unicode standard, and are not likely to be sufficiently stable without it.

Table 1. Standard ASCII & ANSI characters

Sign	Function (in DT2)	Unicode Name	Unicode # (hex)	ASCII I # (dec)	Unicode Subset	IPA #
.	final intonation	FULL STOP	U+002E	46	Basic Latin	1.
,	continuing intonation	COMMA	U+002C	44	Basic Latin	2.
:	lag/lengthening	COLON	U+003A	58	Basic Latin	3.
;	speaker/agent attribution	SEMI-COLON	U+003B	59	Basic Latin	4.
&	discontinuous IU	AMPERSAND	U+0026	38	Basic Latin	5.
=	latching	EQUALS SIGN	U+003D	61	Basic Latin	6.
@	laugh pulse	COMMERCIAL AT	U+0040	64	Basic Latin	7.
#	unintelligible	NUMBER SIGN	U+0023	35	Basic Latin	8.
\$	tier label	DOLLAR SIGN	U+0024	36	Basic Latin	9.
%	glottal	PERCENT SIGN	U+0025	37	Basic Latin	10.
^	primary accent	CIRCUMFLEX ACCENT	U+005E	94	Basic Latin	597 11.
`	secondary accent	GRAVE ACCENT	U+0060	96	Basic Latin	515 12.
'	contraction, orthographic	APOSTROPHE	U+0027	39	Basic Latin	13.
/	head end	SOLIDUS	U+002F	47	Basic Latin	903 14.
\	tail start	REVERSE SOLIDUS	U+005C	92	Basic Latin	15.
<	markup label start	LESS-THAN SIGN	U+003C	60	Basic Latin	16.
>	markup label end	GREATER-THAN SIGN	U+003E	62	Basic Latin	17.
(vocalism start	LEFT PARENTHESIS	U+0028	40	Basic Latin	18.
)	vocalism end	RIGHT PARENTHESIS	U+0029	41	Basic Latin	19.
[overlap start	LEFT SQUARE BRACKET	U+005B	91	Basic Latin	20.
]	overlap end	RIGHT SQUARE BRACKET	U+005D	93	Basic Latin	21.
{	non-vocal event start	LEFT CURLY BRACKET	U+007B	123	Basic Latin	22.
}	non-vocal event end	RIGHT CURLY BRACKET	U+007D	125	Basic Latin	23.
«	vox start (alternate)	LEFT-POINTING DOUBLE ANGLE QUOTATION MARK	U+00AB	171	Latin-1	24.
»	vox end (alternate)	RIGHT-POINTING DOUBLE ANGLE QUOTATION MARK	U+00BB	187	Latin-1	25.
~	constructed example	TILDE	U+007E	126	Basic Latin	26.

Sign	Function (in DT2)	Unicode Name	Unicode # (hex)	ASCII I # (dec)	Unicode Subset	IPA #
!	booster	EXCLAMATION MARK	U+0021	33	Basic Latin	27.
¡	low booster	INVERTED EXCLAMATION MARK	U+00A1	161	Latin-1	28.
?	appeal intonation	QUESTION MARK	U+003F	63	Basic Latin	29.
¿	appeal onset	INVERTED QUESTION MARK	U+00BF	191	Latin-1	30.
×	false start, reparandum	MULTIPLICATION SIGN	U+00D7	215	Latin-1	31.
¬	pseudograph	NOT SIGN	U+00AC	172	Latin-1	32.
	accent unit (tone group)	VERTICAL LINE	U+007C	124	Basic Latin	507 33.
⋮	IU-internal unit (TBD)	BROKEN BAR	U+00A6	166	Latin-1	34.
‡	IU-internal unit (TBD)	DOUBLE DAGGER	U+2021	135	General Punctuation	35.
†	unit boundary (TBD)	DAGGER	U+2020	134	General Punctuation	36.
§	prosodic sentence	SECTION SIGN	U+00A7	167	Latin-1	37.
¶	paratone	PILCROW SIGN [= PARAGRAPH SIGN]	U+00B6	182	Latin-1	38.
÷	chant/stylization	DIVISION SIGN	U+00F7	247	Latin-1	39.
©	co-action (general)	COPYRIGHT SIGN	U+00A9	169	Latin-1	40.
°	piano	DEGREE SIGN	U+00B0	176	Latin-1	41.
¤	whisper	CURRENCY SIGN	U+00A4	164	Latin-1	42.
£	smile voice (alternate)	POUND SIGN	U+00A3	163	Latin-1	43.
¢	continuous event starts	CENT SIGN	U+00A2	162	Latin-1	44.
¥	continuous event ends	YEN SIGN	U+00A5	165	Latin-1	45.
μ	participant gloss	MICRO SIGN	U+00B5	181	Latin-1	46.
*	stress/prominence/beat	ASTERISK	U+002A	42	Basic Latin	612 47.
+	clitic boundary	PLUS SIGN	U+002B	43	Basic Latin	48.
±	non-interactive source	PLUS-MINUS SIGN	U+00B1	177	Latin-1	49.
_	linking [cf. IPA 509]	LOW LINE	U+005F	95	Basic Latin	50.
-	morpheme boundary	HYPHEN-MINUS	U+002D	45	Basic Latin	51.
—	word truncation	EN DASH	U+2013	150	General Punctuation	52.
—	intonation unit	EM DASH	U+2014	151	General Punctuation	53.
...	truncation					
...	pause, untimed	HORIZONTAL ELLIPSIS	U+2026	133	General Punctuation	54.

Table 2. Non-ANSI characters

Sign	Function (DT2)	Unicode Name	Unicode Number (hex)	Unicode Subset	IPA #	
	intonation unit	DOUBLE VERTICAL LINE	U+2016	General Punctuation	508	55.
‖	no intonation unit boundary	NOT PARALLEL TO	U+2226	Mathematical Operators		56.
~	{unit boundary} (TBD)	WAVY LINE	U+2037	Miscellaneous Technical		57.
[phonetic transcription start	LEFT SQUARE BRACKET WITH QUILL	U+2045	General Punctuation		58.
]	phonetic transcription end	RIGHT SQUARE BRACKET WITH QUILL	U+2046	General Punctuation		59.
<	elision start	SINGLE LEFT-POINTING ANGLE QUOTATION MARK	U+2039	General Punctuation		60.
>	elision end	SINGLE RIGHT-POINTING ANGLE QUOTATION MARK	U+203A	General Punctuation		61.
..	hold, micro-pause	TWO DOT LEADER	U+2025	General Punctuation		62.
·	syllable boundary	HYPHENATION POINT	U+2027	General Punctuation		63.
ˈ	primary stress	MODIFIER LETTER VERTICAL LINE	U+02C8	Spacing Modifier Letters	501	64.
ˌ	secondary stress	MODIFIER LETTER LOW VERTICAL LINE	U+02CC	Spacing Modifier Letters	502	65.
aː	phonemic long [IPA=half-long]	MODIFIER LETTER HALF TRIANGULAR COLON	U+02D1	Spacing Modifier Letters	504	66.
a:	long (alternate) [IPA = long]	MODIFIER LETTER TRIANGULAR COLON	U+02D0	Spacing Modifier Letters	503	67.
͡	attenuated IU boundary	UNDERTIE	U+203F	General Punctuation	509	68.
┆	pitch reset, minor	ASSERTS	U+22A6	Mathematical Operators		69.
┆	pitch reset, major	MODELS	U+22A9	Mathematical Operators		70.
┆	amplitude reset, minor	DOWN TACK	U+ 22A4	Mathematical Operators		71.
┆	amplitude reset, major	INTERCALATE	U+22BA	Mathematical Operators		72.
^	reset up	UP ARROWHEAD	U+2303	Miscellaneous Technical		73.
v	reset down	DOWN ARROWHEAD	U+2304	Miscellaneous Technical		74.
„	rush start	DOUBLE LOW-9 QUOTATION MARK	U+201E	General Punctuation		75.
^	rush end	CARET	U+2038	General Punctuation		76.
☺	smile voice with speech	WHITE SMILING FACE	U+263A	Miscellaneous Symbols		77.

Sign	Function (DT2)	Unicode Name	Unicode Number (hex)	Unicode Subset	IPA #
‰	glottal stop, prosodic	PER MILLE SIGN	U+2030	General Punctuation	78.
ʔ	glottal stop, phonemic	LATIN LETTER GLOTTAL STOP	U+0294	IPA Extensions	79.
⊗	co-action: beat (pound, etc.)	CIRCLED LATIN CAPITAL LETTER X	U+24CD	Enclosed Alphanumerics	80.
⊙	co-action: snap	CIRCLED LATIN CAPITAL LETTER S	U+24C8	Enclosed Alphanumerics	81.
⊕	co-action: gesture	CIRCLED LATIN CAPITAL LETTER G	U+24BC	Enclosed Alphanumerics	82.
①	co-action: user-defined type 1	CIRCLED DIGIT ONE	U+2460	Enclosed Alphanumerics	83.
②	co-action: user-defined type 2	CIRCLED DIGIT TWO	U+2461	Enclosed Alphanumerics	84.
③	co-action: user-defined type 3	CIRCLED DIGIT THREE	U+2462	Enclosed Alphanumerics	85.
④	co-action: user-defined type 4	CIRCLED DIGIT FOUR	U+2463	Enclosed Alphanumerics	86.
↑	high	UPWARDS ARROW	U+2191	Arrows	87.
↓	low	DOWNWARDS ARROW	U+2193	Arrows	88.
⤵	downstepped high	DOWNWARDS ARROW FROM BAR	U+21A7	Arrows	89.
⤴	[upstepped low]	UPWARDS ARROW FROM BAR	U+21A5	Arrows	90.
◻	[IPA = 'upstep']	MODIFIER LETTER RAISED UP ARROW	U+A71B	Modifier Tone Letters	518 91.
◻	[IPA = 'downstep']	MODIFIER LETTER RAISED DOWN ARROW	U+A71C	Modifier Tone Letters	517 92.
↑	top	UPWARDS TWO HEADED ARROW	U+219F	Arrows	93.
↓	bottom	DOWNWARDS TWO HEADED ARROW	U+21A1	Arrows	94.
↗	rise [IPA = 'global rise']	NORTH EAST ARROW	U+2197	Arrows	510 95.
→	level	RIGHTWARDS ARROW	U+2192	Arrows	96.
↘	fall [IPA = 'global fall']	SOUTH EAST ARROW	U+2198	Arrows	511 97.
↗ ↘	rise-fall (compound)	[iconic combination: rise + fall]	U+2197	Arrows	98.
			U+2198		
↘ ↗	fall-rise (compound)	[iconic combination: fall + rise]	U+2198	Arrows	99.
			U+2197		
↻	rise-fall (complex)	CLOCKWISE TOP SEMICIRCLE ARROW	U+21B7	Arrows	100.
↺	fall-rise (complex)	BOTTOM ARC ANTICLOCKWISE ARROW	U+293B	Arrows	101.
→↗	truncated level	RIGHTWARDS ARROW WITH STROKE	U+219B	Arrows	102.
↗↗	upkey	NORTH EAST DOUBLE ARROW	U+21D7	Arrows	103.

Sign	Function (DT2)	Unicode Name	Unicode Number (hex)	Unicode Subset	IPA #	
↘	downkey	SOUTH EAST DOUBLE ARROW	U+21D8	Arrows		104.
⇒	midkey	RIGHTWARDS DOUBLE ARROW	U+21D2	Arrows		105.
↑	high-key	UPWARDS DOUBLE ARROW	U+21D1	Arrows		106.
↓	low-key	DOWNWARDS DOUBLE ARROW	U+21D3	Arrows		107.
≈	isotony start	ALMOST EQUAL TO	U+2248	Mathematical Operators		108.
≠	isotony end	NOT ALMOST EQUAL TO	U+2249	Mathematical Operators		109.
≡	isochrony start	IDENTICAL TO	U+2261	Mathematical Operators		110.
≢	isochrony end	NOT IDENTICAL TO	U+2262	Mathematical Operators		111.
1	top	MODIFIER LETTER EXTRA-HIGH TONE BAR	U+02E5	Spacing Modifier Letters	519	112.
1̊	high	MODIFIER LETTER HIGH TONE BAR	U+02E6	Spacing Modifier Letters	520	113.
1̋	mid	MODIFIER LETTER MID TONE BAR	U+02E7	Spacing Modifier Letters	521	114.
1̌	low	MODIFIER LETTER LOW TONE BAR	U+02E8	Spacing Modifier Letters	522	115.
1̍	bottom	MODIFIER LETTER EXTRA-LOW TONE BAR	U+02E9	Spacing Modifier Letters	523	116.
l̊	rising (IPA)	(iconic combination based on Unicode 02E5-02E9)	U+02E9 U+02E5	Spacing Modifier Letters	529	117.
l̋	falling (IPA)	(iconic combination based on Unicode 02E5-02E9)	U+02E5 U+02E9	Spacing Modifier Letters	530	118.
ľ	high rising (IPA)	(iconic combination based on Unicode 02E5-02E9)	U+02E6 U+02E5	Spacing Modifier Letters	531	119.
l̍	low rising (IPA)	(iconic combination based on Unicode 02E5-02E9)	U+02E9 U+02E8	Spacing Modifier Letters	532	120.
l̊l̋	rising-falling (IPA)	(iconic combination based on Unicode 02E5-02E9)	U+02E7 U+02E6 U+02E7	Spacing Modifier Letters	533	121.
ˈ	rise	MODIFIER LETTER ACUTE ACCENT	U+02CA	Spacing Modifier Letters	513	122.
ˉ	level	MODIFIER LETTER MACRON	U+02C9	Spacing Modifier Letters	514	123.

Sign	Function (DT2)	Unicode Name	Unicode Number (hex)	Unicode Subset	IPA #	
`	fall	MODIFIER LETTER GRAVE ACCENT	U+02CB	Spacing Modifier Letters	515	124.
^	rise-fall	MODIFIER LETTER CIRCUMFLEX ACCENT	U+02C6	Spacing Modifier Letters	525	125.
ˇ	fall-rise	CARON	U+02C7	Spacing Modifier Letters	524	126.
ǣ	extra-high tone (top)	COMBINING DOUBLE ACUTE ACCENT	U+030B	Combining Diacritical Marks	512	127.
á	high tone	COMBINING ACUTE ACCENT	U+0301	Combining Diacritical Marks	513	128.
ā	mid tone	COMBINING MACRON	U+0304	Combining Diacritical Marks	514	129.
à	low tone	COMBINING GRAVE ACCENT	U+0300	Combining Diacritical Marks	515	130.
ǎ	extra-low tone (bottom)	COMBINING DOUBLE GRAVE ACCENT	U+030F	Combining Diacritical Marks	516	131.
Ǻ	rising tone [IPA]	COMBINING CARON	U+030C	Combining Diacritical Marks	524	132.
ǻ	falling tone [IPA]	COMBINING CIRCUMFLEX ACCENT	U+0302	Combining Diacritical Marks	525	133.
	high rising contour tone [IPA]	COMBINING MACRON-ACUTE	U+1DC4	Combining Diacritical Marks Supplement	526	134.
	low falling contour tone [IPA]	COMBINING GRAVE- MACRON	U+1DC5	Combining Diacritical Marks Supplement	527	135.
	(?) contour tone	COMBINING MACRON-GRAVE	U+1DC6	Combining Diacritical Marks Supplement	?	136.
	(?) contour tone	COMBINING ACUTE- MACRON	U+1DC7	Combining Diacritical Marks Supplement	?	137.
	rising-falling contour tone [IPA]	COMBINING GRAVE-ACUTE-GRAVE	U+1DC8	Combining Diacritical Marks Supplement	528	138.
	(?) contour tone	COMBINING ACUTE-GRAVE-ACUTE	U+1DC9	Combining Diacritical Marks Supplement		139.
ᳵ	downstepped tone	COMBINING LEFT ANGLE ABOVE	U+031A	Combining Diacritical Marks		140.
a ⁼	same pitch as prior	SUPERSCRIPIT EQUALS SIGN	U+207C	Superscripts and Subscripts		141.
┐	tier event type 1 start	LEFT CEILING	U+2308	Miscellaneous Technical		142.

Sign	Function (DT2)	Unicode Name	Unicode Number (hex)	Unicode Subset	IPA #
┘	tier event type 1 end	RIGHT CEILING	U+2309	Miscellaneous Technical	143.
┘	tier event type 2 start	LEFT FLOOR	U+230A	Miscellaneous Technical	144.
┘	tier event type 2 end	RIGHT FLOOR	U+230B	Miscellaneous Technical	145.
△	prior's pause	WHITE UP-POINTING SMALL TRIANGLE	U+25B5	Geometric Shapes	146.
▽	next's pause	WHITE DOWN-POINTING SMALL TRIANGLE	U+25BF	Geometric Shapes	147.
▷	open pause	WHITE RIGHT-POINTING SMALL TRIANGLE	U+25B9	Geometric Shapes	148.
<	gaze approach start	PRECEDES	U+227A	Mathematical Operators	149.
>	gaze depart end	SUCCEEDS	U+227B	Mathematical Operators	150.
≤	mutual gaze start	PRECEDES OR EQUAL TO	U+227C	Mathematical Operators	151.
≥	mutual gaze end	SUCCEEDS OR EQUAL TO	U+227D	Mathematical Operators	152.
≠	no mutual gaze start	DOES NOT PRECEDE OR EQUAL	U+22E0	Mathematical Operators	153.
≠	no mutual gaze end	DOES NOT SUCCEED OR EQUAL	U+22E1	Mathematical Operators	154.
▣	gesture approach start	SQUARE IMAGE OF	U+228F	Mathematical Operators	155.
▣	gesture departure end	SQUARE ORIGINAL OF	U+2290	Mathematical Operators	156.
▣	gesture stroke start	SQUARE IMAGE OF OR EQUAL TO	U+2291	Mathematical Operators	157.
▣	gesture stroke end	SQUARE ORIGINAL OF OR EQUAL TO	U+2292	Mathematical Operators	158.
≠	no gesture	NOT SQUARE IMAGE OF OR EQUAL TO	U+22E2	Mathematical Operators	159.
≠	no gesture	NOT SQUARE ORIGINAL OF OR EQUAL TO	U+22E3	Mathematical Operators	160.
₂	bracket index	SUBSCRIPT TWO	U+2082	Superscripts and Subscripts	161.
₃	bracket index	SUBSCRIPT THREE	U+2083	Superscripts and Subscripts	162.
₄	bracket index	SUBSCRIPT FOUR	U+2084	Superscripts and Subscripts	163.
₅	bracket index	SUBSCRIPT FIVE	U+2085	Superscripts and Subscripts	164.
₆	bracket index	SUBSCRIPT SIX	U+2086	Superscripts and Subscripts	165.
₇	bracket index	SUBSCRIPT SEVEN	U+2087	Superscripts and Subscripts	166.
₈	bracket index	SUBSCRIPT EIGHT	U+2088	Superscripts and Subscripts	167.
₉	bracket index	SUBSCRIPT NINE	U+2089	Superscripts and Subscripts	168.

Sign	Function (DT2)	Unicode Name	Unicode Number (hex)	Unicode Subset	IPA #
□	comment start	LEFT DOUBLE PARENTHESIS	U+2E28	Supplemental Punctuation	169.
□	comment end	RIGHT DOUBLE PARENTHESIS	U+2E29	Supplemental Punctuation	170.
((comment start	LEFT PARENTHESIS + LEFT PARENTHESIS	U+0028	Basic Latin	171.
))	comment end	RIGHT PARENTHESIS + RIGHT PARENTHESIS	U+0029	Basic Latin	172.
{	syntactic phrase start	LIGHT LEFT TORTOISE SHELL BRACKET ORNAMENT	U+2772	Dingbats	173.
}	syntactic phrase end	LIGHT RIGHT TORTOISE SHELL BRACKET ORNAMENT	U+2773	Dingbats	174.
⌈	syntactic phrase start	LEFT WHITE SQUARE BRACKET =left abstract syntax bracket	U+301A	CJK Symbols and Punctuation	175.
⌋	syntactic phrase end	RIGHT WHITE SQUARE BRACKET =right abstract syntax bracket	U+301B	CJK Symbols and Punctuation	176.
⌈	syntactic phrase start	LEFT WHITE TORTOISE SHELL BRACKET	U+3018	CJK Symbols and Punctuation	177.
⌋	syntactic phrase end	RIGHT WHITE TORTOISE SHELL BRACKET	U+3019	CJK Symbols and Punctuation	178.
□	syntactic phrase start	MATHEMATICAL LEFT WHITE TORTOISE SHELL BRACKET	U+27EC	Miscellaneous Mathematical Symbols-A	179.
□	syntactic phrase end	MATHEMATICAL RIGHT WHITE TORTOISE SHELL BRACKET	U+27ED	Miscellaneous Mathematical Symbols-A	180.
□	syntactic phrase start	MATHEMATICAL LEFT FLATTENED PARENTHESIS	U+27EE	Miscellaneous Mathematical Symbols-A	181.
□	syntactic phrase end	MATHEMATICAL RIGHT FLATTENED PARENTHESIS	U+27EF	Miscellaneous Mathematical Symbols-A	182.
□	syntactic phrase start	LEFT WHITE CURLY BRACKET	U+2983	Miscellaneous Mathematical Symbols-B	183.
□	syntactic phrase end	RIGHT WHITE CURLY BRACKET	U+2984	Miscellaneous Mathematical	184.

Sign	Function (DT2)	Unicode Name	Unicode Number (hex)	Unicode Subset	IPA #
○	unidentified phonetic segment	SHADOWED WHITE CIRCLE [=BALLOON]	U+274D	Symbols-B Dingbats	611 185.
⌘	real name	REFERENCE MARK	U+203B	General Punctuation	186.
√	root (grammar)	SQUARE ROOT	U+221A	Mathematical Operators	187.
**	interesting item	ASTERISM	U+2042	General Punctuation	188.
⌘	interesting item (alternate)	PLACE OF INTEREST SIGN	U+2318	Miscellaneous Technical	189.
:	omission of lines from excerpt	VERTICAL ELLIPSIS	U+22EE	Miscellaneous Technical	190.
⌘	audio filter start	APL FUNCTIONAL SYMBOL DOWN CARET TILDE	U+2371	Miscellaneous Technical	191.
⌘	audio filter end	APL FUNCTIONAL SYMBOL UP CARET TILDE	U+2372	Miscellaneous Technical	192.

[rev. 26-Sep-2009]

Appendix 7

Transcribing Procedures

Introduction

This document gives a broad overview of transcribing procedures. It focuses on procedures, methods, and practical mechanisms that we will use to organize the practice of transcription in this course. A more fine-grained treatment of the details of the transcription process itself can be found in the main textbook for the course (*Discourse Transcription*).

In the approach we take in this course, transcribing a recording entails two main tasks: transcription and checking. During much of the course, you will transcribe one minute of spoken discourse per week. Each week you will also check your partner's minute of transcription from the week before. Each of these tasks involves a variety of subtasks, to be described below.

The Point of Transcription

Why transcribe? Transcription is more than just manual labor—though it surely is hard work. Nor is it merely a means to an end: It's not just a way to get discourse data to use in your research. Rather, the very practice of transcribing has value in and of itself. Many transcribers find that if you put effort and insight into your transcription practice, keeping your eyes (and ears) open and thinking about what you are doing, the transcribing process will teach you more about discourse than any book, article, or lecture ever could.

Recordings

The recordings we will work with in this course will be digitized in a computer format. This facilitates collaboration and exchange of data between course participants. It also makes the transcribing easier and more effective, enabling use of computer-assisted transcription techniques (see below).

It is important to use consistent conventions for creating and naming the computer files for recordings and transcriptions. Consult the relevant Appendices (Ethnographic Recording and Transcription Format) for details.

The recording file that we will be using for most of our work is the Research Segment file (20-40 minutes long). Within this file, we will be mostly working on the Transcription Excerpt, a 4- or 5-minute long subset of the Research Segment. It is very important for you to mark the exact start and end of the Transcription Excerpt within the Research Segment, using your sound editor (see the Appendix on "Digitizing and Editing Recordings"). It is the Transcription Excerpt file that you will use for time-aligning your transcription with the corresponding audio (see "Alignment" below).

Delicacy

In the initial stages of learning about transcription, the transcriptions you will be asked to produce are relatively "broad," as opposed to "narrow". But a broad transcription does not mean a loose one. It can be as rigorous in its own terms as a narrow one is, for what it seeks to represent. As the course goes along, the transcriptions will become more detailed, especially in the area of prosody. For a given week's assignment, the expected level of delicacy will be specified in the assignment schedule handout. See the

handout on “Transcription Delicacy Hierarchy” for specific information on which transcriptional features are to be included at each level of delicacy.

Partners

To gain the benefits of working and talking with your peers about transcribing issues, you will be paired with a partner. All the transcribing and checking work that you do in this course will be done with the support of your partner (and vice versa). Normally you and your partner will be working on different conversations, so that you will gain broader experience through working on two different conversations.

Checking

In addition to transcribing your own conversation, each participant will check their partner’s transcription, on a weekly basis. Because you are transcribing in a team, you will gain the benefit of a second opinion on how to represent what is going on in your conversation. When you check your partner’s conversation, you should offer your insights and alternative interpretations. You should expect to receive the same careful attention in return from your partner.

It is important to take the task of checking seriously. When done well, checking requires about the same time, effort, and attention as does transcribing in the first place. When you turn in your checking of your partner’s transcription to the instructor (e.g. using the OpenOffice function “Edit>Record>Changes,” or the MS Word “Track Changes” function of your word processor), it is expected that there will be lots of marks indicating corrections of your partner’s work, reflecting the fact that you are doing a careful job of it.

Note that you are only responsible for checking those features that your partner has made a serious attempt to transcribe. For example, if you notice that your partner’s transcription contains no pauses at all (even though pauses were supposed to be transcribed in the assigned delicacy level), you are not responsible for identifying all the pauses for your partner. Instead, simply indicate that the necessary feature was entirely omitted, and leave it to your partner to find them in the revision stage. On the other hand, if your partner’s transcription does have pauses marked, and thus displays a good-faith effort to identify the pauses in the recording (with a plausible level of accuracy and completeness), you are then responsible for checking and identifying any pauses your partner may have missed.

Revision

As soon as you have received your partner’s checking of your transcription, you should go back and review your recording in order to update and improve your transcription. Of course, the suggested corrections which are marked in your transcript are themselves interpretations, and hence are not necessarily “correct”. You have to review the recording and evaluate the suggestions for yourself. Consider the suggestions of your partner and your instructor carefully. Think about what the different transcriptions mean, and what justifications exist for your original transcription versus the proposed revisions. If a proposed change seems valid to you, incorporate it in your revision.

Reconciliation

On points where you disagree with your partner, it is very valuable to have a “reconciliation” meeting to go over the corrected transcription together. You can learn a lot by trying to argue for or against a particular transcription decision with your partner—that is, give specific reasons for your position,

naming any relevant auditory cues, and so on. Try to set up a regular weekly time to meet for this purpose. Even if you agree with all of your partner's suggestions or corrections, a reconciliation meeting to discuss the ideas as they apply to your transcription can be valuable.

Updating

You may choose to incorporate changes into your transcription as you go along, so that when you turn in your transcription file for the following week, any earlier portions of the transcriptions will be revised, up-to-date, and brought up to the standard of your current knowledge about transcription. Or, you may prefer to simply incorporate corrections of errors each week, and wait on upgrading the delicacy level of your previous minutes of transcription until you can revise a larger portion at one time.

In any case, when the time comes for turning in the *revision* of Minutes 1-4 of your transcription (i.e. normally around the 6th or 7th week of the course), you will need to fully revise and correct the entire four-minute version. In addition, you will need to bring the whole transcription up to a consistently high level of delicacy.

Explication

A number of class sessions will be devoted to transcription sessions, in which we will not only transcribe, but also discuss the transcription decisions you make, and your rationale for them. Of course, identifying the right transcription category for a phenomenon you observe to be present in the recording data is important, and this decision is overtly expressed by selecting the right symbol for what you hear or see. But what is more important is to be able to understand, and to articulate, the reasons for invoking one particular transcriptional analysis over another. Because a transcription is always an interpretation of reality, it is important to be able to explicate the representations you create. To explain and justify your analysis to other session participants, you need to understand the concept behind each transcriptional category, as well as the specific cues which signal it. You need to be able to motivate and explicate your interpretation.

So that we can make the most effective use of our limited time for in-class transcription sessions, you are expected to be well prepared, understanding the assigned readings, as well as all updated material from the handouts and course packets in a way that lets you explain your decisions. You should try to develop a command of all the transcribing categories, cues, conventions, symbols, and practices described in the materials assigned up to a given date, and be able to explicate and motivate them as we discuss specific transcription problems in class.

Consultation

Because of the number of students in the class and the large quantities of transcription that are collectively generated during the course, it is not possible for the instructor to provide feedback on every word of every student's transcription. However, the instructor will check selected portions of your work at intervals during the course, and offer oral feedback, in class and in semi-private consultation sessions.

To get the most effective input, you (along with your transcription partner) should present your transcription for oral evaluation several times during the quarter. For each consultation session, you are expected to present a portion of the transcription you have recently completed. We will go over it together to talk about relevant transcription issues. Half of the time of each session will be devoted to your transcription, and half to your partner's. (If the number of students in the course is unusually large,

we may need to do some of these instructor consultation sessions in larger groups.)

For each transcription consultation session, you should come prepared with the following:

- Your recording should be available for playback on the instructor's computer, in the form of a digitized computer file (audio should be in .WAV format), accessed either over the Linguistics Department network, or via a USB thumb drive or CD-R, or preferably both. Make sure you locate and *test* your computer file to be sure it works before you come to present your work for instructor consultation. This is so we don't waste time trying to figure out how to access your file and get it to work.
- Pick a *specific section* of your recording that you found challenging or interesting to transcribe, and which you would like to talk about. The amount of material you select should be about half a page (or a page at most), which generally provides more than enough to talk about.
- You should know *exactly* where in the computer file to find each item you want to talk about. Identify the exact time location of the relevant section, and insert a timestamp in your transcription (e.g. <T=127.3>).
- You should bring a paper printout of your selected page of transcription (no more than *one* page), with one copy each for yourself, your partner, and your instructor.

Note that the instructor consultation sessions are an integral and essential part of the course. You are responsible for reserving time for these sessions with the instructor.

Questions

In the instructor consultation session, you will get “live” feedback on your transcribing. In addition, you are likely to be asked impromptu questions about your transcription by your instructor. (You can also raise questions of your own about how to transcribe the specific phenomena in your conversational excerpt, but theoretical discussion of concepts is not the main focus of these transcription consultation sessions.)

Examples of typical questions you may be asked include:

- Is there is an intonation unit boundary between these two words? Why or why not?
- What prosodic cues can you identify, which you could use to argue for the boundary, or against it?
- What intonation unit cues are missing in this case? Does that change your interpretation?
- Who does this pause belong to? Why?
- Where are the primary and secondary accents in this line? What prosodic cues support your position?
- and so on...

Of course no one is expected to know how to answer such questions at the beginning of the course—that is why we're in the course to learn. All of the above are questions we will talk about together in class, so you will have a chance to learn how to master this kind of analysis. When the time comes in the course that we have arrived at this level of transcription, you will be expected to come prepared to explain your transcription decisions, along the lines of the above questions.

Computer-Assisted Transcription (CAT)

As much as possible we will make use of techniques of computer-assisted transcription during this course, learning how to use software such as Sound Forge, Audacity, VoiceWalker, EXMARaLDA, and/or PRAAT. This means that all recordings will need to be digitized for use on a computer. For this reason, you should:

- create a copy of your digitized computer files, along with those of your partner(s), on a portable format, preferably a USB thumb drive (alternatively, on a CD-ROM or DVD-ROM disk)
- bring your computer-readable file containing the relevant recordings with you to class sessions, team meetings, instructor consultation sessions, etc.
- in addition, make sure to keep an up-to-date version of your transcription and recording files in the appropriate folder in the Linguistics Lab network.

Several software tools that we will use for computer-assisted transcription are available free to be downloaded from the web, while others may need to be purchased. Students should refer to the relevant course materials for information on how to obtain these items for use on your own computer.

Alignment

Using a tool like Sound Forge, with its Regions function (or Audacity, Transcriber, Transana, EXMARaLDA, etc.), you should begin to align your transcription with the corresponding recording as soon as possible during the transcription process. This means that for each intonation unit you have defined, you should identify the corresponding “Region” of the recording. This will make it possible to play back the exact portion of the recording which corresponds to each of your intonation units. This makes it much easier for you, your partner, and your instructor to check the accuracy of your transcription. It also makes it easier to gradually refine the granularity of your transcription, as you listen again and add more detail.

For your alignment work, you should use your Research Segment file. (If, after this course is over, you decide you want to expand your work with this Research Segment for your own future research projects, you will be glad you can simply add to all the work you’ve already done on the Transcription Excerpt by extending your transcription and alignment to additional portions of the Research Segment.)

Bleeping

As you time-align your transcription with the corresponding portions of the recording, you should also use the Regions function to mark the location on your recording of any names or other words that could compromise anonymity (see the section on “Anonymity” in the “Transcription Format” Appendix). Before publishing the data, it will be necessary to bleep or otherwise mask these words, so they cannot be recognized.

Language

For doing the transcriptions in this course, you are welcome to work on any language you know very well. For purposes of (unassisted) transcription, this means you should know the language natively, near-natively, or at least fluently.

Schedule

Consult the syllabus and the assignment schedule for the dates when transcriptions and checkings are due. Transcriptions and checkings are always due weekly, on the assigned date. Please note that just before class is probably not a good time to try to email or print out work due on that day, as the computer or printer you want may be in use, or otherwise unavailable. In any case, do not miss class in order to “finish up” a late assignment—the class sessions are too important to miss.

Timeliness

It is important that you do your transcription on time, in order to keep up with transcription activities and discussions in class, and because your partner and instructor need to receive your work on time in order to be able to provide you with feedback. The checking also needs to be done on schedule, in order to give your partner feedback on their transcriptions in time for them to incorporate improvements into their next revision. For these reasons, TRANSCRIPTION AND CHECKING ASSIGNMENTS TURNED IN LATE WILL RECEIVE REDUCED CREDIT.

Communication

For communicating with other participants in the course, we will need to exchange transcriptions frequently during this course, for such purposes as checking, giving feedback, and grading. For such data transfer functions we will use electronic means as much as possible, along with the occasional computer printout.

When the time comes each week to turn in an additional increment of transcription, please do the following BEFORE CLASS on the date due:

- Place a copy of all files relevant to the current assignment in the designated directory for this course. For example, using a computer on the Linguistics Lab network, find the “Everyone” drive, and copy your files into the *Ling 212* directory, under the specific subheading for your class and year (e.g. “*Class 09*” for the year 2009).
- Include the following items:
 - your updated transcription file. In general, the required format is (1) .wav audio file (or Sound Forge.sfk or similar file); and (2) OpenOffice.org Writer .odt file.
 - your updated audio file (.wav format) if appropriate.

By keeping your transcription and audio files up-to-date and accessible, you make it possible for your partner(s), classmates, and instructor to check your transcription, give feedback, and so on.

You should give your transcription to the relevant people each week:

- Give your *transcription* file in the relevant format (may be Sound Forge .sfk, OpenOffice.org Writer .odt, etc.) to two people:
 - (1) one copy to your partner, to be checked, and
 - (2) one copy cc’ed to your instructor, for evaluation.
- Give a *hard copy* version of your transcription in a well-formatted version to the same people. Please note that producing this version is likely to require some additional work, since additional formatting (e.g. tabs, etc.) may need to be introduced when you produce the paper version.

Once you have checked your partner's transcription, you will also need to do the following:

- Give your *checking* file in the relevant format (containing your mark-up of your partner's transcription from the previous week), using the computer or hard copy format in which you received it, to two people:
 - (1) one copy to your partner, for revision and reconciliation, and
 - (2) one copy cc'ed to your instructor.

Format

An important part of transcribing is to use a consistent format for presenting and exchanging information. See the handout on "Transcription Format" for a detailed discussion of this topic, including important conventions for naming your transcription and audio files.

[rev. 25-Sep-2009]

Appendix 8

Transcription Format

Objective

The purpose of this document is to describe the format to be used for producing and checking transcriptions in this course.

Conventions

The main transcription system we will use in the course is described in “Outline of Discourse Transcription” (Du Bois et al., 1993) and *Discourse Transcription* (Du Bois et al., 1992), as updated in Appendix 3 to *Representing Discourse* (2008). Note that, although the theory, discussion, and examples of transcription *categories* as presented in the earlier documents generally remain valid, the transcription *symbols* have undergone a few revisions in the time since the earlier publications appeared. For all the transcribing we do in this course, it is important to use the most recent updated symbol conventions, as indicated in class handouts.

Font

You may wish to use a Unicode font, such as Charis SIL (preferred), Doulos SIL, Gentium, or Arial Unicode MS. (The first three are available for free on the Internet from SIL International, while the last is proprietary but it often supplied with Microsoft software.) This will give you the widest range of options for using special symbols, including characters capable of writing any of the languages of the world in their traditional standard orthography, plus the International Phonetic Alphabet, and many other specialized symbols. (The more familiar Times New Roman font is another option; it has some Unicode characters, but not as complete a set.)

Tabs

Tabs are used in transcribing to help organize the presentation of transcription information, displaying it in a usefully iconic way. For our purposes, tabs are to be used for one purpose only: to distinguish between separate “fields” of data in each line. For a basic transcription, there are two fields. The first field in each line indicates who the speaker is (via a speaker label written in capital letters, followed by a semi-colon). The second field represents the actual utterance, or transcribed speech.

Define your tab location as follows (for the whole file). Define one tab at 2.5 cm (or 1 inch). Insert exactly one tab character in each transcription line:

- 55. if there is a speaker label in the line, the tab is inserted immediately following it
- 56. if there is no speaker label in the line, the tab is the first character of the line

Note that you should NOT use tabs for other purposes, such as inserting blank space in a line in order to align overlaps vertically (just insert spaces instead).

(In a more complex transcription there may be more than two fields, but this issue will be addressed separately.)

Line spacing

Do not use double-spacing, and do not leave blank lines between turns. (The exception is if you are transcribing a language other than English which needs glossing; see below).

Timestamp

Following the last word spoken in each weekly transcription increment, insert a time stamp into your transcription, indicating the location of the corresponding audio in your digitized audio file. For example, if the last word of your first week's transcription ends at 60.6 seconds into the computer audio file, insert the following notation at the end of the line: **<T= 60.6>**. This notation allows you, your teammate(s), and your instructor to quickly find the relevant place in the audio file, in order to keep track of where each week's new transcribed portion begins for checking purposes, and so on. You may wish to add additional time stamps at various points in your transcription, for convenience in locating specific portions of the transcription, for checking, feedback/consulting sessions, presenting segment proposals, and so on. (Or, you may find it useful to time-align the entire transcription with the corresponding audio, using appropriate software. This will amount to time-stamping the start and end of every intonation unit.)

Index

It is useful to index the transcription to make it easier to refer to specific portions of the data. Each line (or other unit) in the transcription is given its own unique index. There are at two methods which are particularly useful for this: numbering each line, and/or timestamping each line.

For line numbering, it is best not to number lines manually, since any revision of the transcription which adds even one new intonation unit could require you to renumber the entire subsequent transcription. It is better to add (temporary) line numbers automatically, at least while the transcription is in progress and has not been finalized.

In OpenOffice.org Writer: To automatically insert line numbers in the entire document, select "Tools/Line Numbering" from the main menu. Check the box for "Show numbering," set the "Interval" to 1, and then click OK.

In Microsoft Word: To automatically insert line numbers in the entire document, select "File" from the menu; select PageSetup/Layout/LineNumbers; check "Add line numbering" and "Continuous".

Timestamping is more stable, but can be somewhat laborious. If you choose this method, you should do it using software specifically designed to facilitate timestamping (e.g. Sound Forge, Transcriber, etc.).

In Sound Forge (etc.): To manually insert time stamps for each intonation unit, one at a time, follow the procedures discussed in the relevant manuals or handouts for Sound Forge, or other similar program.

Anonymity

Names of participants used in the transcripts (both in speaker labels, and when one discourse participant utters another participant's name) should not be the speakers' real names, unless you have received written consent from the participant to use their first name. Otherwise, make up an appropriate pseudonym – one which is comparable sociolinguistically and prosodically – and use it consistently.

Any pseudonym (or other words you have modified because they could compromise anonymity, such as an utterance of a participant's street address or phone number) should be indicated with the pseudograph symbol (i.e. tilde ~). (Note, however, that the tilde is used only for uttered words on the recording, not for speaker attribution labels, even if these are pseudonyms.)

For speaker attribution labels, give the complete first name, all in capital letters, not just an initial (e.g. use JILL: rather than J:).

Before disseminating your recording via the web or CD-ROM, you will need to make sure that any anonymity-sensitive words have been “bleeped” or otherwise obscured. You will learn simple techniques for doing this in this course.

Checking Format

Indicate your suggestions to your teammate by marking your proposed changes in a copy of your teammate's computer file. Use the “Record Changes” function in OpenOffice.org Writer (see below), or the “Track Changes” function in Microsoft Word, or the equivalent function in another program. This will allow you, your partner, and the instructor to easily distinguish between your suggested changes and the original transcription. Your suggestions will normally be displayed in underlined or strikeout font, or in color, or some combination of these. In addition, they may include the name of the person who made the change.

Using the “Record Changes” or “Track Changes” function on a word processor file will also allow you to exchange the checked files easily via email attachments. (If you are unable to use the Track Changes or Record Changes function, you may mark up a paper copy of your partner's transcription, using colored ink or pencil so it will stand out.)

OpenOffice.org Writer: To turn on the “Record Changes” function, select Edit/Changes/Record from the main menu. Your insertions and deletions will now be tracked (using a format similar to that used by Microsoft Word, see below). To review and accept or reject your partner's suggestions, from the main menu select Edit/Changes/Accept_or_Reject.

Microsoft Word: To turn on the “Track Changes” function, select Tools/Track Changes from the menu. Or, as a shortcut, type CTRL-SHIFT-e. To turn off the Track Changes function, do the same again. To view the changes in Word, make sure the “Reviewing” toolbar is displayed. Normally it will be displayed automatically once you select the “Track Changes” function. (To display the “Reviewing” toolbar, if necessary: From the menu, select View/Toolbars, and then make sure the “Reviewing” option is checked.) The reviewing toolbar has several icons representing functions you may find useful for accepting (or rejecting) your partner's suggestions. You may wish to experiment with these.

Glossing

If you are transcribing a language other than English, you may have to gloss it for your instructor. Normally the best way to do this is with interlinear gloss format (two-line or three-line format). Glossing requires some extra work, but the burden is relatively small in comparison with the work of transcribing. (For extensive information on glossing, see the “Leipzig Glossing Rules” at <http://www.eva.mpg.de/lingua/files/morpheme.html>.)

File Sharing

Because of the emphasis on teamwork, this course will involve a lot of sharing of recordings and transcriptions in the form of computer files. We will do as much as possible of this data-sharing via the Internet (and/or the computers and local area network in the Linguistics Lab). Further details about how to exchange data files with your partners will be given separately. The two most important types of files we will be using are the transcription file and the recording file.

Transcription File

You should transcribe each recording in a single computer file (e.g. a word processor file). You will add new material to this file each week, incorporating each additional minute of the recording as you transcribe it.

If possible, use **OpenOffice.org Writer 2.3** (or later), which saves your data in the Open Document format (.odt). This has the great advantage that your data will be saved in an open standard (using open source software). If this is not feasible for you, you may use another word processor such as Microsoft Word (.doc). (If you use a different word processor from either of these, save your file in the “plain text” (.TXT) or HTML formats, to facilitate sharing with other course participants.)

Recording Files

You are responsible for making sure that the appropriate digitized computer files corresponding to your recording (audio and/or video) are available to all relevant participants (e.g. your teammates and the instructor), via course web pages, the Internet, CD-ROM, and/or the Linguistics Lab computers and the Linguistics Department’s Local Area Network. Your computer files should be available in advance of the time they will be needed, e.g. prior to your team meetings, classroom feedback sessions, consultations in office hours, and so on, as well as being ready for your partner to check. Test your file setup before any meeting to make sure it works with the computer you will be using.

To include your files as part of the Class Corpus for Linguistics 212 at UC Santa Barbara for the year 2005 (for example), look in the “Everyone” drive (a.k.a. the “i drive”). There will be a folder with a name like “**Ling212**,” and under this, the “**Class05**” directory. You should make a point of regularly placing your updated files there when you are ready to share them.

Filenames

To make it possible for course participants (especially your instructor and your partner) to reliably keep track of all the computer data files that we will be exchanging, you should use a consistent practice to assign filenames to all your files. This includes your various recording files (called Transcription Excerpt, Research Segment, and Complete), your transcription file, and your checking file. Note that the checking file will normally be a file that was first originated by your partner (as her/his transcription file), which you have subsequently modified by adding your own corrections, suggestions, and comments to it. Each filename should make it clear who the file belongs to (i.e. you, as the one who worked on it most recently), and what it contains, including the relevant minute of transcription and checking.

Construct the filenames for your various digitized *recording files* as follows. Use your name (first or last, but let’s be consistent), in lower case, as a starting point.

1. For your *Transcription Excerpt* recording file (usually 4-8 minutes long): **yourname.wav**

1. For your *Research Segment* recording file (usually 20-30 minutes long): , add “_segment”.
2. If you made a “Complete” digitized version of your recording file (i.e. a digitized file corresponding to the entire original recording), add “_complete”.

Avoid using spaces in the filenames; use an underscore instead. The audio or video editing software should automatically add the appropriate file extension indicating what kind of computer file you are creating, e.g. an audio file (normally .wav) or a video file, as the case may be. The result should look like the following (assuming you are working with an audio recording):

1. your Transcription Excerpt recording file: yourname.wav
1. your Research Segment recording file: yourname_segment.wav
1. your Complete recording file (optional): yourname_complete.wav

The following conventions apply to the filenames for your text-based *transcription files* (i.e. word processor files only). (Note that if you are doing your transcriptions using a sound editor like Sound Forge, you should keep your file name the same during the whole course.) The filenames for your various transcription and checking files should be on the following pattern:

1. your transcription: yourname.doc
1. your checking of your partner’s transcription: hername_checkedby_yourname.doc

To make a 4-minute (or 5-minute) version of your wav file, the best way to do it is as follows:

- Open the wav file you have been using to mark regions.
- Highlight the relevant 4-5 minutes, using the mouse.
- Mark a region that will consist of the whole 4 minutes. Be sure the start and end times for the 4 minutes are included in the region name, and add the word "Source" at the beginning of the region name. Click OK to save the region.
- Copy the excerpt to the Windows clipboard, using the Edit menu or control-C.
- Paste the excerpt from the clipboard into a new file, using control-E.

Exchange

See the handout on “Transcribing Procedures” for a detailed discussion of how transcriptions are to be exchanged between course participants.

References

- Du Bois, John W. 2004. *Representing Discourse*. MS, University of California, Santa Barbara.
- Du Bois, John W., Cumming, Susanna, Schuetze-Coburn, Stephan, and Paolino, Danae. 1992. Discourse transcription. *Santa Barbara Papers in Linguistics* 4.
- Du Bois, John W., Schuetze-Coburn, Stephan, Cumming, Susanna, and Paolino, Danae. 1993. Outline of discourse transcription. In *Talking data: Transcription and coding in discourse research*, eds. Jane A. Edwards and Martin D. Lampert, 45-89. Hillsdale, NJ: Erlbaum.

[Rev. 11-Oct-2007]

Appendix 10.1

Ethnographic Documentation

A researcher who captures a high quality audio or video recording of a naturally occurring interaction may have made a valuable beginning, but the job is not done until essential information about the event and its participants has been collected. Without this, an undocumented recording may lose much of its research value. It is easy enough to avoid this, assuming the researcher makes a plan in advance for asking the right questions, and then makes sure to ask them at the right time. The right time is, preferably, the day of the original recording. This section gives guidance on what kind of information is likely to be most useful for shedding light on the events and the participants being recorded. It also describes the process of getting this information. One side benefit of this process of ethnographic documentation is that the field worker has a chance to answer participants' further questions about what kind of project they are participating in.

Exit Interview

It takes time and patience to do a good job of collecting the various kinds of ethnographic and demographic information that are required in order to properly document a conversation or other kind of spoken interaction. For this reason, it is best to plan to do an exit interview, to be carried out after the recording has been completed. In the exit interview, the field worker presents the person who has just been recorded with a questionnaire asking for general demographic information about such things as age, native language, geographical origins, education, and so on. By asking such questions at the end of the interaction, the researcher avoids putting a spotlight on the individual and their linguistic background before the recording is made. If participants talk first and answer questions later, they are less likely to be made self-conscious during the recording itself. And of course, in an exit interview they can ask any questions they may have arisen for them about the nature of the study being done. (Recall that the participants are always given the right to withdraw their consent to participate, if they wish to do this after finding out more information about the project.)

Forms

The following forms are used to record the necessary ethnographic information about the recorded event, and demographic information about the participants in the event.

1 Consent

Be sure to have each person who appears on your audio or video recording sign a consent form, giving their informed permission for the recording. (See Appendix A.11 on "Consent" for details.)

2 Participant Questionnaire

The purpose of this questionnaire is to gather information about each participant's background, especially those aspects which may influence their form of speaking (such as age, sex, native language, origins, level of education completed, occupation, and so on). Have each person who appears in the recording fill out a questionnaire. You will probably need to help people as they fill out the form, clarifying any questions they may have. In some cases it may be easiest for the field worker to simply fill out the form for them, based on asking them the questions in interview style.

3 Event Description

As soon after the recording as you can, provide a general description of the event recorded, answering the detailed questions on the Event Description form. Include information on any artifacts that may shed light on the event, such as a newspaper that people were pointing to, a printed flyer or course handout they referred to, a specific computer game they were playing, etc.

4 Setting Diagram

This is a simple diagram of the arrangement of the participants within the setting where the recording took place. It is useful for audio recording, but usually superfluous for video recording. Provide the standard header information about the recording, researcher, and date, and then draw a simple diagram, indicating location of speakers, microphones, relevant furniture, and other relevant items.

5 Recording Log

The purpose of making a recording log is to gain a broad overview of the contents of a recording. The log can serve as a rough guide to the recording's contents, especially for locating portions of interest for further investigation. If possible, the field worker who made the recording should do the log, because they may have some relevant insights based on familiarity with the recording context. It is a good idea to make the recording log soon after the recording. (But the time factor is not as critical as it is in the case of the Participant Questionnaire, Event Description, and Setting Diagram.)

At the top of the first page of the log, indicate the following:

- title of the recording
- computer filename of the recording
- name of the researcher making the log
- date the log was made

While reviewing the recording, take notes to describe what is happening in it. For each entry you make, indicate the following:

- the time on the recording that you are referring to
- what is happening at that time
- who is participating (whenever this changes, i.e. when someone enters or exits)

Evaluate each portion of the recording, using a 5-point scale where 5 is high, for:

- video quality (if applicable)
- audio quality
- transcribability
- research potential

The evaluation for research potential is obviously subjective, and will vary depending on the research goals of the analyst. Nevertheless, it is useful for identifying portions of a recording that may be worth pursuing further via a full transcription.

6 Voice Identification Table

The purpose of making a Voice Identification table is to allow researchers who don't know the people on the recording to correctly identify them, that is, to make the right connection between a voice on the

recording and the individual who is represented in the speaker background information which was collected. Even if an arbitrary pseudonym is being used in place of the real name, it is still important to match the transcribed voice to the right demographic information. Note that this task is most important in a large research project where many different transcribers may work with the same data, and where there may be a time lag between making the recording and transcribing it. The task is a fairly quick one, especially if done by the field worker who originally made the recording. It can be easily done at the same time as a recording log.

At the top of the first page of the log, indicate the following:

- title of the recording
- computer filename of the recording
- name of the researcher making the table
- date the table was made

While reviewing the recording, keep track of all the people who appear in it. For each person on the recording, select some distinctive quote in their words. The quote should be one that will allow a future researcher who doesn't know this individual to identify them from the recording. By identify, we do not mean knowing their real name, of course, but simply making the right connection to the (anonymous) paperwork. For each person on the recording, indicate the following:

- the time on the recording when the event you are describing occurs
- who is speaking (use a pseudonym)
- what they said (a sentence is enough)
- how much they participate in this recording (high, low, minimal)
- speaker ID number to match them with the demographic information collected

[rev. 14-Sep-2003]

Appendix 10.4

Setting Diagram

Recording:		Filename:	
Researcher:		Date:	

Draw a diagram below of the setting in which the recorded event took place. Indicate the relative location of relevant aspects of the setting, such as the participants, microphones, table or other salient furnishing, entrance/exit, etc.

Appendix 10.5
Recording Log

Recording:		Filename:		Researcher:		Date:	
-------------------	--	------------------	--	--------------------	--	--------------	--

Time	What's happening	Who	Video	Audio	Trans.	Research

[rev. 14-Sep-2004]

Appendix 10.6

Voice Identification Table

Recording:		Filename:		Researcher:		Date:	
-------------------	--	------------------	--	--------------------	--	--------------	--

Time	Who	Distinctive Quote	Participation	ID #

[rev. 14-Sep-2004]

Appendix 11

Consent Form:

Corpus of Everyday Language Use

Part I: Explanation of Informed Consent

We are asking you (or your child or ward) to be a participant in a research study. This consent document will tell you about the purpose, risks, and benefits of this research study. You should consent only after you have had all your questions answered, and have had enough time to decide whether you wish to participate. Your signature on this form is voluntary and does not waive any of your legal rights, nor does it make the institutions and persons involved in this research any less responsible for your well being.

Part II: Explanation of the Current Study

Why have I been asked to be a part of this study?

We hope to record conversations and other activities of everyday life that you may participate in.

Who is responsible for the data collected in this study?

The persons responsible for the data in this study are:

Professor John W. Du Bois & Professor Mary Bucholtz
Linguistics Department
3520 South Hall
University of California, Santa Barbara
Santa Barbara, California 93106

Why is this study being done?

The purpose of this study is to understand how people use language in everyday life.

What is involved in this study?

We will record your daily activities, including conversations and other kinds of speaking, using an audio recorder or video camera. We will try to avoid interfering with your activities during the recording

process. After the recording has been completed, we will do an exit interview in which we will ask you to give us some general background information about yourself. In some circumstances, and with your permission, we may contact you again after the recorded activity if we have questions about the recording.

Will I know when I am being recorded?

We will record only after obtaining your permission to do so, and we will record only on the dates you have specified to allow recording. Because our audio recording equipment is unobtrusive, quiet, and records automatically without operator intervention, you may not always notice that it is on. You should assume that whenever our recording equipment is present, it is on and it is recording. In addition, we will remind you after the recordings are finished that recordings have been made, and will give you the opportunity to review the recordings at that time.

What about video recording?

Our video equipment requires the presence of a camera operator, who will always let you know when you are being recorded.

What form of recording am I permitting?

You may choose which form of recording you will allow. You can choose to be recorded with audio only, or with both video and audio.

What if children are present?

You may choose for your child (or ward) whether they may be included in this research study. For children under the age of seven, you may use this consent form. For children of age seven or older, there is a special consent form written for them, which presents information about the study in a form that is more understandable for children.

What kind of data will you collect?

We will collect three kinds of data:

- audio and/or video recordings of conversations or other events you are in
- information about the setting of the events, and about participants in the events
- transcriptions of the recordings

How will you gather information about participants?

We have a questionnaire that asks you a number of questions about your background (for example, where you grew up, your age, your work, and so on). You can fill out the questionnaire yourself, or we can help you fill it out. You can choose not to answer any questions you don't want to answer. We also have a questionnaire to gather basic information about the setting of the conversation or other event (for example, where it takes place, how many participants are involved, and so on). Our field worker will fill this out, and will show it to you if you like.

How do you make the transcriptions?

We transcribe by observing the recordings and writing down what is said and done on them.

Who will have access to the data?

Your data will be made available to researchers, students, and others interested in learning about how language is used in everyday life. The data will be available for research, education, and general interest. The video and audio recordings and transcriptions may be used in the following ways:

- Academic lectures
- Classroom teaching
- Research activities
- Scholarly publications
- Instructional materials
- General interest multimedia publications
- Public presentations
- Or other public and private situations where language use may be of interest.

What level of access to the data am I permitting?

By participating in this research, your data will be made available to researchers, students, and others interested. Your data may be used in oral, print, and electronic media (such as books, articles, notes, handouts, multimedia presentations, CD-ROM, and DVD.) In addition, you will be able to choose whether to allow your data to be made available on the Internet. The Internet web sites for accessing the data will require users to register and obtain a password. You will be able to choose whether to allow the use of your name in the data.

How long will I be in this study?

Your consent to be recorded extends only for the dates you specify on the consent form. Normally this will be between a few days and a few weeks. The recordings of your conversations and other activities

which are made during this period will be preserved indefinitely unless you request otherwise (see risks). Even after the recordings have been made, you have the right to withdraw from this study and remove your data. Either you or the investigator may terminate your participation in this study at any time.

What are the risks of this study?

Because of the nature of the data being gathered (that is, audio and video recordings), it may not be possible to conceal your identity as a participant. There is a potential risk that people known or unknown to you will formulate negative opinions of you or your behavior on the basis of their reviewing of these data. We have several safeguards against this:

- You can request that audio or video recording be stopped at any point during the activity, thereby preventing a recording from ever being produced.
- You can review the recording. If you wish, you can have the Principal Investigator or field worker delete or erase particular segments containing your voice or your likeness. We will have a form available for you to request this. You may request this at any time.
- As discussed in the section on confidentiality below, any additional identifying data such as your name will be seen only by researchers directly involved in this project, unless you give us permission to use it publicly.

Are there benefits to taking part in this study?

There are no direct benefits to you personally for participating in this study. The primary benefits from this work are for the advancement of scientific understanding of language use and communication processes. The availability of these data may lead to improvements in the fields of linguistics, communications studies, and related fields, of which you or others may be a direct beneficiary. If you are interested in the results, let us know and we will be happy to share them with you.

Will I receive any payment or other monetary benefits?

You will receive no payment for being recorded. You should not expect any royalties or payments in the future.

What other options are there?

You can choose not to participate in this study.

What about confidentiality?

We will take careful measures to safeguard your identity. Depending on the level of data access you allow, these measures will include:

- using a fake name in place of your name in our transcriptions and participant information records
- replacing any other identifying words such as an address or telephone number
- editing the sound, if your name or other identifying word is spoken on the recordings, in order to make it impossible to hear the name or other identifying information

However, the nature of the data being gathered (audio and video recordings) means that your identity as a participant cannot be concealed under all circumstances. Anyone who happens to know you personally would be able to identify you from the audio or video recording. However, the measures we will take are designed to help safeguard your identity from the general public.

What are the costs?

There are no costs to you for participating in this study.

What are my rights as a participant?

Taking part in this study is voluntary. You may choose to not take part or stop participation at any time. Leaving this study will not result in any penalty or loss of benefits to which you are entitled.

Can I have a copy of this information?

You have a right to receive a copy of this or any other consent form that you sign, and this or any other consent documentation that is used in obtaining your consent.

What if I have more questions?

If you have any questions about this study, you are encouraged to ask them now or at any time during the study, during normal business hours, by contacting either of the directors of this research project:

Name:	Professor John W. Du Bois
	Professor Mary Bucholtz
Address:	Linguistics Dept., South Hall 3520
	University of California, Santa Barbara
	Santa Barbara, California 93106, USA

Telephone: 805-893-3776
Fax: 805-893-7769
E-mail (Du Bois): dubois@linguistics.ucsb.edu
E-mail (Bucholtz): bucholtz@linguistics.ucsb.edu

If you have any questions about your rights as a research participant you may contact:

Kathy Graham, Coordinator
Human Subjects Committee
Office of Research
University of California
Santa Barbara, California 93106, USA

Telephone: 805-893-3807
E-mail: graham@omni.ucsb.edu

Part III: Options

Project Title: Corpus of Everyday Language Use
Conducted by: Prof. John W. Du Bois and Prof. Mary Bucholtz

Please tell us how you wish to participate in this research study. Choose from the following options about (1) dates you are willing to be recorded; (2) types of recording you will allow; (3) levels of research access to the data; and (4) possible contact with you.

Dates you are willing to be recorded:

From: _____ to: _____ (dates)

How may we record your conversations and other activities?
(choose one)

- ☐ Audio recording
- ☐ Video and audio recording

How may we share your data?

By participating in this research, your data may be used in oral, print, and electronic media (such as books, articles, notes, handouts, lectures, multimedia presentations, CD-ROM, and DVD).

May we also use the Internet to make your data available?

- ☐ Yes
- ☐ No

How may we identify you in transcriptions and published data?
(choose one)

- ☐ Data may include my first name AND last name.
- ☐ Data may include my first name.
- ☐ Data may not include my first name OR last name.

May we contact you after the recording is completed?
(choose one)

- ☐ You may contact me after the recording to ask for help with my recording.
- ☐ You may NOT contact me after the recording.

Signed: _____ Dated: _____

Name (PRINT): _____

Appendix 12

Computer-Assisted Transcription: Working with audio recordings (using Sound Forge)

Objective

The objective of this session is to learn about what can be accomplished using audio software designed for working with sound, which we will use for doing research on spoken discourse. We will learn about how to:

- edit sound, including defining regions of sound and annotating them (using Sound Forge)
- aligning each intonation unit in a transcription with the corresponding audio (using Sound Forge)

Sound Forge

This software allows you to edit and play back sound files. It will make permanent changes in your sound files (although there is a partial undo function), so it is best to work with a **copy** of your sound file. The software is primarily designed for people who work extensively with editing sound (like musicians and sound editors who do “post-production” editing for film and television). It is useful to linguists because it is very fast and very capable in dealing with very large audio files (and video too), and because it is ergonomically very efficient. Some of the features intended for sound editors can be put to use for the transcriber’s purposes.

Starting Sound Forge

1. To start the **Sound Forge** program, click on its icon. (If the Sound Forge program has not been installed on your computer, you will first need to install it. If you need assistance in installing Sound Forge on your computer, ask for help.)
2. From the menu, select **File/Open**. Go to the following **Ling212** (or equivalent) directory, select go to the following directory, and select an audio file from those listed:
\Data\English\Audio\SBCorpus
3. You will see a waveform of the audio file.

Setting the Options

1. You should begin by setting some of the Sound Forge options to suit your task.
2. We will want to measure time in our recordings based on seconds (e.g. number of seconds of elapsed time from the beginning of your recording file). To set the time measurement standard to seconds, select **Options/Status Format** from the main menu, then select **Seconds**.
3. Make sure the Regions List and Playlist are displayed. If they are not, from the menu select **View**, then select **Regions List**. Then select **View** again, and select **Playlist**.
4. From the menu, select **Options/Preferences/Playlist**. For Regions List display format, select **Start and length**. For Playlist display format, also select **Start and length**. Make sure the box is checked for **Sort the Regions List alphabetically**.

Using Sound Forge

Some useful functions in Sound Forge are the following:

1. You can play sound by clicking on the appropriate location in any of the three panels: in the waveform display, in the regions list, or in the playlist.
2. To **select** a portion of the **waveform**, drag the mouse across it to highlight it
3. To **play** the current highlighted sound, press the **spacebar**.
4. To highlight the sound **between two lines**, double-click anywhere between the lines.
5. To **move** the cursor to a particular **time** in your sound file, press **<CTRL>g**, and then type the time you want (in seconds) in the dialog box.
6. To **select** the sound in a particular **region**, click on the region label. (You should have the Regions List displayed to do this.) You can then **play** the region using the **spacebar**.
7. To **play** the sounds in the **Playlist**, click on any line of the playlist, and then press the **spacebar**. The cursor will move down through each line of the playlist, while playing the corresponding sound.
8. To **create** your own playlist, use the mouse to **drag items** from the **Regions List**, one at a time, over to the **Playlist**. You can arrange them to be played in any order you like, just by dragging them to a new position. (Note that modifying the Playlist in this way does not modify the original recording file.) [NOTE: Sound Forge Audio Studio lacks the Playlist function.]
9. To **jump** the **cursor** (in the waveform display) from your current position to the **following dotted line** (i.e. to the next region line or marker line), press **<CTRL><RIGHT-ARROW>**.
10. To **jump** the cursor from your current position to the **previous dotted line** (i.e. region line or marker line), press **<CTRL><LEFT-ARROW>**.
11. To **highlight** from the current cursor position to the **following dotted line**, press **<CTRL><SHIFT><RIGHT-ARROW>**.
12. To **highlight** from the current cursor position to the **previous dotted line**, press **<CTRL><SHIFT><LEFT-ARROW>**.
13. To mark a portion of the sound as a **region** (which allows you to name it, transcribe it, and return to it later), **highlight** a portion of the waveform, and then press **r**. (Just plain **r**). When the dialog box comes up, assign the region a name. If your region consists of one intonation unit, for example, you could let the first part of the region name be the beginning time for the intonation unit (which Sound Forge automatically provides), and the rest of the name can be the speaker label followed by a transcription of the words spoken.
14. To **mark** the location of a reference point in your recording, place the cursor where you want to leave a **marker**, and press **m**. (Just plain **m**). You may find it useful to use markers as temporary reference points as you are aligning the intonation units in your recording using **Regions**. Once you are done identifying the Region, you can delete your (temporary) markers if you like.
15. To **copy** an excerpt from a longer audio file into a new audio file, first **highlight** the portion you are interested in. Then copy this sound to the Windows clipboard by selecting **Edit/Copy** from the menu (or just press **<CTRL>c**). Then, from the menu select **Edit/Paste Special/Paste to New** (or just press **<CTRL>e**). This will create a new, smaller file, containing just the sound you have selected.
16. To **undo** the last action, from the menu choose **Edit/Undo** (or press **<CTRL>z**).
17. To **save** your changes (including any regions or playlists you may have created), select **File/Save** (or press **<CTRL>s**). [NOTE: In Sound Forge Audio Studio, each time you save your file, you lose the ability to undo previous changes. In the full version of Sound Forge, you can undo even after saving.]

Creating a Time-Aligned Transcription

18. After you have experimented with the existing examples in English, you should try creating regions in your own audio file. From the menu, select **File/Open**. Go to your directory, and select your audio file. (Remember to make a copy of the file first, to safeguard your data.)
19. It is useful to define a separate region for each intonation unit in your audio file. It is also useful to define a separate region for each pause. This is also a good way to measure the duration of the pauses in your data. (Be sure to do all this work in your Research Segment file; see the Appendix on “Transcribing Procedures” for discussion.)
20. If you drag each of the regions you have defined into the playlist (using the mouse), you can have Sound Forge scroll through the playlist as it plays each intonation unit in sequence.
21. Depending on how you have set up the regions in your recording, you may find that overlapping speech is played twice. By experimenting with setting up additional regions (e.g. for just the *non-overlapped* portion of an IU), you can find a way to have your Regions play back your recording in the right sequence, with your transcription automatically scrolling along with the playback.
22. Note that Sound Forge can be used for editing the audio portion of video files, not just audio files. (It shows the video as it plays, too.)

Digitizing your recording

You can use Sound Forge to digitize your audio recordings, that is, create a computer file which you can use for listening to and transcribing your recording.

Connect your recording device to sound card of your computer. From the menu, choose **Special/Transport/Record**. Choose the following settings: **44,100 Hz, 16 bit, Stereo**.

Burn a CD

You can use Sound Forge to burn an *audio* CD of your recording, which can be played in an ordinary music CD player. From the menu, select **Tools/Burn CD**, and follow the steps.

Of course, you can also use standard CD-burner software to make a *computer* CD of the computer file (e.g. in .WAV format) of your audio recording.

Further Information

For additional information on Sound Forge—which can do many more things than this document describes—you may wish to consult the book **Sound Forge 6.0 Power** (a copy is available in the Linguistics Lab). Most of the information in this book applies just as well to more recent editions of Sound Forge, at least for the basic features that linguists and other discourse researchers are likely to use.

[Rev. 26-Sep-2009]

Appendix 13

VoiceWalker: A discourse transcription utility

Introduction

VoiceWalker is designed to facilitate the transcription of recorded discourse by giving the transcriber efficient control over how sound is are played back.

No longer is it necessary to choose between transcribing on a high fidelity tape deck for sound quality and audibility, or on a specialized foot-pedal transcription machine for convenience of playback control. VoiceWalker combines high-quality stereo sound provided by your computer with far more precise and flexible control over audio playback than even a dedicated transcription tape machine can provide. All is controlled by the transcriber from the computer keyboard (and/or mouse).

Audio quality matters because it gives a degree of accuracy and audibility that allows faster, less fatiguing, and more reliable transcription. It becomes indispensable for deciphering rapid speech, and especially for recognizing words when two speakers talk simultaneously. Precise playback control matters because it allows the transcriber to repeat identically any specified section, as many times as necessary to get it accurately transcribed. With VoiceWalker all key playback functions can be controlled by either the mouse or the keyboard (with function keys). This means the transcriber's hands never have to leave the keyboard, facilitating use of a word processor for doing the transcription.

The simplest way to begin learning to use VoiceWalker is to just try it out: click on all the buttons and see what happens. Just about all the buttons (or function keys) give immediate auditory feedback, so you can easily tell what function you've invoked. And because VoiceWalker never modifies your audio or video file, you can do no harm with it. For a quick introduction that will have you transcribing in a few minutes, see the "Quick Start" section below (§4).

After you have played with the software for a while, you may find it useful to review this documentation, in order to get more out of the software. To give a more detailed picture of what VoiceWalker can do, the rest of this document provides a description of all VoiceWalker functions and options.

What You Need

Computer

To use VoiceWalker you need a computer with a sound card, speakers or headphones to hear the audio, and Windows. (VoiceWalker works with Windows XP, 95/98/Me, 3.1, 2000, NT – almost any version of Windows.) VoiceWalker requires a modest amount of hard disk space on your computer for the program itself (about 4 megabytes). In addition, you need sufficient space on your computer's hard disk (or other computer storage medium, such as CD-ROM, CD-R, CD-RW, DVD, etc.) to accommodate the audio and/or video file(s) you want to transcribe.

Recording Files

VoiceWalker works with audio and video recordings in the form of files on your computer. (Note that as video formats change over time, VoiceWalker may not work with some newer versions of video formats.) So you will need to have your recordings in a digital format appropriate for computer use. For audio, this means audio that has either been digitized for use on a computer (in Microsoft .WAV format), or was

originally recorded in a digital format. For video, this means video in one of the supported video formats (e.g. QuickTime), or video originally recorded digitally (e.g. mini-DV). If the recording you want to work with is not yet digitized in a computer format (for example, if it was originally recorded on a standard cassette audiotape or videotape), this can be easily remedied—see the section below on “Digitizing Your Recordings” (§9).

Getting Started

Installing VoiceWalker

To install VoiceWalker, you will need to download the program from the web, expand the file you have downloaded using WinZip or a similar program, and start the Setup process. The details are as follows:

1. Create a temporary folder on your hard drive.
2. Use your web browser to go to the following web address (URL):
<http://www.linguistics.ucsb.edu/projects/transcription/voicewalker.html>
3. Choose which version of VoiceWalker to download: 1.0a or 2.0. (Depending on which version of Windows you have, you may find that version 1.0a or 2.0 works better for you.)
4. Click to download the VoiceWalker software (which will be in the form of a zip file).
5. When the Windows dialog box appears, choose the option to **Save** the file. (Don’t Open the file, i.e. don’t run it remotely.)
6. Use Winzip (or a similar program) to extract the three files from the vwalker.zip archive, placing them in your temporary folder. (Try double-clicking on the vwalker.zip file, which may take care of this step automatically.)
7. Go to your temporary folder and locate the file called setup.exe. Double-click on this file.
8. The installation routine will begin. Follow the instructions given.
9. Finally, delete your temporary folder (along with its contents).
10. You should now find that there is an icon for VoiceWalker on your “Start” menu or “Programs” listing. Double-click on the VoiceWalker icon to start the program.

Selecting a File

To start using VoiceWalker, tell it which file on your computer you want to listen to or view. From the menu, select **File**, then **Open**, then specify the file you need, using the standard Windows dialog box that appears. (You may need to try clicking on a drive letter or folder name to get the files you want to display.)

Quick start: Transcribing with VoiceWalker

VoiceWalker is actually very simple to use. And because it can do no harm—it only plays back your recording, it never modifies it—the best way to learn what it can do is to simply start using it. This section describes a few simple steps that will allow you to immediately begin using VoiceWalker in coordination with your word processor to transcribe your recordings.

1. Start VoiceWalker by double-clicking on the icon. The program will occupy just a small portion of your computer screen. Drag it to the upper part of your computer screen, near the top.
2. In VoiceWalker, select **File** from the menu, and tell VoiceWalker where to find the audio or video recording that you want to transcribe.
3. Now start your preferred word processor (e.g. Microsoft Word, OpenOffice.org Writer, WordPad, etc.). Resize its window so that the window fills the lower part of your computer screen (i.e. all but the top 20% or so of your screen).
4. You should now have two windows visible: one for VoiceWalker and one for your word

processor, with neither overlapping the other.

5. In VoiceWalker, click on the Walk button (the first button on the left). VoiceWalker will start playing your recording in bite-size chunks, as it “walks” through the recording one “step” at a time.
6. Now switch to your word processor, by clicking on its window. (Or better yet, use the keyboard combination **Alt-Tab** to quickly switch between your two windows.)
7. Go ahead and transcribe what you hear (and what you see, if you're using video).
8. If you need to go back a few seconds in your recording, in order to catch something you've missed, switch back to VoiceWalker, using your mouse or the keyboard combination **Alt-Tab**. Then use the “Step Back” function (or whatever other VoiceWalker function you may need).
9. Once you've learned what all the buttons do in VoiceWalker, by clicking on them with your mouse, you may find it worthwhile to memorize a few keyboard shortcuts in VoiceWalker—especially the function keys, listed below in §7. This makes the transcribing process easier and more efficient, because you can keep your hands on the keyboard the whole time.
10. Don't forget to save your transcription often, using your word processor.

VoiceWalker Functions

The Walk (F5)

The most distinctive feature in VoiceWalker for controlling the playback of recorded sound is called the Walk. The Walk function (F5) plays the recording in manageable chunks so that the transcriber can concentrate on transcribing, as it automatically “Walks” through the recording one “Step” at a time. It plays a brief sound bite consisting of the first four seconds of the recording (one Step), and repeats this portion of audio (and video if applicable) several times to allow the user to transcribe it. Then it steps forward slightly, beginning the second Step about one second after the first. It plays this new four-second chunk of the recording several times, and then moves on to the third Step. Because each new Step repeats part of the material from the previous Step, the transcriber always has enough familiar context to know where s/he is in the recording. And because the Walk is entirely automatic, it leaves the user's hands free to transcribe using his/her preferred word processor in a separate window. (See section below on “Using VoiceWalker with a Word Processor”.)

The Walk sounds to the transcriber something like the following (assuming the number of “loop” repetitions has been set to two):

It's an interesting thing though, because um, I'm wondering, ... you know

It's an interesting thing though, because um, I'm wondering, ... you know

thing though, because um, I'm wondering, ... you know, since it is such
thing though, because um, I'm wondering, ... you know, since it is such

I'm wondering, ... you know, since it is such ...

I'm wondering, ... you know, since it is such ...

... you know, since it is such, ... um,

... you know, since it is such, ... um,

it is such, ... um, ... it's restr

it is such, ... um, ... it's restr

... um, ... it's restricted, you have to go through Me

... um, ... it's restricted, you have to go through Me

... it's restricted, you have to go through Mexico. ...

... it's restricted, you have to go through Mexico. ...

it's restricted, you have to go through Mexico. ... (TSK) No. I don't

it's restricted, you have to go through Mexico. ... (TSK) No. I don't

you have to go through Mexico. ... (TSK) No. I don't think so. No.

you have to go through Mexico. ... (TSK) No. I don't think so. No.

For some transcription purposes, it is enough to push the Walk button once, after which one can spend several minutes at a time transcribing in one's word processor without having to touch another button in VoiceWalker.

Steps

The Walk is composed of Steps. Each Step represents a brief stretch of the recording (by default, lasting exactly four seconds) to be played back. Each Step has a specific beginning time and a specific ending time. Steps are used in three playback functions (in addition to the Walk function), as follows:

Play Current Step (F7). This function plays one Step of the recording, at the current location within the file. Normally this means playing back the most recently played four seconds of the recording. This function can be repeated as often as necessary for transcription purposes, and each time it will play exactly the same four seconds of the recording. You can play one Step, stop to transcribe or to think about what you have heard or seen, and then play precisely the same Step again whenever you are ready. For an explanation of the time displayed when you press the various Step and other playback buttons, see the section below on "Current Time".

Step Forward (F8). This function shifts the Step forward slightly, allowing the transcriber to progress forward at a measured pace through the recording. It shifts forward by one "Offset" (by default, this is a distance of 1.1 seconds) and then plays the Step that begins from this new position in the recording. For example, if the Step just played was from 10 through 14 seconds, the Step Forward button will play from 11.1 through 15.1 seconds.

Step Back (F6). This function shifts the current Step back slightly, allowing the transcriber to reposition to an earlier point in the recording, or to review again a section of the recording that went by too quickly. It shifts back by one Offset (1.1 seconds) and then plays the Step found at this new position in the recording. For example, if the Step just played was from 10 through 14 seconds, the Step Back button will play from 8.9 through 12.9 seconds.

If you think about it, the Walk is really just a combination of the Play Current Step function with the Step Forward function. You can use the Step Forward and Play Current Step functions to move forward through a recording entirely at your own pace, or you can use the Walk function as a kind of automatic pilot to do the same thing without the need to keep pushing buttons, allowing you to concentrate on transcribing.

Notice that once you have begun a Walk, you can still use the Step Forward and Step Back buttons without stopping the “automatic pilot” function of the Walk. For example, you may find that one of the Steps went by too quickly for you to transcribe it, so you need to go back and play another repetition of it. Pressing the Step Back button will take you back to where you want to be. (You can press it several times if you need to move back more than one Offset.) If the Walk is in progress when you do this, it will continue automatically. If no Walk is in progress, the Step Back function will just play the specified Step and then stop. The same applies to the Step Forward function: you can use it to move forward by one Offset during an ongoing Walk, without interrupting the automatic Walk.

Looping (F2)

Sometimes it is useful to repeat the same part of the recording several times over. The Loop function lets you do this automatically, for any of the various Step functions. For example, if you set the number of Loops to four, pressing Play Current Step will automatically give you four repetitions of the current Step. Similarly, the Walk will repeat each Step four times before going on to the next Step. The Step Forward or Step Back functions shift forward or back as usual, and then play four repetitions of the relevant Step. The default number of Loop Repetitions is set to two. You can change this number by clicking the arrows under the Loop display. And you can turn Looping on and off by clicking the Loop Toggle button (or press F2).

Looping only operates in conjunction with the various Step functions (Walk, Play Current Step, Step Forward, Step Back). It has no effect on the Play or Jump functions (described below).

Play (F4)

The Play function plays continuously beginning from the present location in the recording as displayed in the Current Time window. It works like the Play button on any ordinary audio or video playback machine.

Stop (F3)

The Stop function stops all playback. It works like the Stop button on any ordinary audio or video playback machine. It doesn't matter whether playback was initiated by the Play button, or the Walk button, or any of the other functions that play the recording: the Stop button will stop it.

Jump

The Jump functions are designed to move around to different locations in the recording a little faster than the Step Back and Step Forward functions.

Jump Forward moves forward one entire Step (not one Offset) from the Current Time position in the recording, and then plays the Step located at the new position. For example, if the Current Time is at now 10 seconds, pressing Jump Forward will move the Current Time to 14 seconds, and then play the Step that ends at 14 seconds.

Jump Back moves back one entire Step from the Current Time position in the recording, and then plays the Step located at the new position. For example, if the Current Time is now at 10 seconds, pressing Jump Back will move the Current Time to 6 seconds, and then play the Step that ends at 6 seconds.

Because the Jump buttons always begin playing the recording as soon as you move to the new position, you can use Jump Forward and Jump Back to quickly find the place in the file that you are interested in, pressing them as often and as rapidly as necessary to move in the direction you wish.

Current Time

When the Play button is pressed, the Current Time window (above the Time Slider) shows the time of the

recording currently being played, in the format 00:00:00:00, which represents HOURS:MINUTES:SECONDS:HUNDREDTHS. The Time Slider bar moves along automatically during playback, from left to right, to graphically represent the location, relative to the whole file, of the time currently being played.

Note that whenever a Step is being played, the time displayed as the Current Time always represents the *end* of the current Step. This is *near* the actual time that is currently being played -- that is, within one Step's distance of the actual time -- but is not identical to it. During the period that a given Step is playing, the Current Time display does not change. As soon as you begin each new Step (whether by pressing Walk, Step Forward, Step Back, or a Jump button), its endpoint will be displayed as the new Current Time. Thus, in Walk mode, the Current Time changes only when you progress from one Step to the next, increasing each time by the fixed increment of one Offset (by default 1.1 seconds.)

Suppose you are at the very beginning of your recording, with the Current Time displayed as 00:00:00:00. If you press Play Current Step (F7), playback will start at 0 seconds and continue up to 4 seconds, where it stops. But the Current Time display will show 4 seconds (displayed as 00:00:04:00) during the entire period. If you press Walk (F5), the Current Time display will first show 4 seconds, as it plays Step 1, then 5.1 seconds as it plays Step 2, then 6.2, 7.3, 8.4, and so on as each new Step begins. Note that when the Stop button is pushed, the Current Time window will display whatever time it displayed at the moment Stop was invoked.

You can move to any location in your recording by either of two means. You can use the mouse to drag the Time Slider bar to wherever you want it, and release it. The time corresponding to the new Time Slider position will appear in the Current Time window. Or, you can type the time you want directly into the Current Time window, using the same time format (HOURS:MINUTES:SECONDS:HUNDREDTHS). You can also just type in the number of seconds. Thus, typing 66.3 will be understood as one minute and 6.3 seconds (displayed by VoiceWalker as 00:00:06.30).

Notice that you can use the Time Slider to move to a new time even during a Walk, and the Walk will automatically continue from the new time position without interruption. You can do the same during Play function, and playback will likewise continue from the new location without stopping.

VoiceWalker will automatically remember your time location in a given audio or video file after you close the program. Next time you begin a session working on the same file, the time slider will be in the same location, so that pressing Play picks up where you left off in your previous work session.

Volume

Playback volume can be adjusted within VoiceWalker via two small buttons just to the right of the Time Slider. Click on the up-arrow and hold it down to raise the playback volume. Click and hold the down-arrow to lower the playback volume. As you do this, a small bar that appears on the status bar (if this is displayed) will become longer or shorter to show you how much the volume is being raised or lowered. The VoiceWalker volume control does not obviate the need to set your preferred playback volume within Windows. You will still want to set the appropriate overall level for your computer using the Windows controls (e.g. under Settings/Control Panel/Multimedia). We recommend you click the box for "show volume control on the taskbar", for convenience in adjusting playback volume as needed.

Time Stamp (F12)

The time stamp function is designed to help you insert time information into your transcription, if you so desire. When you press the Time Stamp button (F12), VoiceWalker records the Current Time of the audio or video playback, and copies this time into the Windows Clipboard. You can then switch into your word processor and press Paste (or Ctrl-V) to insert the information into your transcription at the relevant place.

Menus

All of VoiceWalker's functions are available from menus. Of course, usually it's easier and more efficient to click the buttons—or even better, to use the function keys (see below). VoiceWalker's main menu has five groups: File, Play, Options, Video, Help.

File Menu

The **File** menu lets you select a new file to work with at any time. Select **File**, then **Open**, and a standard Windows dialog box will appear to allow you to specify the file you want. (Or use the keyboard shortcut **Ctrl-O**.)

To see general information about the file you have selected, choose **File**, then **Info**. This displays the name and path of the file you are working with, along with various technical details about the file (sample rate, number of bits, stereo versus mono, and total running time for the file).

To quit using VoiceWalker, select **File**, then **Exit** (or press **Ctrl-Q**). Or use the standard Windows way to close the program.

Play Menu

Play. All the playback functions of VoiceWalker can be accessed from the **Play** menu. Select **Play**, then choose the appropriate playback function. But there is not much reason to use this menu, as it is far more convenient to use the various playback buttons, or better still, the function keys (see below).

Go to time (F11) is a very useful function that can be accessed from this menu (or via the F11 function key). Pressing the **Go to Time** button brings up a dialog box which allows the user to type in a time; then the Current Time will shift to the specified location in the recording. The time is normally specified in the format 00:00:00:00, representing HOURS:MINUTES:SECONDS:HUNDREDTHS. (Note that this time-setting function is also available on the main VoiceWalker panel, in the Time Slider box.) For convenience you can also just type in the number of seconds (for example, typing 127 will be understood as 2 minutes and 7 seconds, displayed as 00:02:07:00).

Options Menu

The **Options** menu allows users to specify various settings. Most transcribers seem to use the factory default settings given by VoiceWalker, without ever needing to modify them. If this is all you need, you can ignore this section. But the settings can all be modified, if needed for special purposes. Any modified Options settings will be in effect during the current session using VoiceWalker with a given audio or video file. All Options settings are reset to the factory defaults when you leave the program, or when you open a new file.

Note that all times for the settings are given in the same format: 00:00:00:00, representing HOURS:MINUTES:SECONDS:HUNDREDTHS.

Step Duration lets you specify how long a stretch of recording to play for each Step. The default is four seconds (00:00:04:00). This seems to be about the right amount of sound to allow the transcriber to take in what is being said, as well as to provide enough context to see the relation to previously transcribed utterances.

Offset Time lets you specify how far to “offset” the beginning of each Forward Step from the beginning of the previous Step. The default is 1.1 seconds (00:00:01:10). This in effect specifies how much new information the transcriber will take in with each new (Forward) Step. Note that because the default Offset Time is significantly smaller than the default Step Duration, there is a substantial portion of the playback that is shared between one Step and the next. This is helpful in allowing the transcriber to retain a sense of where s/he is in relation to previously heard portions of the recording.

Rest Time specifies the amount of time to pause between Loops during Looping, or between

Steps during a Walk. In other words, it represents a brief silence inserted between successive repetitions or successive Steps (like a “rest” in music). This avoids the overly insistent, breathless character of immediate repetitions with no pause in between them, which can be stressful for transcribers. The default Rest Time is 0.6 seconds (00:00:00:60).

Loop Repetitions specifies the number of times to repeat each Step, as occurs in the various Step functions (Walk, Play Current Step, Step Forward, Step Back). The default number of Loops is two. This is the one option that transcribers do often like to modify, but there’s no need to use this menu to do so: it can be directly accessed on the main VoiceWalker panel, using the arrows under the box that displays the number of Loop repetitions.

Status Bar lets you choose whether to display or hide the Status Bar help information. To find out what any given button does in VoiceWalker, place your cursor over the button. A brief description of the button’s function will be displayed on the Status Bar, at the bottom of the VoiceWalker window (see **Help** below). Selecting this menu item will toggle display of the Status Bar on or off.

Video

The Video menu selection lets you specify what size of window will be created by VoiceWalker for viewing video playback. (Note that as video formats change over time, VoiceWalker may not work with some newer versions of video formats.)

BookMarks

VoiceWalker allows you to “bookmark” up to four locations in the audio or video file you are working on. From the Options menu, select View Bookmarks (or press Ctrl-B), and the VoiceWalker window will enlarge to reveal a display of four bookmarks. Each bookmark has three buttons. As you play back your file, press the first button to set the beginning of the bookmarked location. When the playback reaches the point where you want your selection to end, press the third button to mark the endpoint of the bookmarked location. To play the section you have just bookmarked, press the middle button (“Play Bookmark”).

You can use this function to compare two or more different portions of your recording. Set up the bookmarks you want, then click in turn on the “play” button for each bookmark to make your comparison.

VoiceWalker will automatically remember the locations that you have bookmarked in a given audio or video file after you close the program. The next time you begin a session working with the same file, your previous bookmarks will be available for your use. If you want to clear the bookmarks (erase them), you must specify Clear Bookmarks from the Options menu.

Function Keys

By far the most convenient and efficient way to use VoiceWalker is not to use the mouse at all (or to use it minimally), but to instead control all playback functions via the keyboard, using the function keys. This allows the transcriber using a word processor to keep his or her hands on the keyboard as much as possible. The following list presents a summary of the most commonly used functions each of which can be accessed via a particular function key. Note that all the functions listed here are also available by via the mouse, by clicking the button icons on the main VoiceWalker panel. A full description of what each function listed here actually does is given in the relevant section above.

Key	Function
F1	Help
F2	Looping On/Off
F3	Stop
F4	Play
F5	Walk
F6	Step Back
F7	Play Current Step
F8	Step Forward
(F9)*	Jump Back
(F10)*	Jump Forward
F11	Go to Time
F12	Time Stamp

* not currently implemented

Using VoiceWalker with a Word Processor

Although VoiceWalker can be used just to play back an audio or video file, more often it is used for transcribing. VoiceWalker has been designed to work in conjunction with any word processor you like: Microsoft Word, OpenOffice.org Writer, WordPerfect, WordPad, or any other Windows program you prefer to use. The following discussion repeats much of the information given above in the “Quick Start” section (§4), with a little more detail.

First open VoiceWalker in one window, then open your favorite word processor in a separate window. Resize each window so that you can see both programs at once on your screen. (This should be easy because VoiceWalker takes up very little space on your screen.) You can now control both programs, using the functions of VoiceWalker for audio and video playback, and using your word processor for typing and editing the transcription.

For example, first press the “Walk” button in VoiceWalker to put the playback on “automatic pilot”. Then click on your word processor window, and type the words you are hearing. VoiceWalker will continue to play back sound and video “in the background”, even though your active window is now the word processor. Depending on your transcribing style, you may be able to transcribe for several minutes in this way without having to push another button in VoiceWalker.

But most people will want to use the VoiceWalker controls more often: for example, to repeat the playback of a section that went by too fast to transcribe the first time. To do this you will have to switch between the two programs. The quickest way to do this is to use the keyboard, using the keyboard combination Alt-Tab to switch between the two programs. (We recommend that you close all other windows except these two, leaving just the VoiceWalker and word processor windows open. That way, a single press of Alt-Tab will switch you reliably from one program to the other.)

For example, assume you are typing in your word processor, while VoiceWalker is playing in “Walk” mode in the background. You want VoiceWalker to “Step Back” in the recording because you missed a few words. To do this, type the Alt-Tab key combination followed by the F6 function key. This sequence will first shift you into the VoiceWalker window, and then perform the “Step Back” function. Then type Alt-Tab a second time, and you are back in your word processor, ready to continue typing. You never have to raise your hands from the keyboard. After a little while the three-stroke sequence “Alt-Tab ... F6 ... Alt-Tab” becomes rapid and automatic. Almost every function of VoiceWalker can be accessed in this way with just a few keystrokes, where a function key is the middle part of an “Alt-Tab

sandwich.” The function keys that are probably the handiest to use in this way are the following:

F6	“Step Back”
F7	“Play Current Step”
F8	“Step Forward”

With a few simple Alt-Tab sequences memorized, you can get most of what you need from VoiceWalker without having to stop to reach for the mouse.

Digitizing Your Recordings

Audio

Creating digitized audio files, in order to transcribe them using VoiceWalker, is not much more difficult than copying from one tape recorder to another. In this case, though, the second “tape recorder” is your computer and its hard disk. In broad outline, it typically works as follows: You run an audio cable from your existing audio recorder (“Line Out” or “Phone”) to the sound card on the back of your computer (“Line In” or “Record”). On your computer, you open your sound recording software (e.g. Sony Digital Media’s Sound Forge, etc.), and press the Record button in your software. On your tape recorder, you then push the play button. You are now recording from the tape recorder onto your computer’s hard disk. When you have recorded as much as you want to record, stop the tape recorder, and “stop” the software. That’s it. You should now have a sound file that you can use with VoiceWalker.

VoiceWalker requires the Windows .WAV format, so be sure to use this format. We recommend using a “CD-quality” recording format to get very good sound quality. This means choosing a sampling rate of 44.1 kHz, 16-bits, and stereo. This format requires 10.5 megabytes of hard disk storage space per minute of sound. You can economize on disk space by going down to 22.05 kHz for the sampling rate, with still very good quality. (You could try to save still more on storage space by using a lower sampling rate or 11 kHz, and even lowering the resolution to 8-bits, but the sound quality will start to suffer significantly.) Note that stereo is very important for dealing with spontaneous conversation, because it contributes greatly to the accuracy of hearing simultaneous or overlapping speech. Stereo should be considered a necessity rather than a luxury for working with natural conversation.

Better than digitizing your recordings is to record them digitally in the first place. This becoming much easier with the rapid development of new technologies, such as portable hard disk recorders that record directly in .WAV computer format, and make the transfer to your computer’s hard disk with a standard USB computer cable (e.g. Roland Edirol R-09).

For a full description of the audio digitizing process, please refer to the documentation that comes with your multimedia computer or your sound card, or with the sound recording software that is typically included with the purchase of a computer or sound card.

Video

The easiest way to get a video recording into a computer format that is easy to work with is to make the recording in a digital video format to begin with. Currently the best way to do this for most people will be to use a mini-DV video camera, coupled with an inexpensive DV video-capture card.

The process of creating a digitized video file from a VHS videocassette or other analogue video source is roughly comparable to that described above for digitizing an audio file. However, to digitize analogue video recordings, the process tends to be more complex. It usually requires more specialized equipment, and the details may vary substantially according to the specifics of your video equipment, digitizing card, and so on. If you go this route, please refer to the documentation provided with your

equipment.

Help

To get a brief description of any button on the VoiceWalker panel, move the cursor to what you are interested in, and a description will appear in the Help Status Bar at the bottom of the VoiceWalker window. (You can choose whether to display or hide this Help Status Bar, using the Options menu as described above.)

At present, the Help menu (or F1 key) provides little in the way of substantive information. The role of Help is best served by the Status Bar, plus this manual.

About VoiceWalker

Versions

There are currently two versions of VoiceWalker, which have nearly identical functions, structure, and look-and-feel. Despite the earlier number, version 1.0a may be best for most users. Some users may need to use version 2.0, which however is not as fully tested, and may differ in some minor details from the functions as described in this document. (See “Credits” section below for further information.)

Documentation

Documentation is available at the VoiceWalker website, at:
<http://www.linguistics.ucsb.edu/projects/transcription/tools.html>

Since the documentation on the website is simply a version of the present document, you should check the revision date to see if the version you have is up-to-date (see the date at the end of this document).

Credits

VoiceWalker was created as a spin-off from the Santa Barbara Corpus of Spoken American English, at the Linguistics Department of the University of California, Santa Barbara. VoiceWalker was designed by John W. Du Bois. Programming and design refinements are by Simon H. Corston and Gary Holton, partly based on prototypes by Robert Norris. The original program (VoiceWalker 1) was written using SumTotal’s (formerly Click2Learn and Asymetrix) ToolBook II authoring software. VoiceWalker 2 was entirely recoded by Gary Holton, using different software. Documentation is by John W. Du Bois.

Copyright

VoiceWalker is Copyright ©1998 by the University of California Regents.

VoiceWalker documentation is Copyright ©2006 by John W. Du Bois.

VoiceWalker is free for noncommercial use (as shareware). Support is not available. For further information, contact:

John W. Du Bois, Director
 Santa Barbara Corpus of Spoken American English
 Linguistics Department
 University of California, Santa Barbara
 Santa Barbara, California 93106
 USA

[Rev. 25-Aug-2006]

Appendix 14

Transcriber: A Transcription-Audio Alignment Tool

Objective

The objective of this session is to learn more about what can be accomplished using computer tools designed for working with sound. We will learn about programs for aligning each intonation unit in a transcription with the corresponding audio (e.g. Transcriber). (Sound Forge can also be used for this function.)

Transcriber

Transcriber is a program that lets you play back sound, transcribe the audio, and align your transcription with the corresponding audio. The program was designed for transcribers, although the primary orientation is for transcribing broadcast speech, rather than spontaneous conversation.

After you have selected the appropriate “**Ling212**” directory, click on the **Programs** folder, then the **Transcriber** folder. Then double-click on the icon labeled **trans** (the icon looks like Aladdin’s lamp). This will start the Transcriber program.

A dialog box will appear, with the message: “Open transcription or audio file:” From the files listed, click on the file named **frint980428.trs** (this is a demo file with a transcription in French). (The program should also open the corresponding audio file. If it does not, open a file as indicated below.)

To open an audio file, from the menu select **File/Open audio file**. When the dialog box appears, go to the directory you want. For the demo, select the **demo** directory, and then select the audio file **frint980428.wav**.

To **move** the **cursor**, you can click on a location in the waveform, or click on a sentence in the transcription.

To **play** the sound starting from your cursor position, press the <TAB> key.

To **pause** the sound, press the <TAB> key again.

To **insert** a **breakpoint** (that is, a boundary between one unit of sound and the next), hit the **enter** (or **return**) key while playing the sound.

To **move** a breakpoint (as displayed in the green sections at the bottom of the Transcriber window), use your **mouse** to **drag** the line to the left or right, holding the <CTRL> key and using the **left** mouse button.

For a good description of how to use the program for transcribing, select **Help/User Guide** from the menu. (In Transcriber it’s a good idea to maximize the Help window, making it fill your computer screen, so that it doesn’t disappear on you.) Then read the section on **A typical transcription session**.

To familiarize yourself with the functions of Transcriber, try out the useful functions listed under the menu options **Signal** (for sound playback) and **Segmentation** (for marking regions of the audio file, e.g. as intonation units). You may find it easier to use Transcriber if you learn the keyboard shortcuts listed under these two menu headings, after which you will be able work more conveniently.

As an exercise, improve the transcription of the French demo example given. Insert new breakpoints in the audio file at the intonation unit boundaries. Then go back and type in the transcription in the appropriate place corresponding to the units of sound you have defined.

Sound Forge

Sound Forge can also be used for aligning transcriptions with your recording, using the Regions and Playlist functions (see the separate document devoted to Sound Forge).

[rev. 7-Sep-2004]

Appendix 15

Acoustic Analysis

Objective

The objective of this session is to learn more about what can be accomplished using computer tools designed for working with sound. We will learn about programs for:

- analyzing pitch contours and other acoustic properties of speech (PRAAT, Fujisaki)

The session is not designed to give you complete mastery over acoustic analysis tools, since this merits a course in itself. Rather, it is designed to let you gain enough experience with what they can do for you that you can follow up on your own to learn more, if you find that you are interested in the new research potential which is opened up by these and other similar tools for acoustic analysis.

PRAAT

PRAAT is a program designed for doing phonetic analysis of speech on a computer. For example, it will analyze a sound file to do a “pitch extraction”, that is, it will determine the fundamental frequency of the speech, and draw a corresponding curve. This is only one of many kinds of phonetic analysis that PRAAT is designed to do

Start **PRAAT** by double-clicking on the icon labeled **PRAAT** (the lips-and-tongue icon). (If you do not see it immediately, you may need to do the following to find it. Select the “**Linguistics 212**” (or equivalent) folder, click on the **Programs** folder, then the **PRAAT** folder. Then double-click on the PRAAT icon. This will start the PRAAT program.

When PRAAT starts, it initially appears as two separate windows on your screen.

To **open** a (stereo) audio **file**, from the menu on the left window, select:

Read/Read two sounds from stereo file

To play sound from the **left channel** of the stereo file, select **Sound left** by clicking on it.

Then click on **Play**.

To play sound from the **right channel** of the stereo file, select **Sound right** by clicking on it.

Then click on **Play**.

For further information on how to use PRAAT, click on **Sound help**. Experiment with the various functions you see.

[rev. 7-Sep-2004]

Appendix 16

Preparing a Transcription for Computer Analysis

Objective

The objective of this document is to describe how to prepare a transcription file so that you can work on it using various computer software tools, such as programs for making concordances or aligning transcriptions with sound. This document assumes that you have already made a transcription of some spoken discourse data, and that you now want to analyze it using a computer tool of some kind.

Preparing your transcription

In order to prepare for working with a transcription using computational tools, it is often necessary to first do some further preparation or checking of the transcription file you have made. This is needed to make sure that your transcription is in a format suitable for analyzing with a concordance program or other similar software (e.g. MonoConc Pro, Wordsmith, etc.). This kind of preparation may require some “cleaning up” of your transcription file, to make sure there is nothing in it that will confuse the concordance software (or other computer tool).

Transcription vs. Coding

One way to prepare a transcription for computer analysis is to remove any features that are not transcription per se. In general, transcription (as opposed to coding) properly consists primarily of symbols corresponding to actual, audible or visible events which took place in the conversation or other speech event you are transcribing. In other words, the transcription contains mostly words spoken, vocal noises, gestures, actions by participants, salient environmental events, and so on. In contrast, background information such as a description of the setting, a title, or extensive commentary would normally not be considered to be transcription as such. Of course, this information is very important and should accompany the transcription, in some form, such as a linked database, or comment lines which can be included or removed as necessary for different analytical purposes. But it should be distinguished from the transcription per se, lest computer analysis of this background information produce a nonsensical result.

Nor does transcription per se consist of coding or analytical information such as part-of-speech tagging, syntactic bracketing, and so on. (Of course there are other valid approaches to computational analysis of discourse data, in which the insertion of lots of analytical codes into a transcription file is common practice.)

For doing the kind of file preparation we are aiming for in this session (e.g. for concordance-making), you will normally want to have a more or less “straight transcription”, that is, a transcription which is not encumbered by lots of coding symbols or analytical notations. Note that brief comments embedded in the transcription are usually not a problem. For example, a comment that is included along with some words spoken as part of a single intonation unit, is not something that needs to be avoided or edited out.

In sum, transcription means representing what you hear and see on your recording. The preparation we are doing here mainly involves stripping away everything that is not true transcription.

Procedures

12. If you have not already done so, make sure that all your one-minute increments of transcription are consolidated in a single computer file.
13. It is not so important that this file should include all your minutes of transcription. What is important is that you should do the preparation on whatever transcription data you have ready.

14. Make a copy of your transcription file first. You will be making a few changes to this file in preparation for analyzing it, so it is wise to preserve your original file separately.
15. Remove any lines containing only header information, like the title of the transcription, name of the transcriber, and so on.
16. Remove any lines of text containing nothing but a comment, if possible. Short comments (e.g. those written in double parentheses, etc.) are okay, when they are embedded in (or at the end of) a line with an intonation unit containing words spoken.
17. Remove any lines which consist exclusively of interlinear glosses, or other analytical coding.
18. Remove all blank lines. There should be no blank lines (lines containing no characters) at all in your transcription file. Don't use double spacing. Check the beginning and end of your file to eliminate blank lines there, too.
19. If you have special characters (e.g. Unicode characters) in your transcription, you may need to experiment a bit to find which format best preserves the work you have done. (While the future of corpus and linguistic research may be a Unicode future, the reality is that the future is not quite here yet, at least for some current software. (For example, the current version of the concordance program MonoConc Pro [version 2.2] unfortunately does not handle Unicode characters.) Use the search-and-replace function to replace the IU truncation sign (em dash) with two hyphens (an alternate notation for IU truncation). Other Unicode characters (e.g. mostly the tone and arrow symbols used for intonation transcription) should be replaced with alternate notations, or simply deleted. Note that when you save your file in a plain text format like ANSI or ASCII (see below), any remaining Unicode characters in your file will be automatically replaced by some arbitrary character like a question mark, or a square box. It is better to make a more rational disposition of the Unicode characters yourself, substituting something meaningful for them.
20. When you have finished preparing your file according to the above steps, save it, using your usual format (e.g. if you are using OpenOffice.org Writer, this would be the **.odt** format; if you are using Microsoft Word, this would be the **.doc** format).
21. Now save your file again, this time in a plain text format, such as ANSI or ASCII. This format is required by many concordance programs and other analysis software. This version of the transcription will be the one you will use to make your concordances or other analysis.
OpenOffice.org. If you are using OpenOffice.org Writer, save your file as a plain text, as follows: From the main menu select **"File/Save as"**. When the dialog box comes up, look at the very bottom where it says **"Save as type."** Scroll down in this box until you find **"Text (.txt)"**, and select it. Click **Save**. Your file will be saved with the file extension **.txt**.
Microsoft Word. To save your file in plain text format using Microsoft Word, select **File** from the menu, then **Save as**. When the dialog box comes up, look at the very bottom where it says **"Save as type."** Scroll down in this box until you find the **"Text with Layout (*.ans)"** format, and select it. Then click on **Save**. Your file will be saved with the file extension **.ans**.
 (Alternative file formats that can be used are **"MS-DOS Text with Layout (*.asc)"** or **"Plain Text (*.txt)"**.)
22. Give your file a name that makes clear what is in it. Use the naming conventions specified in the Appendix on "Transcription Format" under the heading "Filenames."
23. Now you have a file that you can use with a concordance program, to do research on questions of discourse and grammar.
24. Preparing your file in this way shouldn't take long, normally less than an hour.
25. Copy your prepared file onto the appropriate directory in the Linguistics Lab network. Be sure to do this **BEFORE** the class session in which we will be using the files.

26. Note that any files which are not prepared according to the above guidelines, or which are not copied into the appropriate directory in advance of the class tutorial session, may be unusable or unavailable for use in the concordance tutorial.

[rev.11-Oct-2005]

Appendix 17

Using a Concordance for Discourse Research

Objective

The primary objectives of this tutorial are:

- ♦ to show what a concordance program is, and what it can do for your discourse research—especially for exploring connections and correlations between patterns in spoken discourse and patterns in grammar and lexicon
- ♦ to show how to take advantage of the systematic representation in discourse transcriptions of such discourse features such as turn-taking, simultaneous speech, vocalizations, pauses, and prosody
- ♦ to introduce users to the specific research potential of the type of discourse transcription information found in the Santa Barbara Corpus of Spoken American English, and in other similar transcriptions

Background

A concordance program lets you search through your discourse data for all the instances of a given word, and then see the surrounding context for each instance that is found. The result is called a key-word-in-context concordance. You type in the word you are interested in (the key word), and in a few seconds you have a list of neatly lined up examples of that word as found in your data. You can then sort this list of words in a variety of ways, in order to get a clearer picture of how it relates to its context. No matter how you sort the words, their context remains available to you.

You can also search for patterns of words, using various "wild-card" symbols to represent the combinations of words you are interested in. With a little cleverness in designing your search (and assuming your discourse data are in a systematic discourse transcription format), you can also look for such discourse features as turn completions, discourse particles or vocal noises word that typically co-occur with turn beginnings, interruptions and simultaneous speech, and so on.

In addition, you can get certain statistics on, for example, what other words most frequently collocate with your key word, and how many instances of each collocation occur in your data.

You can print out your search results, or save them in a file to be printed out with your favorite word processor. Or you can just select specific examples you would like to save (e.g. to include in a paper you are writing), and copy them directly into a separate window that you open for your word processor.

MonoConc is such a key-word-in-context program, designed for Windows. It is easy to use, fast, and intuitively organized, and it is available for use in the UCSB Linguistics Lab. Other comparable Windows concordance software to consider is WordSmith (from Oxford University Computing) and ShoeBox (from the Summer Institute of Linguistics). (ShoeBox also allows interlinear grammatical analysis and glossing; facilitates automatic glossing; and can sort data into "dictionary" format.) (For more information about MonoConc and corpus linguistics in general, contact Michael Barlow at athel@aol.net, or <http://www.ruf.rice.edu/~barlow/corpus.html>, or <http://www.nol.net/~athel/athel.html>.)

General Instructions

From the Windows **Start** menu, start MonoConc by selecting **Programs**, then **MonoConc**. Or if you have an icon for MonoConc on your desktop, you can double-click on that. (To save the discourse examples you find, you may also want to open a separate window for WordPad, or for your favorite word

processor.) Maximize the MonoConc window so that it fills your screen, to allow you to see more examples at once, and more discourse context.

We will be working with transcriptions from the Santa Barbara Corpus of Spoken American English (SBCSAE or SBCorpus) as our discourse data. To access this data, do the following. Within the MonoConc window, load a corpus by telling MonoConc what directory you want, and which files, as follows: From the menu, select **File**, then **Load Corpus**. A dialog box will appear. Make sure the "List files of type" window says "All Files (*.*)" Specify the appropriate drive for directory you want, by clicking on the drop-down arrow for "Drives", and highlighting the appropriate drive. Specify the appropriate folder, by double-clicking on the folder(s) listed in the folder tree window. To use the conversational (and other) transcriptions of the Corpus of Spoken American English available on the UCSB Linguistics Department network, , this would probably be something like **\Data\English\SBCorpus\TextOnly**. Select all the files in this directory, by highlighting them (drag over the listed files with the mouse). Select **OK**.

One or more text files will now be displayed on your screen, each in a separate window. You can look over the transcriptions if you like, but you don't have to. To clear your screen so you can concentrate on concordancing, select **Corpus Text/Hide All**. Even though the file (or files) will no longer be visible, they will still be used in whatever concordance searches you do during this session.

To exert greater control over your searches, you may want to modify the various options that MonoConc offers under the headings of **Concordance/Search Options** and **Frequency/Frequency Options**. (For more information, see the tutorial section on Setup Options.) But for getting started, the MonoConc settings in the Linguistics Lab should be sufficient. (I have set up the various options so as to facilitate working with spoken discourse transcriptions.)

Note: In MonoConc, *spaces matter*. MonoConc treats each space as a "word" boundary: a word is thus any string of symbols bounded by spaces. In the examples of concordance searches given below, you are often asked to type several words or symbols; always type the spaces exactly as they appear on the handout. (When you see a space in an example, it is always exactly one space: there are no cases, in this tutorial, where you will need to type in two spaces in a row.)

You are now ready to start concordancing!

Words and affixes in discourse

The questions which are introduced below are intended to be suggestive of the kinds of research questions you can use a concordance program to ask. With these models to start you off, there are unlimited ways you can use your imagination to extend the questions that you ask in your own research.

When is the discourse marker 'anyway' is used in discourse?

Select **Concordance**, then **Search**. Where it says "enter pattern to search for", type the word **anyway**. What do you see? Looking over the data, what generalizations can you make?

Pick a line that looks interesting to you, and click on this line with your mouse. Notice that a larger amount of context for this line immediately appears in the upper half of the screen. To enlarge this discourse context window so you can see more of the prior (or following) discourse context on your screen, use your mouse to click on the horizontal line separating the two halves of the screen, and drag this line down. Use the scroll bars to look over the discourse context for your example (up to several pages of context).

Try clicking on several other lines. Notice that for each example you click on, the relevant context appears in the upper screen, and the keyword is highlighted in color. Also, each time you select a line, the name of the computer file which that example comes from is displayed on the status line (i.e. in the

lower left corner of the MonoConc window). The total number of words included in your corpus for this search is also indicated on status line, towards the right.

If you find a specific portion of one of these discourse examples that is of particular interest to you, you can cut it out and save it by first highlighting the relevant portion with your mouse, and then copying it into your word processor. (Notice that the tab symbol, which MonoConc displays as a black rectangle, is properly displayed as a tab once you paste the text into your word processor.)

When you are ready to go on to the next search, you do *not* have to close the window containing the previous search results. Just leave it on your screen, so you will be able to compare it with later search results if you like. You can have as many search result windows on your screen as you want, as they do not seem to affect MonoConc's performance.

When is the word 'well' is used in discourse?

Select **Concordance/Search**, and type as the search string: **well**. (Just the word **well**, don't type the punctuation.) What do you see? Use the scroll bars to scroll through the data. What patterns can you identify?

Now examine the data in a more systematic way. Sort the search results you just got for 'well' according to the first word that appears to the left of the word 'well', by selecting **Sort/1st Left**. Scroll through the data. What do you notice now? How does the sorting make the discourse patterning clearer?

Now **Sort** by **1st** (word to the) **Right**. What do you notice? How informative is this sort order compared to '1st Left', for this word?

Now let's look at a more general overview of the discourse patterns involving this word. Select **Collocation**, then **Collocate Frequency**. How does this summarize more systematically what you already saw informally in your sorts of the data? What can you see here that you might not have noticed with just a sort?

What would a more detailed view of the most frequent collocations involving the word 'well' tell me about how it is used in discourse?

Looking over the Collocate Frequency from the previous search, you notice that some words to the immediate left and right of the word 'well' occur with much higher frequency than others. You may also get a sense that the word 'well' has several different functions. To some extent these can be teased apart by looking at in more detail at the specific collocations that 'well' participates in. To do this, decide on three or four words from among the top ten collocations listed under '1st Left', and do a follow-up search for each one. For example, select **Concordance/Search**, then type in the phrase **yeah well**. Next try **okay well**. Then do similar searches for several of the other top ten collocates, including words to the right (e.g. **well I**). How does this let you zero in on the function(s) of 'well'?

What if I want to find out what words collocate with my original collocations?

You can continue the cycle by asking for collocate frequency data on your collocates themselves. For example, first do a **Concordance/Search** for **well I**. Then select **Frequency/Collocate Frequency** (or type **CTRL+f**). What patterns of collocation do you see here? What generalization can you make about the verbs involved? You can also **Sort** the search results by **1st Right** and scroll through the examples to get a more detailed picture. (Always sort the discourse data by making your search results window active. Don't try to sort the statistics window -- it doesn't make much sense to sort these statistics, and MonoConc won't let you do it.)

Of course you could continue this cycle still further by selecting **Concordance/Search** and specifying three words as the string to search for. But as you add more words the number of "hits" (examples of

your search string that are found in your data) will decrease rapidly, and you may not find enough examples to get a generalization. This of course will depend to a great extent on the size of your corpus.

What patterns do you find in the way the noun 'reason' is used?

Search for reason. What do you see? But notice that this leaves out plurals. To remedy this we can make use of MonoConc's "wild-card" symbols. For example, the symbol * represents any string of symbols. (There are also wild-card symbols that allow you to search for exactly one character; for one or zero characters; and for several words, e.g. between 2 and 5 words.)

So **Search for reason***. What do you get? Notice that the use of a wild card means that your search picks up several different search terms. To see more clearly how the various search terms pattern, **Sort by Search Term**. Scroll through the results. What do you see?

To weed out unwanted items, such as (in this case) verb forms, highlight the lines in question with the mouse, and then use **Control-d** to delete them from the listing.

Now examine the remaining data in a more systematic way. **Sort by 1st Left**. What do you notice? **Sort by 1st Right**. What do you notice?

Do **Frequency/Collocate Frequency**. How does this summarize what you already saw in your sorts of the data?

Notice that the statistics provided here cover just the words that remained in your window at the time you initiated the Collocate Frequency calculation. The irrelevant items that you just deleted using **Control-d** are not included, which is of course a good thing. For the most accurate picture of your data, you should do any necessary weeding out of irrelevant examples *before* you calculate your statistics.

When is the verb 'come', in all its forms, used in discourse?

You could get many of the verb forms by **Searching for com***. Try this. Then **Sort by Search Term**. What do you see?

You could get rid of the many unwanted items by highlighting them with the mouse, then deleting using **Control-d**. But notice that this is a bit of a hassle, and you still wouldn't have the past tense form 'came'. Instead, you can specify individually each form of the verb 'come' that you *do* want, using the **Append Search** option, as follows:

Select **Concordance/Search**, and search for **come**. Then, while you are still in this window, select **Concordance/Advanced Search** (or **CTRL+a**), and check the box for **Append Search**. Now search for **comes**. Notice that this *adds* the instances of the word 'comes' to the end of your previous listing: it puts both sets of examples into the same window, rather than making a new window. For present purposes, this is useful.

Now do another Append Search, by selecting **Concordance/Advanced Search** (or **CTRL+a**), and again checking the box for **Append Search**. (You have to check this box each time you do an Append Search.) Now search for **came**. Again, the hits for 'came' will accumulate in the same window as before. Finally, do an **Append Search** for **coming**.

Now that you have all of the verb forms of 'come' in one listing (in one window), examine the data in a more systematic way. **Sort by 1st Left**. What do you notice? **Sort by 1st Right**. What do you notice? Do **Frequency/Collocate Frequency**. How does this summarize what you already saw in your sorts of the data? How does having all the verb forms in one window provide a clearer picture than if the data were separated into four separate windows?

When is the discourse particle 'um' used?

Search for um. What do you see? **Sort by 1st Right**. **Sort by 1st Left**. What do you see?

But this leaves out lots and lots of cases, because of the different ways **um** is uttered -- for example, this word is often lengthened. To fix this we can use the wild card that allows for one or zero characters, that is, %, plus the wild card for an unlimited number of characters (including zero), i.e. *. (You may need to check to make sure that these characters are indeed the current wild cards used by MonoConc, and that they are *not* listed among the "characters to treat as delimiters". Both functions can be specified by selecting **Concordance/Search Options**, and making sure the right characters appear in the appropriate boxes.)

Search for **u%%m***. Then **Sort** by **Search Term**. What's the difference? How does this change your understanding of the role that the word **um** serves in discourse?

There are likely to be a few unwanted items here, which you could weed out by hand. Or better, try searching for **u%%m%%**, then **Sort** by **Search Term**.

Turn-sensitive phenomena

What words tend to occur in turn-initial position?

Recall that the symbol for a speaker attribution label is a colon (in the conventions of Discourse Transcription 1 [DT1]), or a semi-colon (in DT2). So we can expect that this symbol should occur whenever a new speaker starts talking, because of the speaker label. We can take advantage of this fact in our searches. We'll use the wild card * to stand for any string of symbols -- in this case, it will give us any speaker's name, because these always end with a colon (in the Discourse Transcription system).

Do a **Concordance/Search** for ***:** (that is, for asterisk followed by colon, with no spaces). (This is for data transcribed according to DT1. For transcriptions in DT2, do the search with an asterisk followed immediately by semi-colon: ***;** .)

Select **Frequency/Collocate Frequency** (or type **CTRL+f**). Look over the column of the '1st (word to) Right'. This corresponds to the first word immediately following a speaker label, that is, the first word of a "turn". (Notice that this is actually a pretty crude approximation to real turns, since it does not separate out backchannels, as one could argue should be done.) What words are most frequent in turn-initial position?

To see examples of these turn-initial words in context, return to the search results (discourse data) window for ***:**. Then **Sort** by **1st Right**, and scroll to an interesting (frequent) word, as indicated in your Collocate Frequency listing.

What words typically occur in turn-final position?

View the **Collocate Frequency** for speaker labels as above. Look over the column headed '1st Left'. Why does this give you turn-final words? What words are the most frequent in this position? To see examples in context, first go back to the search results (discourse data) window above (i.e. for ***:**). **Sort** by **1st Left**, then scroll to an interesting word

What words occur as single-word turns (e.g. backchannel, etc)?

Search for ***: * *:** (Remember to type the spaces just as they appear here, that is, one space between the first and second "word", and one space between the second and third "word". Because spaces count as word boundaries, MonoConc will interpret this search string as a sequence of three "words" in exactly this order: a speaker label, followed by exactly one word, followed by another speaker label. In other words, a turn of exactly one word in length.) Scroll through the words -- what patterns do you notice?

What words occur in short turns (two-word to four-word turns)?

Select **Concordance/Search Options**. Set the "**matches between**" variable to stand for 2 to 4 words. (You may need to check to make sure that the "**matches between**" symbol is **&**, and that it does not appear on the list of "Characters to treat as delimiters". You can do this at the same time, i.e. by selecting **Concordance/Search Options**.) Select **OK**.

Search for ***: & *:** What do you see?

When done, you may wish to select **Concordance/Search Options** again to reset the "**matches between**" numbers back to what they were before.

Other wild-card uses

When do compound verbs and participles occur?

You could search for everything containing a hyphen (using ***-***), but this would produce too many unwanted items, including lots of truncation symbols (double hyphens). So instead.... (First check **Concordance/Search Options** to make sure that **\$** is the wildcard symbol for "exactly one" character, and ***** is the symbol for "zero or more" characters. Also check that neither symbol is listed under "characters to treat as delimiters".)

Search for **\$*-\$*ing**

Search for **\$*-\$*ed**

Search for **\$*-\$*s**

What do you find this way?

How do genitives and definites interact?

Use **Concordance/Search Options** to set the "**matches between**" numbers to 1 to 7 words. Then, do **Concordance/Search** for **the & of the**. What does this kind of search give you?

How do speakers use special voice qualities and prosodic modifications?

Use **Concordance/Search Options** to set the "**matches between**" numbers to 1 to 10 words. **Search** for **<* & *>** What do you see?

To look more narrowly at just special qualities (i.e. as used by speakers imitating another's speech, making fun of someone's dialect, and so on), **Search** for **<VOX & VOX>** What are these speakers doing?

Setup options

This section offers some suggestions on how to modify the various options that MonoConc offers. The most significant options are those found under the headings of **Concordance (Search Options)** and **Collocation (Frequency Options)**. These options can make a big difference in the efficiency of your searches, and even in their accuracy. So it is well worthwhile to give careful thought to choosing the most effective settings for your research. In many cases this will involve changing the settings slightly as you ask different research questions.

Choosing the most effective settings for your research is easier once you have gained some experience using concordances. You may wish to work with concordances for a while first, and then, when you know better what you are looking for, pay attention to the details of setting the options.

General Setup Options

Select **Concordance** to set up the **Search Options** the way you want them. Some suggestions, oriented toward discourse research:

Set the **"Max search hits"** to the maximum, i.e. 32000

Set the **"Context characters"** to 120 or so, meaning 120 characters to the left of your search term, and 120 characters to the right. (Choose whatever number fills up your screen.)

Set the **"matches between"** character to **&**.

Set the **"matches between"** numbers to **1 to 5**.

Set the **"wildcard character"** for **"0 or more"** to *****.

Set the **"wildcard character"** for **"0 or 1"** to **%**.

Set the **"wildcard character"** for **"exactly 1"** to **\$**.

Skipping Characters

In recent versions of MonoConc Pro, the option of specifying “skipping characters” has been introduced. Select **Concordance/Search Options**, and type the relevant characters into the box marked **Skipping characters**. If you specify a character as a skipping character, this means that a MonoConc search will ignore it, acting for all intents and purposes as though the symbol were not present in the word. This is very useful if there is a laugh symbol in the middle of a word, or the symbol for lag/lengthening, or an overlap bracket. This way you can have your cake and eat it too: you can have detailed transcriptions that acknowledge the reality that in discourse things sometimes happen in the middle of a word, and you can also achieve easy word recognition in your words searches. For this reason, it is useful to experiment with which characters to specify as skipping characters, following the lines suggested by the previous discussion of delimiters. (Note that under **Advanced Search**, you need to check the box marked **Use skipping and equal characters**.)

Delimiters & Skipping Characters

The following are suggested initial *default values* for delimiters and skipping characters for the published transcriptions in the Santa Barbara Corpus of Spoken American English (e.g. for transcriptions using **Du Bois et al. 1993** conventions [=DT1]). Type the characters into the appropriate spaces in the dialog box. (Do not type any spaces between the characters, nor at the beginning or end.)

Characters to treat as delimiters: - [hyphen]

Skipping characters: 0123456789~!.,#+[]@_%^`=/\

Criteria for choosing skipping characters

For certain kinds of searches, you may want to experiment with adding or removing a character from the list of delimiters, or from the list of skipping characters. The following are a few examples of characters that are sometimes good to include, and sometimes good to exclude (this is not an exhaustive list):

Table 1. Characters to be excluded from some searches

<i>Character</i>	<i>Meaning</i>	<i>Exclude when searching for...</i>	<i>Include when searching for...</i>
, . ?	intonation	words	intonation, punctuation
...	pause	words	pause
[]	overlap	words	overlap, simultaneous speech

<i>Character</i>	<i>Meaning</i>	<i>Exclude when searching for...</i>	<i>Include when searching for...</i>
:	length	words	prosodic lengthening
@	laughter	words	laughter, laughing while speaking
%	creak	words	creaky voice
-	(hyphen)	most words	compound words, truncation

Characters to include always

Note that some types of punctuation symbols are okay to include in all searches; that is, they are *not* recommended for exclusion from most searches. For this reason, they should not normally be placed on either the delimiter or the skipping list. These are the symbols which are used to represent the following transcription categories:

Table 2. Characters to be included in all searches

<i>Character</i>	<i>Description</i>	<i>Meaning</i>	<i>Include when searching for...</i>	<i>Irrelevant for...</i>
'	apostrophe	contraction, elision	words (<i>can't, don't, etc.</i>)	
()	parentheses	vocalism	vocalism	word search
()	parentheses	pause timing	pause, event duration	word search
;	semi-colon	speaker attribution [DT2]	speakers, turn beginning	word search
:	colon	speaker attribution [DT1]	speakers, turn beginning	word search*
< >	angle brackets	manner/quality	manner, quality	word search
–	en dash	truncated word	truncation	word search
—	em dash	truncated intonation unit	truncation	word search

[*The one exception or special case concerns the colon. In DT2 this indicates prosodic lengthening of sounds, so it *should* be a skipping character, because it occurs inside words; but in DT1 it indicates speaker labels, so it should *not* be a skipping character. It never occurs inside words in DT1, and it may be useful for searches that target turn boundaries or speakers.]

The colon character is treated differently depending on what it stands for. For transcriptions in which colon represents a participant attribution, it should be left *off* the delimiter and skipping character lists. For transcriptions in which the colon represents lag or prosodic lengthening, it would normally be placed *on* the skipping character list.

Note that it is quite possible that at some point you will want to search for words spelled as contractions, vocalisms, participant attributions, manner/quality features, and so on. If so, you will need to include the characters in Table 2 as part of your search (i.e. maintain them as options available to you in your repertoire of searchable signs). And since they do not normally interfere with ordinary word searches (because they rarely or never occur inside words in a DT transcription), they do not normally need to be excluded (e.g. by listing as a skipping character).

What difference does the choice of delimiters make to the way a search turns out?

To understand what a difference your list of "**characters to treat as delimiters**" makes, first select **Concordance**, then **Search Options**. Erase all the characters found in the box labeled "**characters to treat as delimiters**", and then type the following characters (with *no spaces between them*) into the box in their place: `; ! # ^ + / \ | ~`

Then **Search** for the word **mh**m. How many instances of the word are there in the data, according to this search?

Now add (type in) the following additional characters in the "**characters to treat as delimiters**" box (again without any intervening spaces): **0123456789** Repeat the **Search** for the word **mh**m. Now how many instances of the word are there in the data, supposedly? Why did the change occur?

Now add in the following characters and repeat the search: `[]` What change is there in the number of **mh**m's found in the data, and why?

Finally, add in the intonation-unit final symbols: `, . ?` Repeat the **Search** for the word **mh**m. How many additional instances of **mh**m turn up this time? Why?

To round out the picture, this time keep the same set of delimiters, but broaden the search formula. That is, **Search** for the formula **m%h%m***. Now **Sort** the results by **Search Term**. How many **mh**m's are there that fit this pattern? What are the new instances that appear?

Now let's try a trick to look for some more **mh**m's that may not have been found yet by our previous searches. Search for **hm** (*not mh*m). Then **Sort** by **Search Term**, and look through the data. Do you see some instances of **mh**m here? Were they found in previous searches? Why or why not? Why do they show up in this search for **hm**?

Now remove the numerals and square brackets `[] 0123456789` from the list of delimiters. **Search** for ***m%h%m*** Then **Sort** by **Search Term**. Remove any inappropriate items by highlighting them and deleting with **Control-d**. Are there new instances of **mh**m that were not turned up by previous searches? What kinds?

How does your choice of delimiters affect your research? Why might you need to do a search more than once, with different sets of delimiters?

For interest's sake, you might want to consider how well **mh**m correlates with turn beginning (before and after it). Compare the number of **mh**m's found in the previous search with the number for the following two searches:

: *m%h%m

***m%h%m* *:**

***: *m%h%m* *:**

Skipping characters vs. delimiters

Skipping characters can produce similar effects to those achieved above by using delimiters. Whenever skipping characters are available (e.g. for MonoConc), it is generally preferable to use them, as they are more flexible in certain respects. For software that lacks skipping characters, it may be adequate in some cases to use delimiters. However, the effects of skipping characters often cannot be reproduced with delimiters. The details of the differences are beyond the scope of this document.

In general, the value of skipping characters is highly important for the analysis of prosodically transcribed discourse (e.g. DT1 and DT2).

Stop List

Set up a "Stop List" of "words" to be ignored in your searches. (Depending on your research goals, this could include pseudo-words like ((BEEP)), <<FOOTSTEPS, speaker attribution labels, and possibly even laughter, breathing, and so on.) To do this, first select **Collocation**, then **Frequency Options**. Check the box marked "Content words only", then select **Edit**. Look at the words in the stop list. **Add** or **Remove** words individually if you need to, one to a line.

If necessary, **Load** a stop list from a file. For example, **Load** the file named **stoplist.txt**, which you may find located in the appropriate corpus directory (such as **\Data\English\SBCorpus\StopList**).

If you need to do a lot of work on the stoplist, you may find it useful to use an ASCII text-editor like the Windows programs Notepad or WordPad to edit a list (as an ordinary plain text file in ASCII format).

You can then **Load** this file into MonoConc.

Note that for some purposes you might want to include vocal noises such as **(H)**, **(TSK)**, and so on, as words, for purposes of looking at collocate frequency information. If so, you should remove them from the stoplist you are using, or check the box "Count all words" under **Frequency/Frequency Options**.

What difference does a Stop List make to how the results of a search turn out?

To understand what a difference your Stop List makes, first select **Collocation**, then **Frequency Options**, then check the box marked "Content words only". Then **Search** for the word **um**. Get the **Collocate Frequency** (under **Collocation**). Take note of the most frequent words that co-occur with **um**. What patterns do you see?

Now select **Collocation** again, then **Frequency Options**, then check the box marked "Count all words". Then **Search** again for the word **um**. (It is necessary to make a new search in order to get MonoConc to use the new Frequency Options settings you have just selected.) Again, get the **Collocate Frequency** (under **Collocation**). Now take note of the most frequent "words" that co-occur with **um**. What is the difference here? Which of the two pictures of the data you have just created is more enlightening for the word **um**? Are there other words for which the opposite choice might be better?

Regular Expressions (advanced)

"Regular expressions" represent a powerful way to analyze your discourse data. They are somewhat technical in nature, and may require some getting used to. They are included here as an option for more advanced computer users.

The following "regular expressions" may be useful for searches in the Santa Barbara Corpus, Parts 1 and 2, 1st edition (i.e. transcribed following conventions in Du Bois et al. 1993). The format of the formulas is that used in the concordance program MonoConc, but similar formulas may be useful for carrying out this kind of search in other programs as well.

Regular expressions are part of the "Advanced Search" function in MonoConc, which you can access by pressing **CTRL+a**. For more information, consult the help function in MonoConc: **\Info\Help**, and search for "Regular expression".

Please note that it will make a big difference in these searches whether or not you have checked the box "Use skipping and equal characters". Try it both ways to see what happens. Also, you may wish to experiment with the choice of characters to be specified as skipping characters, and with the choice of delimiters.

In the table below, the first column specifies the target, that is, the symbol (or an example of it) that you are looking for. The second gives the regular expression, what you type into the search box in MonoConc

Pro. The third column gives a gloss, naming the category being searched for.

Target	Expression	Gloss
&	&	IU continued
~	~	pseudo-graph
#	#	real name (LDC Part 1)
!	!	real name (LDC Part 2)
=	=	prosodic lengthening (DT1)
@	@	laugh sign, single
<@	<@	long feature: laughter begins
[\[overlap begins
]	\]	overlap ends
@@@	@+	one or more laugh signs
@@@	\b@+\b	one or more laugh pulses, surrounded by word boundary
XXX	\bX+\b	one or more unintelligible syllables, w/ word boundary
@word	@[a-z]+	laughing while speaking a word
,	,	continuing intonation (comma)
?	\?	appeal intonation (question mark)
.	[^\.]\.s	final intonation (period)
...	\.{3}	pause, medium or long (3 dots)
..	\s\.{2}\s	pause, short (2 dots, with surrounding whitespace)
--	--	truncated IU sign (two hyphens)
--	\s--	truncated IU sign (with whitespace preceding)
--	\S--	truncated IU sign (with NO whitespace preceding)
-	[^]-\s	truncated word (followed by whitespace)
- --	-\s--	truncated IU in which last word of IU IS truncated
--	[^]-\s--	truncated IU in which last word of IU is NOT truncated
-	[^]-\s[^--]	truncated word NOT followed by IU truncation
-	[a-z]+-[a-z]+	hyphenated word
Target	Expression	Gloss
((MIC))	\([A-Z]+\)	comment, one-word
(SNIFF)	[^\(][A-Z]+\)	vocalism
X-ray	[a-wyz]*x-[a-wyz]+	word containing one letter X, but not only X (e.g. XX)
(TSK)	\(TSK\)	click
:(TSK)	:s*\.*s*\[*[0-9]*\)(TSK\)	click at the beginning of turn (ignoring pause, overlap)

The following notations are for the version of the Santa Barbara Corpus which is time-stamped in seconds (i.e. the LDC published version).

Target	Expression	Gloss
13.27	\s[0-9]+\.[0-9]+\s	one timestamp (with surrounding whitespace)
@ 13.27	@+\s[0-9]+\.[0-9]+\s	laugh followed by whitespace, then timestamp
@	@\d*\]\s[0-9]+\.[0-9]+\s	IUs ending in a laugh, with or without overlap
(H)	\(H\)\d*\]\s[0-9]+\.[0-9]+\s	IUs ending in (H), with or without overlap (NB: square brackets and numbers are removed from the “delimiters” list]
(Hx)	\(Hx\)\d*\]\s[0-9]+\.[0-9]+\s	IUs ending in (Hx)
((MIC))	\)\)\d*\]\s[0-9]+\.[0-9]+\s	IUs ending in comments ((COMMENT))

Expression	Gloss
\s[0-9]+\.[0-9]+\s[0-9]+\.[0-9]+\s	two timestamps in a row (w/ whitespace)
\.\s[0-9]+\.[0-9]+\s[0-9]+\.[0-9]+\s	final intonation (period, then two timestamps)
[?,&_\. -]\s[0-9]+\.[0-9]+\s[0-9]+\.[0-9]+\s	IUs ending in ? , . _ - &
[^?^,&^_\.^ -]\s[0-9]+\.[0-9]+\s[0-9]+\.[0-9]+\s	APPROXIMATELY: IUs with no final boundary tone marker (but note the substantial number of false positives)
\([0-9]*\.[0-9]*\)	will include all and only pauses in DT2; for example (.) (..) (...) (0.7) (13.29)

[Rev. 27-Sep-2009]

Appendix 18

Resources:

Access to Computational Tools for Discourse Research

Overview

This document describes resources available to support research on spoken and written discourse. The resources are mainly of three types:

- software for working with sound, including sound editing and phonetic analysis programs
- software for working with transcriptions, especially for making concordances
- web pages that provide access to lots more information about corpora, concordances, linguistic analysis software, etc.

Several of the programs mentioned below are introduced in this course, and are available for use in the Linguistics Lab. If you find the software presented in this course to be useful, you may wish to obtain a copy for your own use. Among tools of interest to students of spoken discourse, two kinds are of special interest: programs for editing or analyzing sound, and for creating concordances.

Sound Editing and Analysis

Sound Forge 8.0 is a very useful program for working with sound. Most of its features which are of interest to students of spoken discourse are also found in **Sound Forge Audio Studio 8.0**. The Audio Studio version is significantly cheaper, and nearly as good for most purposes of spoken discourse research. Note that both programs require Windows XP.

Sound Forge 8.0 (about \$380) or Sound Forge Audio Studio (about \$70) can be ordered from Sony Media Software at:

<http://mediasoftware.sonypictures.com/products/soundforgefamily.asp>

But the cheapest way to obtain either Sound Forge program is through a purchase at an educational price (e.g. from www.CreationEngine.com). (For some educational software purchases there may be a minimum order of 10 copies, so it may be best to try to do make such purchases jointly as a class.)

(However, if you are on a strict budget, some of the useful functions can be found in the non-commercial program PRAAT – although, for the features where Sound Forge excels, PRAAT may seem much slower to work with, and not as user-friendly.)

A number of programs for working with sound are available for free on the Internet. Perhaps the most widely used is PRAAT. This is oriented to phoneticians, and has a significant learning curve. It requires some technical sophistication in phonetics to get the most out of it.

PRAAT can be downloaded from www.praat.org.

A related tool is the **Fujisaki editor**, which can be downloaded from:

www.tfh-berlin.de/~mixdorff/fujisaki_analysis.htm

Transcription

XTrans is a useful tool for transcribing recordings and aligning the transcription with audio. It can be downloaded for free at: <http://www ldc.upenn.edu/tools/XTrans/>

Transcriber is another useful tool for transcribing recordings and aligning the transcription with audio. It can be downloaded free from www.etca.fr/CTA/gip/Projets/Transcriber/. Normally you should choose one of their “binary distributions,” which are available for Windows, Mac, Unix, and Linux.

Elan is a very powerful tool for transcribing and analyzing audio and video. It is available from the Max Planck Institute, Nijmegen at: <http://www.lat-mpi.eu/tools/elan/>

VoiceWalker is useful for playing back recordings to transcribe them. Version 1.0a is recommended, though you may also wish to try version 2.0. VoiceWalker is available for free from the UC Santa Barbara Linguistics Department, on the Web at:

<http://www.linguistics.ucsb.edu/projects/transcription/tools.html>

Concordance-Making

MonoConc Pro (about \$85.00) can be ordered at <http://www.athel.com/>

In addition, there are several other concordance programs that you may wish to consider.

Wordsmith Tools has a large user base and a good reputation. It is available at www.lexically.net/wordsmith/

AntConc is free, and available at:
<http://www.antlab.sci.waseda.ac.jp/software.html>

Concordance by R.J.C. Watt claims to work with Unicode. I have not tried it, but it does allow you a 30-day free trial. It is available at:
www.concordance.ukgateway.net/Concordance200Manual.htm#top

If you want to work with British English, one of the very best concordance tools available anywhere is the **BNCWeb**, available at: <http://homepage.mac.com/bncweb/home.html>. But this software requires that you first set up the **British National Corpus** (BNC) on a UNIX or Mac OS X computer. See: www.natcorp.ox.ac.uk/

Another corpus of British English is **ICE-GB**, available at:
<http://www.ucl.ac.uk/english-usage/projects/ice-gb/>

To use the World Wide Web as a source of linguistic data – a giant, unruly corpus – check out:
<http://www.webcorp.org.uk/>

Corpus Linguistics Resources on the Web

The following URL's and home pages have lots of material of interest for researchers wishing to use computers to work on corpora of spoken or written discourse. Information is available on numerous and diverse languages (though there is far more material on written language, and on English).

David Lee's "Bookmarks for Corpus-based Linguists"

<http://tiny.cc/corpora>

Stefan Gries

<http://www.linguistics.ucsb.edu/faculty/stgries/other/links.html>

Manuel Barbera

<http://www.bmanuel.org/clr/index.html>

Corpora List

<http://helmer.aksis.uib.no/corpora/>

Texts, text centres, resources and programs on the Web

<http://helmer.aksis.uib.no/text.htm>

TalkBank

<http://www.talkbank.org/>

Linguistic Data Consortium

<http://www ldc.upenn.edu/>

Language Archives

<http://www ldc.upenn.edu/exploration/archives.html>

Consent, Ethics & Human Subjects

Human Subjects Committee (including Human Subjects Training Module) (UCSB):

<http://research.ucsb.edu/compliance/index2.shtml>

Informed Consent Guidelines (TalkBank):

<http://www.talkbank.org/share/irb/>

Code of Ethics (TalkBank):

<http://talkbank.org/share/ethics.html>

"Can We Tape?" -- Legal aspects of audio recording (phone calls, conversations, etc.) for journalists:

<http://www.rcfp.org/taping/>

[rev. 26-Sep-2009]

Appendix 19

File naming conventions for media excerpts

This Appendix explains file naming conventions that are useful for identifying and locating media excerpts (audio and/or video clips) used to illustrate phenomena in spoken discourse. For example, these conventions may be used for excerpts illustrating transcription categories and conventions.

In general, media excerpts (audio and/or video clips) represent brief excerpts taken from a longer recorded source. For *Representing Discourse*, most of the original source files are already published as part of the Santa Barbara Corpus of Spoken American English (<http://www.linguistics.ucsb.edu/research/sbcorpus.html>).

The file naming convention used for these audio clips is as follows: *SourceFileName_StartTime-EndTime*. For example, the interpretation of the filename *SBC028_0117-0184* can be broken down as follows:

<i>Item</i>	<i>Notation</i>	<i>Meaning</i>
Source file (Discourse ID#)	SBC028	The <i>discourse ID</i> number in the Santa Barbara Corpus is number 028. (This identifies the source audio file, e.g. the published Santa Barbara Corpus .WAV file, most of which are between 20 and 30 minutes long.)
Start time	_0117	The <i>start time</i> for this clip is 117 seconds from the beginning of the source file.
End time	-0184	The <i>end time</i> for the clip is 184 seconds from the beginning of the source file.

The numbers are rounded off to the nearest relevant second. The start time is always rounded down, while end time is rounded up (so that the numbers, while approximate, are fully "inclusive" of the actual example on the original recording). In most cases, opening the file for the audio clip will provide more precise information showing the exact start and end times (to multiple decimal places), in case that level of precision becomes relevant.

Whenever possible, filenames following these explicit conventions have been used to label discourse examples presented in the book *Representing Discourse*. They may also be used for published articles, term papers, course handouts, and conference presentations. The timestamp numbers provide information that allow users to reference, identify, and listen to the source for any cited audio clip. With explicit reference information linked to the Santa Barbara Corpus, for example, readers can look up and listen to the audio for any examples of interest.

Because the examples in *Representing Discourse* are mostly short, and already published as part of the Santa Barbara Corpus, these clips may be used (e.g. played) under standard "fair use" terms. In particular, they may be played in public without additional concern for anonymity issues, given their prior publication with consent. Note, however, that they are copyrighted material (copyright University of California, Santa Barbara and Linguistic Data Consortium), so the appropriate restrictions on copying and resale (beyond fair use) still apply.

[rev. 27-Sep-2009]

Project 1

Transcription

Objective

The goal of the transcription work in this course is to help you learn how to develop an understanding of the nature of discourse—of language in use—by creating a systematic written representation of it, i.e. by transcribing it.

Procedures & Format

The primary source of information on the details of how to do the transcription work for this course is in Appendix 7 and Appendix 8 of *Representing Discourse* (Du Bois n.d.).

- For *transcribing procedures*, carefully follow the specifications in Appendix A.7.
- For the *transcription format*, carefully follow the specifications in Appendix A.8.

While the present document adds a few logistical details, it does not attempt to summarize or replace those Appendices, which should be consulted directly.

Data

For the very first introduction to transcribing practice, you may be asked to transcribe a recording supplied by the instructor (the so-called “Minute 0” recording). This will be made available to course participants in digitized format (as .WAV computer files). More details will be given in class regarding the initial transcription.

For most of the course, participants will transcribe a conversation they have recorded themselves during the first week or two of the course. Each participant will transcribe a total of 4 minutes from their own conversation, at the rate of one minute per week. We will refer to these weekly increments of transcription as “Minute 1,” “Minute 2,” “Minute 3,” and “Minute 4.”

Overview

- *Transcription.* Transcribe one minute of your conversation per week, according to the Assignment Schedule.
- *Checking.* Check your partner's transcription (one minute per week). This is due one week *after* your partner transcribed it, according to the Assignment Schedule.
- *Alignment.* You should time-align each intonation unit in your transcription with the corresponding portion of the recording, using sound editing software.
- *Levels.* The level of delicacy for *transcribing* a given minute is specified in the Assignment Schedule. The level of delicacy for *checking* a given minute is determined by the transcription you are checking. You should check the transcription at the same level as the de facto level of the transcription as presented to you by your partner.
- *Revision.* When the full 4-minute Transcription Segment is revised, all 4 minutes should be brought up to the same level of delicacy, and fully time-aligned with the recording.
- *Schedule.* The due dates are as specified in the Assignment Schedule.

[rev. 27-Sep-2009]

Project 2

Ethnographic Recording

Objective

The objective of this project is to help course participants learn how to:

- record a naturally occurring spoken interaction
- document the interaction ethnographically
- prepare the recording for transcription
- enhance the usefulness of the recording for research

Assignment Summary

This project is composed of twelve distinct steps, which are summarized here:

1. Learn about working with human subjects.
2. Record a spoken interaction.
3. Document the participants and the event ethnographically.
4. Digitize your recording.
5. Make a Recording Log.
6. Make a Segment Proposal.
7. Make a Research Evaluation.
8. Select your Research Segment.
9. Select your Transcription Excerpt.
10. Present your Segment Proposal and your recording.
11. Make a Voice Identification table.
12. Write an Ethnographic Sketch.

Each step is explained fully below.

1. Learn about working with human subjects.

Before making your recordings, you should first educate yourself about the appropriate ways of conducting research about “human subjects”. While not all forms of ethnographic recording fall into the category of research requiring official approval by the local human subjects committee (or internal review board), sooner or later most researchers (including graduate students) who make audio and video recordings of people are likely to find themselves involved in research which does fall under the ethical and legal guidelines of these committees. It is a good idea to educate yourself about the issues as early as possible in your research career.

For example, the University of California, Santa Barbara provides a Human Subjects Training Module, which takes about half an hour. The training is done over the Internet, at the following address:

<http://research.ucsb.edu/compliance/index2.shtml>

When you go to the web site, you can read “Training FAQ” if you like. Then select “Human Subjects Training Module”, then “Login”. You will be presented with a form that asks you for some information. The codes you will need to sign up for the Human Subjects training in this course are as follows:

I.D. Number: LING-DU-JO-003

Your last name:

Your first name:

Your e-mail:

Your sponsor's e-mail: dubois@linguistics.ucsb.edu

Last name of I.D. number owner: dubois

Other notification:

For “Your last name,” etc., give them just what they’re asking for. Under “Other notification,” you could, optionally, supply the email address of another professor who requires you to do the Human Subjects training for their course (e.g. bucholtz@linguistics.ucsb.edu).

2. Record a spoken interaction.

Record a spoken interaction. Your recorded interaction should have the following attributes:

1. a naturally occurring interaction (not staged or contrived)
2. 3 to 5 speakers (preferably)
3. contains at least some lively back-and-forth interaction
4. contains some overlap, but not too much (say, about 15-20% of lines being overlapped)
5. lasts at least 40-60 minutes (whether speakers talk the whole time or not)
6. conversation may be in any language you know well
7. excellent audio quality, recorded on high quality equipment

You can make your recordings working individually, or working together with your partner. You should produce one usable recording per person. Since you will be working with this recording throughout most of this course, be sure to make it a good one! (In addition, you will publish the recording as part of the Class Corpus.)

If possible, you should try to capture the entire natural speech event from beginning to end in your recording. Ideally, the conversation will turn out to be a reasonably lively one, to provide adequate challenge in transcribing. The minimum of three speakers is intended to increase the likelihood of lively interaction and turn-taking, including stretches with overlapping speech. The limit of not more than five or six speakers is to avoid *too much* overlap, in the interests of feasibility—a reasonably easy transcription. (In certain kinds of more formal speech event such as a class discussion, however, there may be more participants and yet an acceptable level of overlapped speech.)

(A recording with just two speakers *may* be viable under certain circumstances, if the recording is highly interactional. But get permission from the instructor *in advance* if you want to record a 2-person interaction.)

Sound quality. You should record the interaction using a recorder capable of high quality stereo sound,

with high quality (external) stereo microphones. For example, you may use a stereo digital audio recorder, preferably one that records directly in a standard computer file format (e.g. a flash recorder such as the Ediorol R-09 or a hard disk recorder that records uncompressed audio files, such as WAV format), or a digital video camera (e.g. mini-DV), preferably with external stereo microphones. Try to get the best audio quality you can, under the most natural circumstances. Especially for the segment you select for publication in the Class Corpus, the audio quality should be very good to excellent.

Video quality. Those who may want to use their recording in future research in sociocultural or interactional linguistics (i.e. after this course) should consider making a video recording. However, you will have to make your own arrangements for obtaining a video camera and learning how to use it. If you do use video, be sure to get a good image that will contribute effectively to your ability to do research with the recording. Choose an appropriate camera angle, pay attention to the lighting on your subjects, include all the relevant interactional information in the picture you shoot, and use good equipment, including external microphones for good sound quality.

Language. The speech event you record may be in any language you know well enough to transcribe (i.e. for which you have native or near-native fluency). If you speak a language other than English natively you are encouraged to use it for this project. You may even use a language that your teammate and instructor do not know. (If you choose this option, you will have to provide an interlinear free translation for your transcriptions, but this is not too much extra work.) Most of the speakers on your recording should be adults with native or near-native command of the language they are speaking. (But if one or a few in a multi-party conversation are non-native speakers this is okay.)

3. Document the participants and the event ethnographically

At the time of the event you are recording—normally immediately afterwards—you should gather relevant information about the participants and the event itself. Normally this will take the form of an “exit interview,” which is organized around appropriately designed research protocols, as represented in forms. The relevant forms include a Participant Questionnaire for each speaker, and an Event Description and Setting Diagram for each event recorded. These forms are designed to record relevant ethnographic information about the speech event and each participant in it. The reason for doing this as an exit interview is to avoid having the participants spend a lot of time on your recordings talking about the minute details of your Participant Questionnaire or other aspect of your research logistics.

If you are present at the time of the recording as an observer (e.g. video camera operator), you may find it useful to take notes on the ethnographic context, if you can do so unobtrusively. (But if you are present in the role of a participant in the conversation, you should not take notes—just “act naturally,” as they say.) Obtain signed consent forms from each speaker on your recording. Note that if you are one of the speakers on your recording, you must document yourself in the same way, including signing the consent form.

It is important to note that your recording will be approved for use *only* when you have completed the following:

- *fully documented* your speech event ethnographically
- secured *signed consent forms* for each participant who appears in your Research Segment (§8) and your Transcription Excerpt (§9)

For a more complete discussion of how to use the relevant forms to collect ethnographic information, see Appendix A.10.1 on “Ethnographic Documentation.”

4. Digitize your recording

If your recording was made using a WAV-format flash recorder (such as the Edirol R-09 or similar recorder), you don’t need to digitize it, since it’s already in the right format. Just transfer it to a computer hard drive for safe-keeping, using a standard USB cable. (And backup your data in a safe place!)

But if the recording is made in another format, you will need to digitize your recording, using appropriate hardware and software. The resulting computer-digitized form of the recording (i.e. a .WAV file on a computer hard disk) is the basis for what we will use exclusively for all work in the course, including transcriptions and assignments. You should digitize either just the relevant portions for your research, or the entire recording. Normally we will mostly work with a “Research Segment” of about 20-40 minutes, rather than from the full original recording, or with the “Transcription Excerpt” of about 4 minutes (see discussion of the Segment Proposal below).

5. Make a Recording Log

Listen carefully to your entire recording (or the research segment), and log the contents (see Appendix A.10.1 on “Ethnographic Documentation”). To do this, start playing the recording and just leave it running, jotting down informal notes as you go. There are two ways to manage this:

(1) *Written log*. You may write the information on paper or in a computer text file. You can use the Recording Log form, or design your own, registering the following information at regular intervals during the recording:

- Time
- What's happening (e.g. gossip, arguing, silence, preparing dinner, etc.)
- Participants (comings and goings of participants)

In order to make reference to a reliable and reproduceable timeline for all events on your recording, you should use the digital time code from the computer audio file or other digital source, if possible. This will specify the time elapsed from the beginning of the computer file of the recording, which will allow you to return to the exact location whenever you need to find something.

In addition to the above, rate each section of the recording for the following criteria, on a 5-point scale (5 is excellent, 1 is poor):

- video quality
- audio quality
- transcribability
- research potential

(2) *Annotated media log*. You can log the same information in a more usable format, by attaching annotations to specific locations in your recording. To do this, start with the computer file of the Research Segment of your audio or video recording. Use the appropriate software, such as a sound editor or transcription tool, to record your annotations. For example, you can use sound editing software to place “Markers” at various interesting time locations in your recording. For each marker, you then write

the relevant information for what is going on in its neighborhood, using the software to take notes. Your annotated media log should include the same information as indicated above for the written log.

6. Make a Segment Proposal

In advance of the meetings described below (for presenting your proposed Research Segment and Transcription Excerpt to your team and to the class; see §8, §9, and §10), make two specific proposals regarding which segments of your recording you will work with during this course. Use the Segment Proposal form, or make your own. Make sure you accurately record the following information on each Segment Proposal form:

1. the start and end times of your proposed segment (using a reproducible digital time code if possible, preferably from your computer-digitized file)
2. the first few words of your proposed segment (as a mnemonic and confirmation of the correct location)
3. the last few words of your segment (for the same reasons)
4. your evaluations of your recording according to the specified criteria (5=excellent, 1=poor.)
5. your reasons for starting and ending the excerpt where you do

Make photocopies of your Segment Proposals to give to each member of your research team at your team meeting.

7. Make a Research Evaluation

Do a Research Evaluation of your recording, judging its research potential on the relevant criteria. Either with your teammates or on your own, listen to and evaluate each recording. Record your own judgments, and a summary or consensus of the group's judgments, on a Research Evaluation form. (If you like, this can be done in a group as part of the team meeting described below.)

8. Select your Research Segment

Select an appropriate "Research Segment" from your recording, and present it to a meeting of your research group, as a candidate for inclusion in the Class Corpus. Your Research Segment should be about 30 minutes long, give or take (normally a minimum of 20 and a maximum of 40 minutes long). Choose the start and end points carefully, so as to capture a meaningful unit of interaction if possible. You should know your recording well, so that you are prepared to present your case in favor of the particular segment boundaries you are proposing. Provide photocopies of all relevant documentation to your teammates. (That is, distribute your Research Evaluation and your Segment Proposal to all. Your Recording Log should be available for reference but need not be photocopied for everyone.)

For your meeting, you will need to reserve a location suitable for talking about and evaluating your recordings, as well as facilities for listening to the recordings on good quality stereo loudspeakers (e.g. on a home stereo system, or in the Linguistics Seminar Room). Using headphones won't work because everyone needs to hear in order to provide input. And cheap computer speakers are ill-advised, because they disguise the qualities (positive and negative) of the recording you are trying to evaluate.

Based on your listening and discussion as a group, each participant should rate each segment presented on a 5-point scale (5 is excellent, 1 is poor) for the following criteria (see Section 1):

1. naturalness
2. audio quality
3. video quality (if applicable)
4. transcribability
5. research potential

While “research potential” is obviously a subjective judgment which will depend greatly on the goals of the researcher, for our purposes one of the key factors to consider is that the segment should be interactionally lively (i.e. including some back-and-forth between speakers, and probably some overlap during competitions for the floor, etc.). If the recording documents a hard-to-find type of interaction, or one of special interest to you, this can also be relevant to your assessment of its research potential.

Treat the ratings as a significant factor in your decision about what to choose as your Research Segment. (You will need to use the same evaluative criteria, applied even more rigorously, for your Transcription Excerpt; see §9). In the end, the decision is up to you, but it should be one that will stand up well when you make your class presentation (§10).

Some time after the meeting where you decide on your Research Segment boundaries, you should prepare a computer file of the relevant portion of your original recording. Use a sound editor to mark the exact start and end in your original recording file. Then select the exact Research Segment, and copy it to a separate file.

9. Select your Transcription Excerpt

At the same meeting, you should also present your candidate for selection as your Transcription Excerpt. This is a 4-minute stretch of continuous discourse, which must be a subset of your longer Research Segment. Choose your Transcription Excerpt carefully, because you and your partner will be listening to and transcribing this segment very carefully over the next five or six weeks of the course. The excerpt you choose should excel on the various evaluative criteria discussed above.

Try to avoid choosing a Transcription Excerpt in which one speaker tends to dominate the conversation. For learning purposes, it is important to be exposed to several different voices and speech styles. This is especially important if you are a participant in the conversation: don’t pick a section in which you are the dominant speaker (although it is fine if you are a participant, of course). This is mostly relevant to the initial stages of learning about transcription, as some people find it hard to listen to their own recorded voice in an objective and analytical way. (This should not be taken as a general injunction against recording yourself; it’s entirely possible to produce excellent research materials in which you yourself are a participant. And under no circumstances should you try to limit your own participation while you are present in a recording. This will produce an unnatural recording of you, and is likely to throw everyone else off the track as well.)

When you have decided on your Transcription Excerpt boundaries, use a sound editor to mark the exact start and end in your Research Segment file. Then copy just the Transcription Excerpt into a separate file. This is the file that we will be passing back and forth as you transcribe and check throughout the course.

10. Present your Segment Proposal and your recording

During this class session, all course participants and the instructor will (briefly) listen to and critique your recording.

In advance of the class session, prepare to orally present your Research Segment (§8) and Transcription Excerpt (§9) proposals, as approved in your team meeting, to the class at large. Be prepared to justify your start and end points for the Research Segment, and the quality of your Transcription Excerpt on the relevant evaluative criteria.

You will be asked to play one or two brief, well-chosen illustrative excerpts from your recording (totalling about 30-40 seconds or so). Choose your illustrative excerpt(s) carefully to support your presentation to the class. Note the exact start and end time of your illustrative excerpt *in the computer file you will be using*, and mark it as a region in this file, so you will be ready to play the right piece when you are asked. If your computer file won't play, your proposal cannot be approved.

Based on what you say and what you play, the instructor will decide on the spot whether your proposed segments are suitable for use in this course.

Documentation. On the appropriate day (according to the Assignment Schedule), turn in all the documentation for your recording, including the various ethnographic information sheets, original signed consent forms, Segment Proposals, etc., as well as computer files for both the Research Segment and Transcription Excerpt. (Computer files should be turned in on CD-ROM and/or in the appropriate computer file location as designated for this course.)

11. Make a Voice Identification Table

As a guide to future transcribers who may not know your recording as well as you do, identify by voice each individual who speaks on your Research Segment (see Appendix A.10.1 on "Ethnographic Documentation"). (This is important for making clear which speaker on the recording goes with which speaker in the database of ethnographic documentation, especially where names are changed and bleeped to preserve anonymity.) To do this, write down a few words as uttered by each individual, along with the time on the recording where the words were spoken. Choose a distinctive and easily identifiable utterance. Normally you should choose a sentence or an intonation unit uttered shortly after the first appearance of each individual on your recording. (In most cases, much or all of this work will be automatically taken care of as part of your first minute or so of transcription.)

12. Write an Ethnographic Sketch

Write up a brief ethnographic sketch for your Research Segment. This should be a readable prose description, drawing on the information you gathered in your ethnographic documentation plus listening to the recording. You should provide a narrative that gives readers the overall picture of the nature of the event and the identity of the participants, plus any specific cultural background they may need in order to understand what is going on in the recording. The work of gathering the information for this sketch should be begun at the time you originally make the recording (e.g. in filling out the various forms, etc.). However, you will most likely want to complete the sketch after you have transcribed some or all of your segment, later in the course.

The length of your sketch may vary from a paragraph up to a page or two, at a maximum (single-spaced with normal margins and fonts). Turn it in as a computer file, and with a paper copy as well.

You may wish to include a brief glossary, in the form of a table listing the meaning of any obscure words or phrases (e.g. specialized knowledge exclusive to these participants that would not be available to the general reader, or in standard reference works such as dictionaries).

(Optional: In the database for the Class Corpus, enter the relevant ethnographic and demographic information.)

Schedule

For information about relevant due dates for the tasks described above, consult the Assignment Schedule.

[rev. 27-Sep-2009]

Project 3

Developing Examples for Teaching Transcription

Objective

The goal of this project is to find and develop examples from your own conversational data which can be used to teach concepts and methods in the transcription of discourse. Such examples can be used in two ways:

1. as an *illustration*, to show how a specific transcription category, concept, or method is to be perceived or interpreted
2. as a transcription *problem*, to be solved by choosing the right category (and symbol) to transcribe it

Often the same example can serve either or both functions.

The Problem

You should begin by setting a problem, based on your experience in transcribing discourse. Have you encountered something in your recording which is interesting and/or challenging to transcribe? How might the phenomena you encounter, and the transcription decisions they require you to make, shed light on the process of learning how to analyze and transcribe spoken discourse?

What you consider “interesting” in a discourse example is, of course, open-ended in principle. It depends in part on what theory of transcription you are using to frame your orientation, and what your research orientation and goals are. Thus, in addition to drawing on problems arising in your discourse transcription experience, you should think about how to frame your examples theoretically, in light of a some broader conception of the goals which lie behind the particular transcription categories you are illustrating. Even if you are a novice in transcription, it is useful to think about the analytical assumptions, implicit goals, and theoretical framing which lie behind the transcription categories you are illustrating. Why is it important to transcribe these things, in this way? You may get some ideas for this from class lectures, discussions, readings, and so on.

Selecting Examples

There are many ways to approach the task of illustrating an aspect of discourse transcription. If you are seeking to illustrate a rare discourse phenomenon, you may find only one or two examples suitable for use, in which case you will have to take whatever you can find. For high-frequency discourse phenomena, on the other hand, you are likely to encounter an embarrassment of riches—a vast number of instances to choose from. In this situation, you can afford to be choosy, selecting examples that go beyond the minimum requirements to meet additional criteria. The following are some useful criteria to consider:

1. The example should normally provide a clear instance of the phenomenon in question.
2. On the other hand, in some cases, the point is precisely to have an unclear case (e.g. a difficult choice of one intonation contour vs. another, or of presence vs. absence of an intonation unit boundary).
3. The example should be short and clearly focused on the problem at hand.
4. The example should be well recorded and easy to hear.
5. The example should avoid difficult or time-consuming transcription problems that distract from

the question at hand (e.g. overlapping speech should normally be avoided—unless, obviously, what you are illustrating is overlapping speech).

6. The example should not be so difficult that it cannot be processed in the time available.
7. Anonymity-compromising proper names and other such pseudograph words should be avoided, if they violate anonymity agreements. (Even if such words have been bleeped, this too should be avoided if possible, because it can be distracting.)

In addition to the above primary criteria, there are some secondary desiderata that are worth taking into consideration. If the example has one or more of the following features, this is considered a plus:

1. The example includes a response from another speaker, demonstrating how the phenomenon in question has been heard and interpreted by a participant (e.g. appeal contour followed by a response, long pause plus some reaction to it, question and answer, false start and repair initiator, disfluency and mimicry of it, stance and counter-stance, and so on).
2. The example includes a “minimal pair,” allowing for direct comparison of contrasting transcription categories (e.g. normal voice vs. vox or ‘voice of another’, with both produced by the same individual in the same exchange).
3. The example includes a minimal pair (or triplet, etc.) with more than one member of a contrasting paradigm within a short distance (e.g. continuing vs. final intonation in two successive intonation units).
4. The example illustrates the same words used twice, but realized with different intonation (e.g. cases of dialogic syntax, where two different speakers may use the same words and/or syntactic construction, but with different intonation).
5. The example illustrates the same intonation used twice, but realized on two different strings of words or syntactic constructions (e.g. isotony, the same tune repeated in successive intonation units).
6. The example is inherently interesting, for whatever reason (e.g. content, affect, conflict, humor, etc.). This may be relevant to why the participants are doing what they are doing. Even when it is not directly relevant to the transcription question at hand, it tends to increase the student’s interest and motivation to arrive at a good understanding of how to transcribe it.
7. The example is relatively self-contained, and retains some degree of coherence even when extracted from its larger discourse contexts.

In developing your example, it is a good idea to consult with others to make sure your example is effective in what it seeks to do. Ask your transcription partner and other course participants to look at and listen to your example, to see that it works in the way you intend.

Types of Transcription Problems

The following are some suggestions of general types of transcription categories or problems that could provide the focus for a useful exercise:

- a single transcription category, with varying instances of the category (e.g. different kinds of continuing contour)
- a choice between presence or absence of a transcription category (e.g. presence or absence of intonation unit boundary)
- a choice between presence or absence of the functional cues for a transcription category (e.g. presence or absence of intonation unit boundary)

- a choice (paradigmatic contrast) between two or more transcription categories (e.g. continuing vs. final contour)
- a problem in functional interpretation (e.g. who does this pause belong to)
- a problem in complex transcription notation (e.g. how to represent multiple overlaps)
- and many more kinds of problems...

Full Transcription

Normally your teaching example will include two versions of each transcription, which may be termed the “full” and “reduced” transcriptions.

The full transcription represents a more or less complete transcription of the selected conversational excerpts, at the level of delicacy/granularity appropriate to the problem as you have defined it. One might consider this to be “the answer” to the transcription problem you have posed. However, acknowledging that there may be more than one valid answer—more than one way to transcribe the excerpts you have selected—your full transcription nevertheless represents at least your best understanding of how to transcribe the excerpts, with respect to the transcription questions you have posed.

Reduced Transcription

The reduced transcription is the same as the full transcription, except that some transcription information has been removed from it. For example, the reduced transcription might deliberately leave out pauses, even though you have already transcribed them. This reduced transcription is provided as both a challenge and an aid to the learner. It is a challenge in that something is left out which the learner must provide. It is an aid in that some features are already transcribed for the learner. This is so that learners faced with the transcription exercise can focus their attention on the new task at hand, that is, learning how to hear, interpret, and transcribe the currently targeted transcription category (e.g. pauses), without being distracted by spending time figuring out what words the speakers are uttering, or identifying other transcription categories not currently being focused on.

What is left out from a reduced transcription will depend on what transcription level and what specific transcription issues are being addressed in your exercise. At the more basic levels of transcription (relatively low on the delicacy/granularity hierarchy), for example, the transcription exercise might be addressed to the problem of identifying intonation unit boundaries. In this case, the reduced transcription would obviously leave out the marking of intonation unit boundaries, but also any prosodic features that might serve as strong cues for the presence of an intonation unit boundary (e.g. a long boundary pause). What remains will be mostly just the words uttered, along with the attribution labels indicating which speakers uttered them (i.e. a “Level 0” or “Lexical” transcription). Specifically, you would remove all symbols for pauses; intonation contour, including truncation of intonation units and words; vocalisms such as in-breaths and clicks; lag or prosodic lengthening; and of course the line breaks (carriage returns) that mark intonation unit boundaries (except where a turn change requires a line break). At a higher transcription level, it might be useful to create a transcription exercise focused on a specific cue or group of cues for intonation unit boundaries. In this case the question posed would be whether a particular cue (such as lag, rush, pitch reset, etc.) was present or not, or which cues in general were present or absent. Here the reduced transcription would retain the indication of the intonation unit boundary itself (i.e. the line break), but omit the specific cue(s) being targeted, which the learner would be asked to provide.

What to include

Your transcription example should be prepared in the form of a handout including the following:

1. A statement of the objective of the exercise, including the transcription phenomenon to be targeted
2. Specific problem(s) and/or question(s) relevant to the targeted transcription phenomenon
3. A “reduced” transcription of the relevant discourse example(s)
4. A “full” transcription of the same examples (on a separate handout, or on the back page)
5. Some discussion of the point of your examples, if appropriate

In addition, you need to isolate the recordings needed to support your example. You will need to create computer file(s) containing just the excerpt(s) that are needed in order to present your transcription example(s). This means one or more small file(s) in computer format (e.g. .WAV audio files, or video files), each of which may only be a few seconds long. You should copy the computer files onto a CD, and also place them on the web or on the local area network.

Presentation

When presenting your example in class, you should do the following in advance:

6. Set up the necessary recording files (audio and/or video) in preparation for playing them in the classroom.
7. TEST your setup to make sure everything works. Make sure that you will be able play your examples in the class when the time comes.
8. Bring photocopies of your handout(s) to distribute to all course participants.
9. Decide on a definite “lesson plan”, a planned procedure for your presentation:
 - j. How will you present the example to the class?
 - k. What will you ask the students to do?
 - l. How much time will it take? Will it fit in the allotted time?
13. Make sure your examples are self-sufficient, so that students can consult them on their own. Make them available in the lab or on the web.

[rev. 6-Sep-2005]

Project 4

Intonation Unit Boundary Cues

Objective

The focus of this analysis is on recognizing the individual cues that characterize the *boundary zone* between the end of one intonation unit and the start of the next. In addition, we consider how to weight the various cues as part of the total complex of cues which motivate an overall conclusion about the presence of an intonation unit boundary.

Procedure

1. Evaluate the boundary zone between two (potential) intonation units, listening for each relevant intonation unit boundary cue. For each cue, mark your decision about whether it is *present* (X) or *absent* (no mark). You can write a single stroke to indicate a cue that seems to be “half” present.
2. Based on your overall evaluation of the complex of boundary cues, decide whether an intonation unit boundary is present. Mark your decision.
3. On a 5-point scale (5 is high, 1 is low), indicate how strong/reliable you consider the presence of an intonation unit boundary to be.
4. You may find it useful to take notes (using the Comments section) to justify or clarify the basis for your decision about a given cue, and about the intonation unit boundary overall.
5. Note that each page of the chart provides space for the analysis of *two* boundaries. Write the words of *two pairs* of (potential) intonation units in the spaces provided.
6. Since the focus is on intonation unit *boundaries*, a single intonation unit may be analyzed twice, once for its initial boundary and again for its final boundary. To do this, write the words of one intonation unit in two places: first, in the upper right box (as the second member of the first pair of intonation units); and second, in the lower left box (as the first member of the second pair of intonation units). This will allow you to analyze both the initial and final boundary cues for the same unit.
7. Use an erasable format to allow for changes and corrections (e.g. a pencil on paper, an erasable marker on acetate, or a computer).

Data

You should find and analyze at least three kinds of cases:

- (7) a clear case, in which the intonation unit(s) are supported by presence of most of the cues;
- (8) an intermediate case, in which the intonation units lack several of the well-known cues (e.g. pause, breath, and turn);
- (9) a doubtful case, in which some cues are present, but where you think there probably is NO intonation unit boundary. (For example, consider a place where a pause occurs, but it's a word-search pause, and no other supporting cues for an intonation unit boundary are present.)

[rev. 6-Sep-2005]

Project 5

Intonation Analysis

Objective

The objective of this project is to explore ways of analyzing and representing intonation. We will focus on intonation contours as they are deployed in naturally occurring spontaneous conversation.

The focus here is on taking a small, carefully selected set of previously transcribed conversational excerpts, and subjecting the data to in-depth intonational analysis. This will involve two kinds of analysis:

- ◆ so-called “impressionistic” transcription of intonational categories at higher delicacy
- ◆ computer-based acoustic analysis of pitch and other acoustic cues for intonation

Through the impressionistic analysis we will explore in greater depth some of the literature on intonational analysis (e.g. ToBI, Hirst & Di Cristo, Cruttenden). The computer-based work may involve learning some new software (e.g. PRAAT; see Appendix on “Acoustic Analysis”).

Data

We will be starting from conversational data which you have already transcribed to a fairly high level of delicacy (e.g. Level 5.2). It is assumed that the appropriate intonation unit boundaries and their cues will have been previously identified, along with basic intonational categories such as boundary closure, appeal, truncation, and primary and secondary accent. For the present project, you should identify some interesting examples of intonation that meet the following criteria:

- ◆ Select 2-5 excerpts from your transcription of your recording, with a total duration of about 15-20 seconds
- ◆ Each excerpt should represent a continuous stretch of discourse of 2-10 intonation units in length

An ideal topic for this project is the intonation of appeal and questions.

If you wish, you may use data that you have worked with previously for other projects (e.g. “Developing Examples for Teaching Transcription” or “Intonation Unit Boundary Cues”). However, your choice of material to analyze for the current project should be well motivated by its interest and relevance to the issues of intonation contour analysis, since you will be making a close analysis of the intonation contours in these excerpts.

A good strategy is to look for cases of “minimal pairs” of contrasting intonational phenomena. For example, the following kinds of data would make an interesting starting point for analysis. (The descriptions below are quoted from the project on “Developing Examples for Teaching Transcription.”) Relevant criteria include:

8. The example illustrates the same words used twice, but realized with two different intonation contours (e.g. cases of dialogic syntax, where two different speakers may use the same words and/or syntactic construction, but with different intonation).
9. The example illustrates the same intonation used twice, but realized on two different strings of words or syntactic constructions (e.g. isotony, the same tune repeated in successive intonation units).
10. The example includes a minimal pair (or triplet, etc.) with more than one member of a contrasting paradigm within a short distance (e.g. continuing vs. final intonation in two successive intonation units).

It's a good idea to choose at least some excerpts that provide nice, clean, well-behaved examples illustrating some of the intonation categories we are learning about. Admittedly, not all conversational data cooperate so nicely in this regard. Naturally occurring conversation can present some very challenging prosodic phenomena which are difficult to transcribe in terms of presently available models of intonational transcription. This may ultimately demand that we revise the current crop of intonation theories. In the meantime, however, as novices we would like to experience some analytical gratification. So choose at least some examples that the intonation models we presently have seem to do a good job of describing.

Analysis

In preparation for the Lab Tutorial, do the following:

1. Select your excerpts according to the guidelines given above.
2. Using the Regions function of Sound Forge to work on your Transcription Segment recording file, make sure all your intonation units in the relevant excerpts are properly aligned. In addition, mark any other regions of particular interest in your selected excerpts that you may want to look at in the Lab.
3. Use a sound editor (e.g. Sound Forge) to make a copy of each of your excerpts from your Transcription Segment recording file into its own small WAV file. (Note that the regions you have marked in your Transcription Segment file will be automatically copied into the short excerpt files.) Make a folder, give it an appropriate name, and copy all your small audio files into the it. Copy the folder onto a CD-ROM and bring it with you to the lab session.
4. Make a careful transcription of your selected excerpts at Level 5.10 (Intonation), and bring it with you to the Lab Tutorial.

At the Lab Tutorial, do the following:

5. Using PRAAT, make a pitch curve of one or more of your selected examples. How does the pitch curve correspond to your prior impressionistic analysis and transcription?
6. Make some other relevant PRAAT analyses of your examples, e.g. an amplitude curve.
7. What do these two instrumental analyses add to your understanding of the intonational patterns you are hearing?
8. Conversely, how does your previous “impressionistic” listening to, and interpretation of, the intonation contours inform your understanding of the instrumentally-derived acoustic information?
9. Can you identify any “mistakes” that PRAAT has made in deriving a pitch curve? How is it possible for you to tell that the program has made a mistake?

After the Lab Tutorial, write up the following:

10. Make a final revision of the intonational transcription (Level 5.10) of your excerpts, incorporating what you have learned in doing your instrumental analysis.
11. Include some presentation of the PRAAT analysis of pitch curves for several intonation units, or other relevant instrumental analysis.
12. Briefly discuss how the analysis you have done relates to the pragmatic function in discourse of the intonational contours you present. Here you may choose whether to draw more on your impressionistic analysis or your instrumental analysis.
13. The sum of your transcription, instrumental analysis, and discussion should be 2-3 pages.

[rev. 6-Sep-2005]

Project 6

Folk Transcription Practices

Objective

The goal of this project is to find an example of a “folk” transcription practice, and figure out how it works. This will allow us to talk about:

14. how spoken discourse is represented in visible form
15. how the different ways of representing spoken discourse are designed to serve the goals of the intended users of the representation

This is a bit of field work – or a scavenger hunt. You will be looking for an example of how other people represent spoken discourse on paper (or on a computer screen), and thinking about how it is done, and why.

Background

The act of representing spoken discourse is performed by any number of professions that deal with spoken language, such as cartoonists, magazine interviewers, newspaper reporters, court reporters, playwrights, screenwriters, novelists, and many others. Each group represents spoken discourse according to its purposes, and the result is a wide range of different systems, conventions, and practices for representing spoken discourse.

For this exercise we are going to avoid the professional academic researchers (e.g. linguists, sociologists, anthropologists, etc.) who transcribe spoken discourse in order to study it. Rather, we are looking at what we might call folk traditions of representing spoken discourse, which are typically used for other goals than studying discourse as such. (What linguists and other academic researchers do when they transcribe discourse is of course an interesting topic in its own right, but this is a topic for another kind of investigation.)

Procedure

Go out and find:

- one example (per student) of a folk transcription, which exemplifies...
- some recognizable practice for representing spoken discourse in visual form.

You may find your example in a comic book, newspaper, magazine, on the Internet, or anywhere you like. The only requirement is that it be authentic (don’t make it up yourself), and that it include a visual representation of spoken discourse, as done by somebody *other* than linguists (or other such academic researchers).

You will want to compare your example with those found by other participants in the course. You may want to team up with a specific partner to do this, and present your findings jointly as a pair. The two examples you select should differ from each other in some interesting way: with respect to their goals, conventions, and practices, for example. They can be drawn from any of the professions listed above, or

from any other you can think of—use your imagination. In general, it is more interesting to cast a wide net when thinking about folk traditions of transcription.

Your examples should be substantial enough to give an idea of what the systems are like. That is, each of your two examples should probably include at least a few dozen words of represented speech. You can use any language you know well enough to analyze.

For each of the two transcription samples and their (implicit) transcription practices, consider the following questions.

- (10) What is the overall purpose or goal of the transcription practice?
- (11) Who is the intended audience, the intended users of the transcriptions?
- (12) What is the difference between the represented discourse and the representing discourse? What is included and what is omitted from the transcription? How did the transcriber abstract away from the original discourse being represented in order to arrive at the transcription?
- (13) Identify a few interesting conventions or practices used in the transcription system. How do these transcription categories, symbols, or practices serve their intended purpose?

Prepare a *brief* (five minutes per presenter) and very simple class presentation about *selected aspects* of each of your two transcription systems, based on the above questions. Time your presentation so that you will be able to say what you need to say within the allotted time. Make a one-page handout, with facsimile reproductions, for each of the sample transcriptions. Just reproduce (e.g. photocopy or print out) your example as you found it, do not retype! You may wish to annotate your examples with arrows or numerical indexes in order make it easier for you to point out selected features when you talk about them in your presentation. And you may wish to include on your handout a few of the analytical points you are going to make. We will take a few minutes for questions and discussion after your presentation.

So that the examples you bring in may be added to an archive of folk transcription examples (available to any course participant), be sure to give:

- the source for each transcription, with *full bibliographical reference*
- if you found your example on the Internet, include the website's URL
- your name (as the collector, labeled "Collected by...")
- the date of publication or collection (e.g. the date you consulted the website)

NOTE: Be sure to bring enough photocopies of your handout to give to all course participants.

Some Observations

The following observations are offered for your consideration. If you find them useful, that's fine; if not, just ignore them.

Notice that, in order to have a representation of discourse, there are actually *two* distinct discourse events involved: first, the *represented* discourse event, and second, the *representing* discourse event (cf. Jakobson).

represented discourse event <=====> *representing* discourse event

The *represented* discourse event is (in general) “what really happened”, that is: some people talked. The *representing* discourse event is a later attempt to depict the first event in written form.

Similarities between the two discourses are to be expected, given that one seeks to depict the other, but the differences are also instructive. Normally the represented discourse event is an act of speaking of some kind (conversation, interview, speech, etc.) which takes place on some definite occasion. The representing discourse is usually created later, on the basis of, but in a different medium from, the represented discourse. The transcription is generally created, and interpreted, in accordance with certain representational conventions, which may be well-established or ad hoc, explicit or implicit, conscious or unconscious. The representing discourse attempts to depict some aspects of the represented discourse, but never all its aspects. What is included and what is left out can tell us a lot about what the transcription is trying to do.

It might seem that in the case of theatrical productions such as screenplays or film scripts, transcription as such is not involved, since the written version of the discourse often exists before the spoken version, and the written may in fact be taken as a recipe for performing the spoken. To be sure, this is a crucial difference in intention, and one could debate whether the word transcription is appropriate for such cases. But the fact remains that many of the same issues are involved: conventions for representing one medium by way of another medium, decisions about what to include and what to exclude, and so on.

[rev. 6-Sep-2005]

Project 7

Transcription System Comparison

Objective

The goals of this project are:

- (1) to make a systematic comparison of two distinct transcription systems, in such a way that each system sheds light on the other
- (2) to think about which new or alternative transcription categories, representations, and practices you may want to borrow from, and incorporate into your future discourse research

Procedure

1. Select a transcription system to analyze. This should be a system which: (a) is used by academic researchers (from any field) for doing research on the properties of spoken discourse (construed broadly), and (b) differs substantially from the Discourse Transcription system used in this course (Du Bois et al., 1992, Du Bois et al., 1993, Du Bois 2005).
2. Each course participant must present a different transcription system. First come, first served.
3. To get the broadest coverage of alternative systems for discourse transcription, we will choose only *one system per group* of similar transcription systems as listed below (unless there are enough course participants to go back for a “second round” of picks).
4. You may propose to present a transcription system other than those listed below, subject to prior approval by the instructor.
5. Having selected a transcription system (and secured the sole rights to present it!), read about your system. Learn about what it can do, and how you can use it.
6. Take a half-page of your own conversational transcription, and retranscribe it in the new system. For some transcription categories, you will simply substitute different symbols for what you have previously written. For more profound differences between the systems, you will have to revisit your recording in order to retranscribe it anew, interpreting what you hear and/or see in terms of the new or different transcription categories employed by your selected transcription system.
7. Make a handout, consisting of no more than two pages. The first page of your handout should present your transcription in two versions. On the top half of the page, present your own transcription, as you originally made it using the familiar Discourse Transcription format. On the bottom half of the page, present your re-transcription of the same excerpt, using the new system.
8. Include line numbers for both transcriptions. To make it easier to refer to specific points of interest, you may want to include labeled arrows, footnote markers, or other such annotation.
9. On the second page of your handout, present a brief comparison of the two systems. You should include:
 - a) a key, explaining any special or unfamiliar symbols used for transcription categories, with their meaning/function
 - b) a comparison in terms of differences, e.g. additional transcription categories included in the other system, transcription categories or distinctions not marked, matters of display, iconicity, format, practice, theoretical motivation, etc.
 - c) summarize what you see as the pros and cons for each system
 - d) give any conclusions you may have, especially about what you consider worthwhile to adopt as a transcribing practice, based on your encounter with the new system
10. Prepare a recording excerpt containing just the portion of the recording corresponding to your half-page of transcription. Bring the file to class in a portable, computer-readable format (e.g.

- a .WAV file or video file, as appropriate, on a CD).
11. Make a brief presentation to the class about your comparison of the two transcription systems, using your handout. Be sure to play the recording so that everyone can see and hear what the new system is about.

Theory

In comparing different transcription systems, it is useful to have a theoretical framework to guide you. Works which present a theoretical framework for designing and evaluating systems of transcribing spoken discourse include (Du Bois, 1991, Edwards, 1993, Ochs, 1979). You should read at least one or two of these articles before you begin this project, and make reference to the general theoretical considerations they introduce in presenting your evaluation of your transcription system.

Transcription Systems

The following are some transcription systems of interest. The “focus” section indicates what are the primary concerns driving the transcription system, based primarily on the theoretical interests of the transcribers and their audiences. For each transcription approach, one or two seminal articles or books are given which describe the system.

Table 1. Some systems for transcription of discourse

	System	Focus	References
A 1.	Conversation Analysis	talk in interaction	(Jefferson, 2002, 2004)
A 2.	Conversation Analysis	gaze, body, talk in interaction	(Goodwin, 1981)
B 3.	Gumperz	Contextualization cues, prosody	(Gumperz and Berenz, 1993)
C 4.	HIAT	timeline, score notation	(Ehlich, 1993)
C 5.	Overlap markup	sequence, score notation	(Meyer, Morris, & Blachman 1994)
C 6.	EXMARaLDA	annotation graph, score notation	see URL: Schmidt, Thomas
D 7.	Ochs	child language, theory	(Ochs, 1979)
D 8.	CHAT (CHILDES)	child language, general	(MacWhinney, 2000)
E 9.	London-Lund corpus	prosody	(Svartvik and Quirk, 1980) 9-25, 152ff
E 10.	London-Lund corpus	prosody	(Peppé, 1995)
E 11.	London Teenagers	prosody, interaction	(Haslerud and Stenström, 1995)
F 12.	British tradition	intonation (as contours)	(Cruttenden, 1986) ch. 3
F 13.	Bolinger	intonation (iconic meaning)	(Bolinger, 1986)
F 14.	French	intonation	(Blanche-Benveniste and Jeanjean, 1987)
F 15.	INTSINT	intonation (phonetic)	(Hirst and Di Cristo, 1998)
G 16.	ToBI	intonation (phonological)	(Beckman and Ayers, 1994)
H 17.	TEI	computer formatting	(Johansson, 1995)
H 18.	Annotation Graphs	computational representation	(Bird and Liberman, 1999)
H 19.	MATE	computational representation	see URL

Table 2. Key to Grouping of transcription systems

Group
A Conversation Analysis
B Contextualization cues
C Score
D Child language
E Survey of English Usage
F Intonation as contour
G Intonation as tone sequence
H Computers

Table 3. Web sites for selected transcription systems

System	Web site URL
8. CHAT (CHILDES)	http://childes.psy.cmu.edu/manuals/CHAT.pdf
9. ToBI	http://www.ling.ohio-state.edu/~tobi/ame_tobi/
10. Annotation Graphs	http://www ldc.upenn.edu/Papers/DTAG1999/dtag.pdf
11. MATE	http://www.ims.uni-stuttgart.de/projekte/mate/mdag/
12. EXMARaLDA	http://www1.uni-hamburg.de/exmaralda/index-en.html

References

- Beckman, Mary E., and Ayers, G.M. 1994. *Guidelines for ToBI labeling. Version 2.0*. Columbus, OH: Department of Linguistics, Ohio State University.
- Bird, Steven, and Liberman, Mark. 1999. Annotation graphs as a framework for multidimensional linguistic data analysis. In *Towards standards and tools for discourse tagging, Proceedings of the workshop*, 1-10: Association for Computational Linguistics.
- Blanche-Benveniste, Claire, and Jeanjean, Claire. 1987. *Le français parlé: Transcription et édition*. Paris: Didier.
- Bolinger, Dwight L. 1986. *Intonation and its parts*. Stanford: Stanford University Press.
- Cruttenden, Alan. 1986. *Intonation*. Cambridge: Cambridge University Press.
- Du Bois, John W. 1991. Transcription design principles for spoken discourse research. *Pragmatics* 1:71-106.
- Du Bois, John W., Cumming, Susanna, Schuetze-Coburn, Stephan, and Paolino, Danae. 1992. Discourse transcription. *Santa Barbara Papers in Linguistics* 4:1-225.
- Du Bois, John W., Schuetze-Coburn, Stephan, Cumming, Susanna, and Paolino, Danae. 1993. Outline of discourse transcription. In *Talking data: Transcription and coding in discourse research*, eds. Jane A. Edwards and Martin D. Lampert, 45-89. Hillsdale, NJ: Erlbaum.
- Du Bois, John W. 2005. *Representing discourse*. Ms., Linguistics Department, University of California, Santa Barbara. Santa Barbara.
- Edwards, Jane A. 1993. Principles and contrasting systems of discourse transcription. In *Talking data: Transcription and coding in discourse research*, eds. Jane A. Edwards and Martin D. Lampert, 3-31. Hillsdale, NJ: Erlbaum.
- Ehlich, Konrad. 1993. HIAT: A transcription system for discourse data. In *Talking data: Transcription and coding in discourse research*, eds. Jane A. Edwards and Martin D. Lampert, 123-148.

- Hillsdale, NJ: Erlbaum.
- Goodwin, Charles. 1981. *Conversational organization: interaction between speakers and hearers*. New York: Academic Press.
- Gumperz, John J., and Berenz, Norine. 1993. Transcribing conversational exchanges. In *Talking data: Transcription and coding in discourse research*, eds. Jane A. Edwards and Martin D. Lampert, 91-121. Hillsdale, NJ: Erlbaum.
- Haslerud, Vibecke, and Stenström, Anna-Brita. 1995. The Bergen Corpus of Longdon Teenager Language. In *Spoken English on computer: Transcription, mark-up and application*, eds. Geoffrey Leech, Greg Myers and Jenny Thomas, 235-242. Harlow: Longman.
- Hirst, Daniel, and Di Cristo, Albert eds. 1998. *Intonation systems: A survey of twenty languages*. Cambridge: Cambridge University Press.
- Jefferson, Gail. 2002. Is "no" an acknowledgement token? Comparing American and British uses of (+)/(-) tokens. *Journal of Pragmatics* 34:1345-1383.
- Jefferson, Gail. 2004. Glossary of transcript symbols with an introduction. In *Conversation Analysis: Studies from the first generation*, ed. Gene H. Lerner, 13-31. Amsterdam: Benjamins.
- Johansson, Stig. 1995. The approach of the Text Encoding Initiative to the encoding of spoken discourse. In *Spoken English on computer: Transcription, mark-up and application*, eds. Geoffrey Leech, Greg Myers and Jenny Thomas, 82-98. Harlow: Longman.
- MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk, 3rd edition. Vol. I: Transcription format and programs*. vol. 1. Mahwah, NJ: Erlbaum.
- Meyer, Charles F., Morris, Robert A., and Blachman, Ed. 1994. Can you see whose speech is overlapping. *Visible Language* 28:110-133.
- Ochs, Elinor. 1979. Transcription as theory. In *Developmental pragmatics*, eds. Elinor Ochs and Bambi B. Schieffelin, 43-72. New York: Academic Press.
- Peppé, Sue. 1995. The Survey of English Usage and the London-Lund Corpus: Computerising manual prosodic transcription. In *Spoken English on computer: Transcription, mark-up and application*, eds. Geoffrey Leech, Greg Myers and Jenny Thomas, 187-202. Harlow: Longman.
- Svartvik, Jan, and Quirk, Randolph eds. 1980. *A corpus of English conversation*. Lund: Gleerup.

[rev. 27-Sep-2009]

Project 8

Politics, Ethics, & Orthography in Transcription

An often-overlooked fact about transcription is that it affects people's lives—in particular, the lives of the people whose speech is being represented in the transcription. Often the effects may seem minor or even negligible, but at times the impact can be surprisingly powerful. For this reason, transcription necessarily invokes both ethical and political issues. Many of these repercussions tie in with more general issues in the ethics and politics of how people create representations of other people. This is because transcription creates a representation of one of your most important attributes: your speech. To the extent that your speech is tied up with your identity, its representation may become a lightning-rod for ethical and political sensitivities.

While ethics and politics can touch on almost any aspect of transcription, a particularly sensitive point concerns the seemingly simple matter of spelling: how a speaker's words are spelled. Because the treatment of spelling has often been a point of heated contention, a number of authors have come to recognize the connection that spelling has with the ethics and politics of transcription.

Some key writings on ethics and politics of transcription are listed below. The entire special issue of the *Journal of Sociolinguistics* edited by Jaffe (Vol. 4, no. 4, 2000) is of particular interest.

The following are some key points to consider in evaluating and debating the ethics and politics of representing discourse:

1. Why does transcription matter so much to people's lives?
2. How does spelling participate in creating representations that impact the represented speaker's perceived identity?
3. What other aspects of transcription (or "representing discourse") raise ethical and/or political issues?
4. How does the transcription's target audience (professional scholars, college students, general public, potential voters, the community of represented speakers, and so on) affect the ethical issues?
5. What about "unintended overhearers"? What happens if the transcriptions are circulated outside the originally intended target audience?
6. What are the ethical and political responsibilities of transcribers?
7. As an exercise, create two transcription versions of one stretch of conversation. The two transcriptions should reflect substantially different ways of representing the speech of the participants (for example, with regard to spelling of non-standard speech), as reflected in the typical practice of two different approaches to transcription. What are the pros and cons of each transcription, in terms of (1) ethics and (2) science?
8. Who has the right to decide how a person's speech will be represented?

References

- Bucholtz, Mary. 2000. The politics of transcription. *Journal of Pragmatics* 32:1439-1465.
- Preston, Dennis. 1982. 'Ritin' fowklower daun 'rong: Folklorists' failures in phonology. *Journal of American Folklore* 95:304-326.
- Preston, Dennis. 2000. Mowr and mowr bayud spellin': Confessions of a sociolinguist. *Journal of Sociolinguistics* 4:614-621.
- Jaffe, Alexandra, and Walton, Shana. 2000. The voices people read: Orthography and the representation of non-standard speech. *Journal of Sociolinguistics* 4:561-588.

[rev. 6-Sep-2005]

Project 9

Concordance Analysis

Objectives

The objective of the concordance project is to learn how to use concordance software to:

- ◆ search spoken discourse data for particular words, patterns, and contexts
- ◆ create concordances of spoken discourse data
- ◆ identify basic quantitative generalizations about spoken discourse data

1. Preparing the transcription

You should arrive at the tutorial with a version of your transcription file (in computer format) already prepared in a format suitable for computer analysis. Concordance software typically operates only on “plain text”, which may be in either ASCII (DOS) format or ANSI (Windows) format. It is necessary to first save your data in this format in order to use MonoConc. (See Appendix A.16 on “Preparing a Transcription for Computer Analysis”.)

2. Starting the concordance program

- a) The concordance program we will use is called MonoConc Pro. This program makes concordances of discourse data (whether spoken or written), assuming the data are in appropriate computer-readable format (see previous section).
- b) To start MonoConc Pro, double-click on its icon. (For example, double-click on the icon labeled **MP2.2** [or equivalent]. Or select **Programs** and then click on **MonoConc Pro**. (If you can't locate the program, ask for help.)
- c) From the MonoConc menu, select **File/Load Corpus File(s)**. When the dialog box appears, find the folder with the data you are interested in. (For details, see Section 6 below on “Copying the data.”) Highlight all of the files in the folder, by dragging the mouse over all of them. (Or do “Select All” by typing <CTRL>-a.) Then click **Open** to open the files for searching.

3. Setting up MonoConc: Highlights

- a) From the menu, select **Concordance/Search Options**. Specify the following features:
 - ◆ Max search hits: **32000**
 - ◆ Context type: **Characters** Size: **150**
 - ◆ Special characters: **&** matches between **0** and **1** words.

4. Skipping characters and delimiters

The following are suggested initial *default values* for skipping characters and delimiters for the published transcriptions in the Santa Barbara Corpus of Spoken American English (e.g. for transcriptions using **Du Bois et al. 1993** conventions [=DT1]). In MonoConc, select **Concordance/Search Options**. Then type the characters into the appropriate spaces in the dialog box. (Do not type any spaces, neither between the characters, nor at the beginning or end.)

Characters to treat as delimiters: - [hyphen]

Skipping characters: 0123456789~¬?! ,# + [] @ _ % ^ ` = / . \

Choosing the skipping characters

For certain kinds of searches, you may want to experiment with adding or removing a character from the list of delimiters, or from the list of skipping characters. The following are a few examples of characters that are sometimes good to include, and sometimes good to exclude (this is not an exhaustive list):

Table 1. Characters to be excluded from some searches

Character	Meaning	Exclude when searching for...	Include when searching for...
, . ?	intonation	words	intonation, punctuation
...	pause	words	pause
[]	overlap	words	overlap, simultaneous speech
:	length	words	prosodic lengthening
@	laughter	words	laughter, laughing while speaking
%	creak	words	creaky voice
-	(hyphen)	most words	compound words, truncation

Characters to include always

Note that some types of punctuation symbols are okay to include in all searches; that is, they are *not* recommended for exclusion from most searches. For this reason, they should not normally be placed on either the delimiter or the skipping list. These are the symbols which are used to represent the following transcription categories:

Table 2. Characters to be included in all searches

Character	Description	Meaning	Include when searching for...	Irrelevant for...
'	apostrophe	contraction, elision	words (<i>can't, don't, etc.</i>)	
()	parentheses	vocalism	vocalism	word search
()	parentheses	pause timing	pause, event duration	word search
;	semi-colon	speaker attribution [DT2]	speakers, turn beginning	word search
:	colon	speaker attribution [DT1]	speakers, turn beginning	word search*
< >	angle brackets	manner/quality	manner, quality	word search
—	en dash	truncated word	truncation	word search
—	em dash	truncated intonation unit	truncation	word search

[*The one exception or special case concerns the colon. In DT2 this indicates prosodic lengthening of sounds, so it *should* be a skipping character, because it occurs inside words; but in DT1 it indicates speaker labels, so it *should not* be a skipping character. It never occurs inside words in DT1, and it may be useful for searches that target turn boundaries or speakers.]

The colon character is treated differently depending on what it stands for. For transcriptions in

which colon represents a participant attribution, it should be left *off* the delimiter and skipping character lists. For transcriptions in which the colon represents lag or prosodic lengthening, it would normally be placed *on* the skipping character list.

Note that it is quite possible that at some point you will want to search for words spelled as contractions, vocalisms, participant attributions, manner/quality features, and so on. If so, you will need to include the characters in Table 2 as part of your search (i.e. maintain them as options available to you in your repertoire of searchable signs). And since they do not normally interfere with ordinary word searches (because they rarely or never occur inside words in a DT transcription), they do not normally need to be excluded (e.g. by listing as a skipping character).

5. Search

Once you have set up these features, you are ready to begin working. Follow the suggestions in the Appendix titled “**Using a Concordance for Spoken Discourse Research**”. (If you don’t have your paper copy with you, you may be able to locate an electronic copy of the file, in Word format, in the folder for **\Programs\MonoConc.**) You can return to the present document after you have worked through the cited Appendix.

6. Copying the project data

In order to have some corpus data to analyze, it may be necessary to copy it into a directory for your own use.

Before the Lab Tutorial begins, you will need to have made your own separate electronic copy of the corpus data, which you will use during the session. Because MonoConc does not allow multiple users to access the same data at the same time, you will need to copy several transcription data files into your own working folder first. You should do this IN ADVANCE of the Lab Tutorial on concordances. Do this as follows:

- d) The main source of data we will be using is located in the folder labeled **Ling212** (or **Discourse**, or equivalent), under the following folders: **\Data\English\SBCorpus**. This folder should in turn contain folders labeled **\TextOnly** and **\TextTime**. We will mostly use the data files in **TextOnly**, but you should copy both folders, as follows.
- e) Create another new folder on your own drive, label it **SBCorpus**, and then copy all the necessary files into it.
- f) Create a new folder on your own drive, label it **ClassCorpus**, and then used the mouse to drag all the necessary files into it (i.e. transcription files prepared by your fellow students in the class).

7. Further topics

The following are more specialized topics, which may be of interest for more advanced users.

Stoplist

The previous point remains valid even if you want to *exclude*, for example, all participant attribution labels. For some purposes, you may wish to create a “stoplist” of all the participant attributions (speaker labels), which excludes them from such things as the collocation frequency counts. But to do this, you will need to have the participant attribution symbol available to you, in order to allow you to specify the distinction between the attribution label and an actual pronunciation of the same name by a speaker in the conversation itself. Similarly, you should normally keep truncated words separate from full words, to avoid confusing a truncated word with a full word. The truncation symbol must be included in your search repertoire to allow such a distinction.

Timestamped files

- a) For some purposes, you may wish to do a concordance of transcriptions containing timestamps. This will give you the information you need to go from an example you find in your concordance search to the corresponding audio, using the “Go to” function in a program like Sound Forge (see next section).
- b) To include the timestamps in your concordance searches, first **unload** the “TextOnly” versions of the corpus files that you loaded when you started MonoConc, by selecting from the menu **Files/Unload Corpus**.
- c) Then go to the following folder: `\Data\English\SBCorpus\TextTime`. Select all of the files. Note that having the timestamps in the file means you may have to design some of your concordance searches differently.
- d) For the (optional) exercise on “**Regular Expressions**” (see ****handout**), many of the searches have been designed with the timestamped version of the corpus files in mind, as found in `\Data\English\SBCorpus\TextTime`. (Of course, a “regular expressions” search can also be used with non-timestamped files—once you're familiar with the technique, you can design your own queries..)

8. Project

In following the tutorial in the Appendix on “Using a concordance for discourse research,” you will have tried out many different searches of the spoken discourse data, using the concordance program MonoConc Pro. Your experience in making these concordance searches (and interpreting the results) may have given you additional ideas about other searches you could do.

Based on your concordance searches as part of this tutorial, write up a brief description of one or more *linguistically interesting* searches you can do with this kind of tool. How can the concordance results be used to pursue linguistic research? Discuss how this kind of tool could be useful to someone who wished to explore the grammatical and/or discourse patterns further. You may especially wish to comment on how specific aspects of the representation of *spoken* discourse (e.g. transcription symbols for turns, speakers, intonation units, pauses, overlap, lengthening, laughter, etc.) can be exploited in your concordance research.

Note that the focus here is not on discovering original research results, but on learning about the value of concordances as a potential tool for your future research.

Write this up in a brief report (NOT a paper), which should include at least a few examples of results turned up by your concordance search.

[rev. 27-Sep-2009]

Project 10

Data Publication

Objectives

The goal of this project is to help you learn how to publish your data on the Internet (or on CD-ROM). This means disseminating to a broad audience the following items:

- ◆ your time-aligned transcription
- ◆ your recording
- ◆ the ethnographic information that users will need to interpret the transcription and recording

Among the reasons for publishing your data are:

- ◆ disseminating and preserving the knowledge represented in your transcription
- ◆ gaining experience in integrating transcriptions with sound and/or video in a web format
- ◆ learning about the web publication process
- ◆ getting credit for publication

Preparation

Prepare the following items:

- ◆ final version of your time-aligned transcription of your Transcription Segment
- ◆ recording file for your Transcription Segment
- ◆ recording file for your Corpus Segment (checked for anonymity issues)
- ◆ ethnographic sketch
- ◆ other ethnographic data, if appropriate (and checked for anonymity)
- ◆ put all text files in web format (Open Document text file [.odt] and/or Adobe Acrobat file [.pdf]). (You can use the free open-source software OpenOffice.org to do this.)

Arrange with other class participants to group these together in the appropriate folders on the Linguistics Lab computers, ready for publication.

Anonymity

It is important that you take appropriate measures to safeguard the anonymity of the speakers on your recording. This means you should check your data before publishing it, to make certain that any identifying information (such as participants' last names, addresses, telephone numbers, etc.) are:

- ◆ eliminated from your transcription, and substituted with appropriate pseudographs
- ◆ eliminated from your ethnographic sketch and other accompanying demographic information, and substituted with appropriate pseudographs
- ◆ bleeped from the audio of the recording file versions to be published

The precise items to be suppressed will vary according to the consent agreements you have with the participants on your recording. (Note that not all names in your conversation are necessarily at issue, only those of the participants you have recorded.)

In some cases, the consent agreement you have with the individuals on your recording may preclude publication to the web. In this case, you should publish your data on CD-ROM.

Product

Publish the full version of the Class Corpus on the Linguistics Department network, the Internet, and/or to CD-ROM.

[rev. 27-Sep-2009]

1 However, most descriptions of transcription systems for discourse or conversation do not provide any detailed account of the delicacy issue. As far as I am aware, the present presentation is the most detailed one, perhaps the only one.

2 A more nuanced discussion of turns will be presented in the chapter on Sequence.

3 For presentation transcripts, it may be possible to rely on the vertical alignment, and thus dispense with the numbered bracketing. However, for database transcriptions, it is preferable to index overlap brackets by numbering them, to avoid unrecoverable disruption of overlap information during editing.

4 In addition to these audible cues, Intonation Unit boundaries tend to show significant correlations with various non-audible but overtly observable behavioral cues including gesture, gaze, and other aspects of body behavior. Little is known about these correlations with Intonation Units at present, but the issue is ripe for investigation.

5 Correlations of Intonation Unit boundaries with non-phonetic features such as word quantity, information quantity, information structure, syntactic structure, and so on, are of considerable interest in their own right. But they do not represent *audible* cues for Intonation Units as such.

6 In addition to these audible cues, Intonation Unit boundaries tend to show significant correlations with various non-audible but overtly observable behavioral cues including gesture, gaze, and other aspects of body behavior. Little is known about these correlations with Intonation Units at present, but the issue is ripe for investigation.

7 Correlations of Intonation Unit boundaries with non-phonetic features such as word quantity, information quantity, information structure, syntactic structure, and so on, are of considerable interest in their own right. But they do not represent *audible* cues for Intonation Units as such.

8 Some transcription systems do use a distinct notation for overlapping speech if the two speakers start at the same time, e.g. {Atkinson, 1984 #983}; the motivation for doing so is unclear.