Beat Generation:
Utterance length and how it changes throughout conversation

Eric Eisenberg

Swarthmore College

Abstract

In linguistic literature, little attention has been paid to quantifying UTTERANCE LENGTH, or the number of words or other units of length used by a speaker per turn, through the course of conversation. Using a playwriting exercise by Philadelphia playwright Bruce Graham as a window into the issue, this paper discusses the how one can go about studying such a surprisingly complex issue. As observed in an original experiment eliciting two-person, spontaneous conversations from college students, Graham's template for utterance length proves inaccurate. Instead, instances of a multi-turn pattern of utterance lengths, given the name BEAT, emerge. When the speakers' beats are analyzed through the course of dialogues, they outline a beat-level MACROTURN-TAKING SYSTEM that shares many of the features of the utterance-level TURN-TAKING SYSTEM described in linguistic literature. Using the concepts of beats and macroturns, I construct a new template for utterance lengths throughout the course of a twenty-five-turn dialogue that more accurately reflects real-life speech than Graham's template.[1]

Table of contents

---

[1] This paper was completed with the help and input of Donna Jo Napoli, David Harrison, Jack Hoeksema, Eliza Cava, and Matthew Woodbury, for which I am very grateful.

Note to the reader:

Although this is a Linguistics thesis, I have taken special care to consider the playwright's interests even at the expense of linguistic ones: I used English orthography for linguistic examples for the most part instead of IPA, eschewed references to linguistic literature unless I could explain the relevant points in the paper, and tailored my analysis for the playwright's needs. I suggest that playwrights and casual readers read sections 1.1 and 1.3-15 to get a sense of the background of the project; skip directly to sections 2.4, 4.11 and 4.15 in order to understand what my analysis entails; skim section 5, especially sections 5.1 and 5.34, to understand how to look at the graphs; and finally read sections 6-7, containing the results, analysis, and conclusions that might be of interest.

The other portions of the paper explicitly detail my reasoning, methodology and opinion of what a study of the type I have presented here should adopt as relevant considerations. These are targeted to the linguist or language-enthusiast interested in carrying out a similar study or interested in exploring the complexity of the question in full. As far as I can tell, no one has ever

attempted to examine the lengths of utterances over the course of a conversation for the same reasons and to the same extent that I have. Thus I have included material not strictly necessary to understand the results and conclusions – discussions of interview strategies one should reject, description of my transcription conventions, etc. – as a reference for further projects in the field. I intended the experimental design, data-structures, and analysis presented below as a jumping off point for further, more formal or narrow research, not as the be-all-end-all of linguistic inquiry into the matter.

1. Introduction and Background

1.1 A playwright's exercise

The impetus for this project came from a playwriting exercise created by Bruce Graham, a local Philadelphia playwright. Graham has received numerous accolades, awards, and grants for his work in theater, film, and television, and teaches both film and theater courses at the University of Pennsylvania and Drexel University (Graham 2005).

The exercise consists of a dialogue between two characters, whom I shall call A and B, concentrating on the rhythms created by their speech. The writer numbers the lines on a page from one to fifty. When the writer fills in the words spoken by each character, the characters should alternate lines; i.e. character A speaks all of the odd-numbered lines and character B all of the even-numbered lines. Graham then provides a template for the number of words each character may speak per line, stipulating only that contractions count as a single word. The course of the characters' minor conflict or disagreement, as Graham called it, is mapped out in figure 1.

Figure 1: The number of words per line in Graham's playwriting exercise (1995)

Lines 1-20: 2-4 words	Lines 37-38: 20 or more words
Lines 21-30: 1-2 words	Lines 39-45: 4-6 words
Lines 31-36: 1 word	Lines 46-50: 1-2 words

Graham explains that the point of this exercise is to choreograph speech rhythms between two distinct voices. "Since plays are meant to be heard and not read, rhythms are extremely important. They establish mood, built [sic] tension, heighten conflict and expand characterization" (Graham 1995). The template Graham provides is supposedly not mere fabrication for the sake of interesting rhythm, but rather one derived from observing conversations in real life.[2]

Despite this, upon reading the example scene Graham provided, which followed the template above, I was struck by how stagy the dialogue sounded. That is to say, though it had a characteristic rhythm, it did not sound "natural" to me in the way that everyday speech might. So, I began to listen more carefully to two-person conversations around me – my own conversations, conversations between my friends, exchanges I happened to overhear – and sure enough, I noticed discrepancies between what I heard and what Graham had proposed.

The most striking difference to note was that Graham's template outlines a BALANCED exchange between speakers, with one speaker using roughly as many words as the other at all times. It seemed to me that conversations are usually UNBALANCED, with one person using far more words and the other responding with short replies. One speaker takes a turn monopolizing the conversation in the short term, and then the speakers switch, so that the other person can have a say. Such a pattern is absent in Graham's template. This is not to say that I thought no two people would ever speak with balanced turns: formulaic speech seemed to follow such a tack. An example is illustrated in figure 2, below.

---

[2] I have no source for this claim other than Donna Jo Napoli's (p.c. September 2005) recollection of the presentation of the exercise. However, the question of whether the exercise reflects actual conversation remains even if Graham did not explicitly design it to do so. Graham is a decorated and distinguished artist – he would not intentionally create dialogue that sounded fake or forced. His exercise therefore acts as a case study of the perception of spontaneous speech on stage as compared to actual spontaneous speech.

Figure 2: Formulaic speech sample illustrating balanced speech patterns

A: Hi
B: Hey
A: How's it going?
B: I'm good. You?
A: Good. Nice day out.
B: Yeah, it's really nice.

On a related matter, Graham requires a five-line section (lines 31-36) in which each character uses only one word per line, which I found jarring. I could imagine such an exchange only if the characters were shouting names at one another. Perhaps in a play the characters might awkwardly interrupt one another and produce such choppy dialogue, but I did not believe that actual speakers would interrupt one another so thoroughly that five single-word lines in a row would occur. Lines 31-36 were a single instance of a larger problem: overall, Graham's template appeared to allow too few words per line to sound like natural speech.

Of course, these were just my impressions, based on a limited number of uncontrolled observations. The question remained: does Graham's template accurately characterize spontaneous conversation? If not, why not, and is it possible to construct another template that would be more accurate?

1.2 MEAN LENGTH OF UTTERANCE (MLU)

The topic at issue in Graham's exercise – the number of words used by each speaker per line – has been researched by linguists in some subfields under the heading MEAN LENGTH OF UTTERANCE (MLU). It also falls under the larger sociolinguistic category of TURN-TAKING, the study of how conversation is organized among speakers, discussed in section 1.3, below. Unfortunately, neither field has fostered linguistic experiments germane to the issue presented in this paper.

Surprisingly enough, there is a limited amount of literature on MLU, perhaps because of the difficulty in defining it (discussed below in section 4). However, two groups of linguists

have persisted in quantifying it: those who study child language acquisition, and those who study human-computer interaction with an eye to having machines converse with and understand a human speaker.

1.21 Child Language Acquisition

The former group, informed by Roger Brown's work in the 1960s, believe that the average length of a child's utterances while he or she is learning to speak is a more effective indicator of the amount of grammatical complexity he or she has acquired than age. Brown monitored the course of language development in three children, nicknamed Adam, Eve, and Sarah, in a long-term study that lasted for four years. His analysis centered, however, on only the first 8-10 months of the data, attempting to describe the early stages of language acquisition. Brown states that MLU outstrips age as a measure of language development because, while children acquire language at very different rates, for those in the early stages of language acquisition "...almost every new kind of knowledge increases [utterance] length" (Brown 1973, 53). Examining data from other linguists (notably De Villiers and De Villiers 1972) as well as his own, Brown ranked fourteen grammatical structures in order of their presumed complexity, which was determined by the order in which children acquired them. The correlation between age and these complexity rankings was only .68, while the correlation between MLU and the complexity rankings was .85. Age and MLU combined had a correlation of .92, showing that age does add a small amount of predictive power (Brown 1973, 273-4).

Building off this research, Brown demarcated five stages of first-language development, now famous in the field, based on MLU and the maximum number of morphemes present in any given utterance (called UPPER BOUND). His stages are summarized in figure 3, below.

Figure 3: Brown's Stages of Language Development (Brown 1973, 56)

	Target Value		Approximation Attained	
Stage	MLU	Upper Bound	Maximum Distance from MLU	Maximum Distance from Upper Bound
I	1.75	5	.31	2
II	2.25	7	.05	1
III	2.75	9	.25	1
IV	3.50	11	.20	1
V	4.00	13	.06	1.67

Brown's analysis stops at MLU = 4.0 (Stage V) because:

> ...by the time a child reaches Stage V...he is able to make constructions of such great variety that *what* he happens to say and the MLU of the sample begin to depend more on the character of the interaction than on what the child knows, and so the index loses its value as an indicator of grammatical knowledge. [Brown's emphasis] (Brown 1973, 54)

Quite rightly, then, there are no comparable studies of normal adults attempting to quantify the complexity of their grammars through consideration of MLU.

The child nicknamed Eve in Brown's study reached Stage V at the age of 26 months, and that the other children did so at about 40 months. In light of this fact it seems reasonable that adult conversation should contain an average utterance length that exceeds 4.0, barring some situation in which it would be preferable to keep one's utterances short. After all, adults have much greater facility with language than children under 5 years old, as a rule. If we assume that a speaker follows Graham's template, speaking the maximum number of words allowable per line and 20 words on line 37, speaker A (even-numbered lines) will speak 101 words in 25 lines, or 4.04 words/line. That does not appear to be significantly higher than a child who has just finished the early stages of language development; could this be a piece of evidence indicating that Graham's template is inaccurate? Since Brown's MLU counts morphemes such as the plural

marker *–s* and the past tense marker *–d* as separate morphemes, we cannot say. Adding the additional morpheme tallies to Graham's template might make the MLU of the characters considerably greater than the words per line figure cited above. Thus, Brown's data, although of interest – in particular, his method for counting morphemes will be discussed below in section 4 – cannot solve the problem for us.

1.22 Human-Computer Language Interfacing

The second group of linguists who have expressed interest in MLU devote their research to HUMAN-COMPUTER LANGUAGE INTERFACING. That is to say, they want to teach computers to understand and simulate human speech. The appeal of perfecting a computer program than can converse naturally with the user is obvious: instead of typing specific instructions or navigating complicated menus, one could simply talk to the machine to get it to perform the desired task. Of course, creating such a program is extremely difficult. But, in certain, pre-defined arenas, conversation may not be as complicated as the free-for-all that confronts the casual listener in most situations.

For example, customer service interactions may have rather well-defined limits outside which the conversation is unlikely to stray. Buying an airline ticket or retrieving information about a flight is such a task. A trio of Danish linguists (Dybkjær et al. 1993), in an effort to create a system capable of fielding phone calls to an airline, performed a series of simulations in which a human impersonated a computer response system with which test subjects would interact as they attempted to make flight plans. (This type of experimental design is comically known as "WIZARD OF OZ".) However, due to the processing constraints, the machine would be unable to handle utterances greater than 10 words long, and would function in real time only if the user's MLU was between 3 and 4. This was so in part because a longer sentence is more likely to contain syntactically-complex structures. More importantly, even without complex syntactic structures, a long sentence requires a more sophisticated system of parsing the text than

a short one, since the machine will need more strategies for grouping the correct words together. The experimenters therefore designed the responses of the Wizard (machine impersonator) to minimize the length of the users' utterances. At first, user MLU could be as high as 12, far in excess of the acceptable bounds. By the seventh iteration of the experiment, the Wizard's communication strategy was honed to the point where users' utterances were reduced to 1-2 words per turn: the machine literally asked all the questions, requiring only simple fill-in-the-blank responses from the user. The linguists behind this experiment claimed that conversations between travelers and their (human) travel agents usually settled into a similar pattern, with the agents speaking solely in questions and the customers solely in clipped responses. They cited no evidence in favor of this claim, however.

While the MLU data from the experiment above comes from adults, and so is closer to spontaneous conversation among fluent speakers, the scenario is so constricted that the findings are almost orthogonal to the current question. In consciously reducing the user's MLU the experiment above undermines the study of any spontaneous pattern that may exist in normal conversation.

There have been less restrictive studies, though. AT&T hired a team of five linguists to continue some previous research on machine comprehension of natural speech for its phone service needs (Gorin 2003). The research deals with call classification: the machine needs to understand what the customer wants well enough to transfer him or her to the appropriate support department. In order to design the speech-interpreter, they examined human-human service interactions. The previous research considered calls lumped by AT&T into a category called Operator Services. This included making telephone calls, stipulating whether the call would be collect or paid for with a calling card, and retrieving information about those calls (e.g. the cost per minute). These few, specific tasks leave little room for improvisation, and seem comparable to the airline ticket experiment above. The new system under research, however,

fielded questions about customer's phone bills (Customer Care). There are far more than a few questions a customer could ask about his or her bill, and there is no means of predicting what the customer will ask about. Thus, the new domain of research is far more spontaneous and complex than the old, in theory. As evidence of this greater complexity, the AT&T researchers pointed out that the MLU of speakers calling in for Customer Care was considerably higher, at 39 words per turn, than the MLU of speakers calling in for Operator Services, at 19. Furthermore, speakers tended to use a larger vocabulary in Customer Care calls than operator services calls.

The data from the phone service experiment pertains to conversations between two adults in which the subject is open ended, so it pertains to the issue of natural conversation. However, although the customer may have leeway in what to ask, the customer service paradigm is still quite distinct from everyday speech. Both the customer and the AT&T technical representative would be aware, for example, that the purpose of a Customer Care call is for the customer to describe his or her problem or question in as much detail as necessary. The customer is likely to have a premeditated speech, of sorts, to deliver to the representative; this may explain the seemingly-sky-high average of 39 words per utterance. The situation remains the same for the Operator Services calls, except that the speaker delivers a request instead of a concern. Seeing as the request is generally straightforward, the MLU drops compared to Customer Care calls, but still remains rather high.

Both of these situations deviate from normal conversation, in which the subject arises spontaneously, as does most of the dialogue. There is no predetermined convention of who should talk first, and there is not necessarily a predetermined objective to be completed, after which the conversation will terminate. This data, like the data presented beforehand, is also ill-suited to provide insight into whether Graham's template for two-person dialogue reflects natural speech.

1.3 The significance of utterance length

Having established that Graham's exercise points to an area never specifically addressed in Linguistic literature, it is worth the time to pause and consider if, perhaps, the length of a speaker's utterances is not a subject worth pursuing. What does studying such a thing tell us, and why is it interesting?

1.31 What utterance length might mean for the linguist

The easiest answer to the question is to say simply, "Because it's there." Linguists study the self-organizing system of human language as a window into human cognitive processes, so any aspect of language provides much-needed evidence of what might be happening inside the black box of the mind. Ideally, the amount of information that a speaker conveys per utterance should be as important as the way he or she pronounces each word, chooses which terms to use, inserts pauses or changes of intonation, etc. Linguists study all of the latter without hesitation, so there is no reason that they shouldn't allot time for the former.

More specifically, if there were rule-governed behavior regarding the utterance lengths in two-person dialogue, or specific templates that a conversation should follow, it might shed light on a few seemingly-unrelated branches of the analysis of language. For example, utterance-level intonation, including pitch variation, stress, and the like – typically referred to as INTONATION CONTOUR – could be a function, in part, of the length of the utterance (Prieto 2004, Wang and Hirschberg 1991).[3] If sociolinguistic factors such as the economic class or race of a speaker could influence the average length of the speaker's utterances in a given conversation, a causal link with predictive power might be found between such social factors and intonational patterns. Perhaps the notion of REGISTER, or the mode of speaking as dictated by social factors such as the

[3] Prieto actually argues that it is not utterance length but sentence-type that really affects intonation. However, her data shows that short utterances of only one pitch accent differ markedly from longer utterances, irrespective of sentence-type. Also, she uses number-of-pitch-accents as her measure for utterance length, and is not concerned with how those pitch accents may map to utterances of different lengths as measured by tokens or time. Wang and Hirschberg point out that intonational boundaries may be based on utterance length as measured by time.

relative age, power, and class of the speakers, could be attributed to something as concrete as a socially-prescribed minimum or maximum turn length for one or both speakers.

Obviously this is a speculative and probably unlikely possibility. On the other hand, the relative merits of such a position cannot be addressed unless one examines utterance length as a phenomenon in order to debunk it. We will never know unless we look into it.

1.32 What utterance length might mean for the author

Independently of the theoretical linguistic concerns surrounding utterance length, there are a number of authors who might very much like to know what spontaneous, natural speech sounds like. Dialogue makes up a significant percentage of most novels and certainly of most plays, and perceived skill or clumsiness with literary dialogue can be of great artistic and commercial consequence. I have heard my father, to name one, describe some of his favorite novel writers as "having a great ear," by which he is complimenting their natural-sounding dialogue. I assume that many readers feel similarly, and playgoers even more so.

I do not wish to argue the case that all dramatic dialogue (i.e., in plays) should necessarily match what one hears in life as closely as possible. As a budding playwright myself, I recognize that such an argument completely misses the point of drama: a play necessarily presents a version of reality that is more exciting, more interesting, or more vivid than life. Otherwise, why go see the play? You could get a bigger thrill from simply living for a couple of hours and save yourself the ticket price. The draw to come see the show is that it presents a "heightened" sense of reality, although sometimes abstracted so far from reality as to seem fantastical.

Nevertheless, if an author wishes to capture some essence of life in his or her work, it may be useful to know what real dialogue sounds like as a starting point. At the very least, it may serve as something with which to compare successful literary dialogue in order to answer the following question: if natural dialogue as it normally unfolds is not dramatically engaging,

what about it must be "heightened" for it to be effective? Perhaps Graham's exercise may be effective playwriting even if it turns out to allow too few words to be considered natural – or because it allows too few words (!) – as this requires that the action and the ideas come quicker, with less noise.

1.4 Addressing the question

I have discussed the fact that no one has attempted to create a conversational template à la Graham with utterance length as its major focus. Yet such a template and such a focus seem worthy of study. The best way to address the question appeared to be an original experiment that would elicit two-person, spontaneous dialogue approximating natural speech as closely as possible. Graham's exercise can be considered one hypothesis that the experiment tests. However, I questioned Graham's template, and hypothesized the following four generalizations, listed as figure 4. These hypotheses were suggested by my initial, informal observations of dialogue, and solidified by the background research presented above.

Figure 4: Hypotheses on the form of spontaneous, two-person dialogue

1) Utterance length will not, for the most part, be balanced:
   - 1a – For almost any given portion of the conversation, one speaker will employ considerably more words per utterance than the other
   - 1b – The distinction of having longer utterances in the short term will abruptly switch from one speaker to the other at points throughout the discourse
2) If utterance length is roughly equal between the two speakers over a given period, then the nature of the conversation will either be formulaic speech, or periods of emotional intensity for both speakers (e.g. name calling).
3) The MLU of each speaker over the course of an entire dialogue should be larger than 4 and smaller than 19, probably between 7-12.

1.5 Turn-taking

Much of the discussion of the results of this project relies on an understanding of turn-taking. I have decided to introduce the concept here in the introduction, rather than later, so that it will fall within one of the recommended sections for playwrights.

The study of TURN-TAKING, or how humans spontaneously organize their (speech) interactions, does not specifically address the issue of utterance length. Rather, it focuses on the transitions between speakers' utterances, and the emergent patterns associated with the system for affecting these transitions. If conversation is viewed as a long train consisting of many train cars, utterance length is the study of how long each car is, while studies of turn-taking examine the hitches that connect the cars to one another. Of course, the train metaphor simplifies things: speech in not composed of large, physical objects that cannot overlap, and train cars are all hitched in much the same way regardless of their size and function. Utterances, on the other hand, can overlap, interrupt one another, and otherwise interact in ways that physical objects cannot. And, there is no reason to assume that utterances of different forms or functions should interact with one another in exactly the same ways.

It is possible that, since an utterance's length is one aspect of its form, length of utterance could have effects on turn-taking between speakers. There is, however, a more pressing reason to describe some basic principles of turn-taking in the body of this paper than some far-flung chance that utterance length will affect it. As will become clear much later in section 6, many of the properties of the turn-taking system that govern how one utterance moves to the next may have correlates in a system of MACROTURN-TAKING that governs how groups of utterances by one speaker move to groups of utterances by the other speaker.

Three linguists, Sacks, Schegloff, and Jefferson (1974), provide an accessible and comprehensive introduction to turn-taking in conversation in their work, "A Simplest Systematics" (a shortened title). Their paper describes 14 observations about turn-taking in conversation that any proposed analysis should capture, and then sets out to account for the observations. I have chosen nine of these as relevant to the topic at hand, displayed in figure 5, below. I shall briefly explain what these principles mean, in case they are unclear, and then shall

foreshadow how they may apply to my analysis and/or the organization of macroturns, discussed in section 6.

Figure 5: Relevant principles of turn-taking (Sacks et al. 1974: 700-701)

(1) Speaker-change recurs, or at least occurs

(2) Overwhelmingly, one party talks at a time

(3) Occurrences of more than one speaker at a time are common, but brief

(4) Transitions (from one turn to a next) with no gap and no overlap are common. Together with transitions characterized by a slight gap or slight overlap, they make up the vast majority of transitions.

(5) Turn size is not fixed

(6) Relative distribution of turns is not specified in advance

(7) Talk can be continuous or discontinuous

(8) Turn-allocation techniques are obviously used. A current speaker may select a next speaker (as when he addresses a question to the other party); or parties may self-select when starting to talk.

(9) Repair mechanisms exits for dealing with turn-taking errors and violations; e.g. if two parties find themselves taking at the same time, on of them with stop prematurely, thus repairing the trouble

Consider a speaker's TURN to be when he or she is talking, and the end of that turn as when he or she stops talking for more than a small pause. The principles in figure 5 describe features of these turns, as observed by Sacks et al. Principles 1-2 describe a situation in which one speaker is talking and the other(s) is/are silent, and then the distinction of who is talking switches, so that another speaker begins talking and the other(s) is/are silent. Principles 3-4 describe the transition between one speaker's period of talking and the next speaker's period of talking: many times speakers will smoothly affect this transition so that no two speakers are talking at the same time, but sometimes, briefly, two speakers will talk at the same time. Also, sometimes there will be a short period in which no one is talking. Principle 5 means that the length of time that one speaker will talk before another speaker begins to do so may vary.[4]

[4] Principle 5 refers to utterance length. As I said, studies of turn-taking are only concerned with utterance length insofar as it relates to the transitions between turns, and so the vague assertion

Principle 6 says that the number of turns (periods of talking) that one speaker takes does not need to equal the number of turns that the other speaker takes; this is determined during the course of conversation. Principle 7 addresses the fact that another speaker may begin speaking just before, just as, or just after the current speaker stops talking (CONTINUOUS), or there may be an extended gap, or LAPSE, between when the current speaker stops talking and the next speaker begins (DISCONTINUOUS). Principle 8 means that, for example, if the current speaker directs a question to another speaker, he is selecting that other speaker as one who should take the turn after his (CURRENT-SPEAKER-SELECTS-NEXT); however, if the current speaker does not select the next speaker, any of the other speakers present who decide to jump in may do so (SELF-SELECTION). Finally, principle 9 states that, should anything go awry with the turn-taking process, speakers will tend to fix the problem by altering their behavior.

I based one aspect of my analysis on principle 8 that does not bear on the discussion of macroturns that immediately follows. When attempting to parse the conversational data I collected on a turn-by-turn basis, I incorporated principle 8 by treating direct questions differently than all other utterances. Because these questions involve the current speaker nominating the next speaker (current-speaker-selects-next), I treated the end of a direct question as the end of the current speaker's turn and the beginning of the next speaker's turn. This is reflected in my analysis of questions, below, in section 4.231.

In section 6, I shall describe a MACROTURN-TAKING SYSTEM for two-person dialogue that shares many of the features above. The domain on which this system operates will not be whether a speaker is talking or not talking (UTTERANCE-LEVEL), but whether a speaker's utterances in a given period are substantially longer than the other speaker's utterances. Specifically, I will develop the notion of a BEAT, or pattern of utterance lengths over the course

---

that utterance length may vary is all that is necessary. I adopt the same strategy when describing macroturns in section 6, allowing them to be expandable without examining the extent to which they are expandable in detail. A later study on beat-length could address this issue.

of a few turns, and propose a macroturn-taking system to describe how speakers' beats are organized (BEAT-LEVEL).

The results in section 6 show that, when it is orderly, macroturn-taking follows the principles listed in figure 5. However, the data also show that macroturn-taking is much less pristine than utterance-level turn-taking with respect to these principles. Unlike turn-taking, which rarely deviates from the principles for more than a small amount of time or spoken material, macroturns violate these principles not infrequently: one speaker's macroturn may last for a very long time, precluding the other speaker's macroturn (bending/violation of principle 1); macroturns may overlap for more-than-brief durations (violation of principles 2-4); and speakers do not always seem to change their behavior to initiate repairs for macroturn errors and violations (violation of principle 9).

Despite these differences, macroturn-taking does seem to follow principles 5-8 relatively uniformly, and sometimes follows principles 1-4 and 9. Further exposition of the similarities between the utterance-level turn-taking system and the beat-level macroturn-taking system appears in section 6, below.

2. Designing an Experiment

I knew from my informal observations that it would be impossible to simply listen to conversations as they occurred in real time and transcribe them for analysis. Though this method might work for G.B. Shaw's Henry Higgins, the flow of speech was simply too fast for me to capture it accurately without some technological aid.

Because of this, I found myself faced with the observer's paradox: in order to study "natural" and spontaneous conversation, I would have to record it. But, most feasible means to record speech would necessitate that it be in a studio or other controlled setting, which would be neither natural nor truly spontaneous: I would have to artificially bring together speakers and

make them talk. How should I reconcile these practical considerations with my intended subject?

2.1 Eliciting spontaneous speech

As one would expect, I was not the first linguist eager to observe spontaneous speech, nor the first to face the problems associated with doing so. There has been considerable research in the area. Linguists use the term VERNACULAR to describe the "natural," spontaneous, uncensored speech that we use in our everyday lives. In recognition of the observer's paradox mentioned above, it is generally accepted by linguists that one can never observe completely vernacular speech, just as biologists recognize that one can never observe nature completely independent of human influence. However, like biologists, linguists attempt to minimize the skewing effects of their observations and experiments. Experimental design, then, is based on the notion of capturing speech that APPROACHES THE VERNACULAR.

The most respected authority on the subject is William Labov. Sometimes known as the father of QUANTITATIVE SOCIOLINGUISTICS, the statistical study of how language interacts with social factors, his experiments, methodology, and findings have revolutionized the way that linguists and non-linguistics alike look at language in our society (Seabrook 2005, Linguistics 2005). Perhaps his most famous experiment disproved that certain low-prestige dialects found in Manhattan were regional, by showing them to be class-based (discussed later in this section).

Labov has laid out what he considers to be the best practices for obtaining vernacular speech along with all of the social data (race, birthplace, class, etc.) pertinent to the holistic analysis of an individual's speech. He argues that linguists should form long term relationships with linguistic communities, building up trust to ensure that all subjects are comfortable and unguarded during sessions, and therefore reveal vernacular speech to the extent that it is possible.

The preferred method: the SOCIOLINGUISTIC INTERVIEW. This is an in-depth interaction, often one-on-one, in which the interviewer attempts to elicit between one and two hours of

material from a/each speaker. The interviewer must guide the conversation to ensure that the proper background (demographic) information is obtained, as well as linguistic information of interest to the study in progress. This may be obtained overtly, with experiments or elicitations of phrases or judgments of phrases, or it may be picked up incidentally as the interview progresses. In either case, particular care is taken to steer the exchange toward topics of interest to the interviewee, in order to encourage the speaker to divulge a personal narrative. Labov contends that a speaker's personal narratives, concerned most expressly with the personal style of social interaction and community, will approximate vernacular speech (Labov 1984).

Immediate problems forced a compromise between my project and these established elicitation practices. First, in a single semester, a long-term regimen of interviews would be impossible. Second, the interview methodology Labov describes requires a trained linguistic interviewer: one needs experience in constructing MODULES, or question-maps for conversation topics, that are sufficiently informal and open, and one needs experience improvising within the a set of modules to find the topics of interest to the interviewee and draw them out. I have no such training, and would not have had time to construct and carry out a pre-experiment in which I tested various interview strategies in addition to conducting interviews, analyzing the results, and writing this paper. Third, it would have been extraordinarily difficult to convince the any of the subjects to whom I would have access (ultra-busy Swarthmore students) to participate in a two-hour long interview. The incentive required would have been too expensive to be feasible.

Besides, most of the sociolinguist reasoning behind Labov's suggested means of elicitation is beyond the scope of this paper. Demographic information is not specified in Graham's playwriting exercise; to test it, or attempt to construct a similar template, would not require knowing the sociolinguistic details of the conversation (although they might be relevant).

Beyond that, the aim of the study was to observe two-speaker dialogue, which I believe is particularly unsuited to interview elicitation if the interviewer does not wish to count him- or

herself as one of the speakers. Labov includes group sessions under the umbrella of sociolinguistic interviews, but his interview methodology seems to be aimed at producing "monologic" rather than "dialogic" speech. My personal bias says that the personal narrative may be the most natural form of speech for a single-speaker, but it is only natural in two-person speech if it arises spontaneously from otherwise back-and-forth dialogue. If the interviewer addresses a question to two speakers, such as, "Have you have ever gotten into a fight?" they may each have a story to tell that arises artificially from the question, rather than through the self-organization of their conversation. This would pose no problem if we were examining their pronunciation of certain words or the inflection of their sentences, but if the point of the experiment is to study the turn-structure of dialogue as it arises naturally, a formal "you tell yours, then I'll tell mine" setup undermines the object studied.

In a large group session it might be possible that the speakers would become involved in a group conversation after one or two initial responses, and therefore forget that the interviewer was present and slip into a more vernacular mode of speech. Labov briefly describes a few such sessions, observing that "...the best records of vernacular speech have been obtained in group sessions, where the effects of observation are minimized through the controlling interaction of peers." (Labov 1984: 48). Within the relatively transparent format of a session with two speakers and the interviewer, however, it seems unlikely that the speakers would muddle up the formal turn order, forget the presence of the interviewer, or slip into anything approaching spontaneously-arising two-person dialogue for more than a few turns. Again, an experienced interviewer might be able to goad two speakers into speaking only to each other, but the task seemed too daunting for me.

Labov suggests two alternate types of linguistic investigation that he feels have merit. One is a telephone survey, the other, a "Rapid and Anonymous" survey. Each of these methods lends itself to testing for specific features of speech. For example, the telephone survey Labov

describes consisted of 15-minute interviews (a manageable size for me) that revealed enough of the speaker's spontaneous speech for his or her vowels to be analyzed, and well as targeted questions about Philadelphia dialects. The famous experiment alluded to above was Rapid and Anonymous survey. Labov and his fellow experimenters fired questions at passers by in Saks Fifth Avenue, asking for an object that was located on the fourth floor. They found that customers on the upper, more expensive floor were less likely than those on the lower floor to engage in R-DROPPING (pronouncing words without the *r* sound where is it should appear in Standard English, e.g. *waituh* instead of *waiter*). Sales reps were more likely to r-drop on the bottom floor as well, and cashiers and other low-totem-pole employees more likely still. Most shocking of all, it didn't seem to matter where the subjects came from or lived; only their social status affected their speech (Labov 1973).

Each of these alternative methodologies is farther from, not closer to, satisfactory for the aims of my project. Rapid and Anonymous surveys do not produce extended dialogues at all; by definition they should rapidly test for a single feature. Telephone surveys would not fix the problem of needing two speakers to talk to each other naturally. In fact, all the same problems apply, except that it would be more difficult to gauge the extent to which two speakers on a phone were comfortable and engaged in conversation than two real-life speakers in front of you in the same room.

What about eavesdropping or candid recording, the easiest and most direct ways to access spontaneous speech? Of course, Labov devotes some time to these endeavors, but in his discussion of ethics, not of methodology. There are many respectable linguists who believe that, if a person is conversing in a public space loudly enough to be overheard, one may ethically analyze their speech for the purposes of analysis. However, overhearing something, even writing it down afterwards, is far less invasive that actually recording a speaker without his or her knowledge. Labov strongly disapproves of such practices on two grounds: he believes it is

unethical or at least is perceived as such, and he believes it is ineffective. "It is our [Labov's organization's] opinion that researchers who engage in candid recording will eventually cause repressive legislation," he states flatly (Labov 1948: 51). He also points out that candid recordings are typically of poor quality and not under controlled conditions, and therefore can contribute little to linguistic research. On top of all that, underhanded recording techniques may undermine the sense of trust within the community that Labov's sociolinguistic interview methodology is designed to foster.

2.2 An improvised solution

It seemed that Labov's best practices were not well-suited to my experiment. The closest approach to what I needed was the group-session interview, but I needed to find some means to minimize the interviewer's effect and encourage the subjects to speak only amongst themselves, not to the interviewer. It seemed artificial and unreasonable to assume that I could simply instruct the subjects to pretend I wasn't there. Surely that would not produce the same kind of speech that the subjects might use if they were merely speaking to one another under normal circumstances.

The compromise I came up with would be the following: a two-person "interview" session in which the interviewer was, for some reason, not present. This would verge on candid recording, specifically discredited by Labov, but should be acceptable because it would not succumb to the disadvantages of candid recording as outlined above. Subjects could be told that they would be recorded as part of the interview; this would take care of the ethical issues surrounding eavesdropping, etc. The recording device could be placed in the open, as well, ensuring that no one would feel hoodwinked, while raising the general quality of the recorded material. Without the interviewer present, any dialogue captured on the recording would be relatively spontaneous on the part of the speakers and subject to their own turn-taking practices, thereby being closer to the aim of the project than any formal interview might be.

2.3 Eliminating the interviewer

Of course, my solution immediately begs the question, "How does one conduct an interview session without an interviewer?" I could not schedule two subjects to go to an empty room at a certain time and then simply not show up. That would never work: how could I instruct the subjects as to how long they should stay, how could I inform them that I was recording the session, how could I ensure that they would know they had come to the correct room? In such an uncontrolled setting, I imagined that most students would quickly leave to go check their email and make sure that they had come at the right time.

2.31 A Red-herring experiment

A much more tenable solution would be to invent some red-herring experiment that would serve as an excuse for the subjects to be in the testing room. I could provide instructions, reassuring the subjects that they were in the correct place, etc., and then find some pretense for eliminating myself as a presence. Once again, this type of design tiptoes toward the distrustful practices Labov counsels against. But, I felt that, since I was already a member of the Swarthmore student community, there was little danger of me alienating myself from the student body. This would not be a long-term study, so there would be no danger of subjects in later experiments expecting some "twist." In terms of ethics, the subjects would still know that I was recording; they just might not know what would be important to me about their session. I gambled that Swarthmore students, rather than be upset or indignant that I had conducted an experiment about which I was not specifically concerned, would find it interesting and perhaps exciting that they had not known the point of the experiment. Thus Labov's criticisms do not apply.

The seemingly simple task of eliminating my presence from part or all of the session had some sticky points. There remained the looming threats that the subjects would sit in silence, attempt to find me to ask a question, or desert the experiment altogether. Furthermore, the more

awkward or involved my strategy for exiting, the less likely that I would be able to repeat the process identically at each session, create a relaxed atmosphere, and not give away the fact that my leaving the room was a deliberate attempt to encourage conversation between the two speakers. The latter would have the same effect, in my opinion, of me sitting the two speakers down, pointing them at one another, and commanding them, "Converse!" Surely this would not simulate normal, spontaneously-arising conversation.

2.32 The INTERRUPTION technique

I considered three techniques for eliminating my presence as interviewer in the sessions. I call the first the INTERRUPTION technique. The scenario would unfold as such: two speakers and I would sit down to conduct my experimental sessions; I would explain the basics of the (red-herring) experiment and reveal that the proceedings would be recorded. After starting the recorder, I would conduct some of the experiment. At a predetermined point, an outside influence (a friend of mine) would interrupt the session and urgently request that I leave the room for some reason, perhaps for a phone call or to help them move something extremely heavy. I would apologize and leave the session for a few minutes, without having turned off the recorder, but ask that the subjects stay in the testing room and wait for my return. In my absence, any conversation that arose between the speakers would constitute my data from the session.

Only an unreasonable optimist would accept the design above, given the concerns I raised about repeatability, simplicity, and serenity in the recording session. The subjects, if I fooled them, would believe the experiment had gone awry; they might certainly leave the room to find me after a brief interval, or give up and return to their busy work schedules. If I did not fool them, they would be extremely suspicious of the remainder of the session, and perhaps unlikely to speak normally and unreservedly with each other. The extended skit would require the help of

another person and some acting skill on both that person's and my part. The interruption technique entailed too many variables, and so I rejected it.

2.33 The JUST-STEPPING-OUT technique

The second option was what I would like to call the JUST-STEPPING-OUT technique: I would design my red herring experiment so that some aspect of it would need to be randomly selected. For example, I might read a passage from a book and ask the two subject's opinions on the material, and which passage I would read (of say, 12 possible passages) could be determined randomly. It wouldn't matter if I actually needed to randomly select a passage; the conceit would be that the software that would randomly select the passage was installed on a computer outside of the testing room. After explaining details about the experiment (including recording) and starting the recorder, I would need to leave at some point to randomly select the passage. In other words, I would leave the testing room with the excuse that I would be "just stepping out for a second." In the time that I was outside of the testing room, any conversation between the subjects would constitute my data.

This technique would eliminate some of the problems associated with the interruption technique, in that the experiment would appear to be progressing normally. It would also be far more repeatable, because the pretense for me leaving the room would be less chaotic, and it would not require a second person to help conduct the experiment. However, it has a few problems of its own. It is implausible: why wouldn't I simply conduct the session in the room in which the computer in question was located, and why would I need a computer at all? If the subjects believed that it would only take a minute or two to conduct the randomization, they might sit quietly and wait; after a certain amount of time had passed they might speak, but certainly as time went on they would begin to think either that something had gone wrong with the experiment or that there was something funny going on. This would put us right back where we were with the interruption technique, with the danger of the subjects coming to find me or

leaving. In fact, it seemed that the just-stepping-out technique would work only if the subjects attributed the long delay to my incompetence. I believed that, at Swarthmore, this would be more likely to irritate the subjects than the interruption technique. I therefore rejected it, as well.

2.34 The TOO-MUCH-TIME technique

The technique I did adopt was suggested to me by my housemate, Matt Woodbury. He proposed that, if I were to give two subjects in the same room a simple task and far too much time to do it in, they might talk during the extra time out of boredom. I improved upon this idea by couching it in the terms usually associated with experiments: I would explain that the easy task I presented to the subjects would take very little time, but as a matter of CONTROL, i.e. to make all of the sessions uniform, I would stipulate that the subjects not be allowed to leave the testing room until all of the time allotted for the experiment had passed.

This technique addressed all of the issues associated with eliminating the interviewer's presence: I could meet the subjects in the testing room at the beginning of the experiment, situate them as I wished, explain any details I wished (including that they would be recorded), start the recorder, and have a plausible reason for leaving. All of these things could be done calmly and repeatably. Since my presence would not be necessary for the subjects to carry out the task, it seemed not only plausible but reasonable that I should leave to minimize my impact. The duration of my absence need not be attributed to incompetence or the breakdown on the experiment; it was a plausible if not obvious device to ensure that the sessions were uniform. Finally, because the length of the experiment was fixed and the subjects were informed that they should stay for the entire length of it, the problem of subjects leaving the testing area was minimized. (Though not eliminated; if subjects did not read the instructions or understand my explanation, they might still come to find me to ask if they could leave.) The too-much-time technique – thank you, Matt! – seemed the way to go.

2.4 Choosing a task

In order to carry out my experiment, I had to select a suitable task to present to subjects. The task needed to have a number of features. It should encourage, but not require, speech: the target was dialogue that rose spontaneously from speakers, not through a command present in the experiment. In order that the subjects' speech approach the vernacular, the task and its trappings should seem as informal and humorous as possible; I thought this would relax the subjects and make them feel more comfortable. The task should be something familiar enough that each subject might have personal information to bring to bear upon it, hopefully stimulating conversation. However, it should also be relatively unexpected, to (hopefully) shock the subjects into temporarily forgetting the presence of the recorder. An unusual task would also prevent the subjects from guessing the point of the experiment, which I hoped would keep them engaged in it. Finally, the task needed to be something easy enough that any subject could participate in it and finish it quickly, yet open-ended to the extent that there would be no "right" answer. If the subjects were to talk about the task, as I expected they might, I did not want there to be a concrete answer on which any conversation would converge; that would run counter to observing the flow of conversation as regulated only by the speakers themselves.

The task I chose to fit these specifications is given on the next page as figure 6. Rather than approximating it, I have decided it would be more efficient to present it exactly as my experimental subjects encountered it, except that it appears on a single page. In the experiment, the paragraph of instructions appeared on one page, with the image and the question appearing on a second page, stapled to the back of the first. The image has been sized down accordingly.

Figure 6: My experimental task

The picture on the next page is a comic strip that I, Eric Eisenberg, think is funny. It comes from a webcomic by David Willis entitled "Shortpacked." Please read the webcomic and answer the question that follows it. You may talk to the other person in the room if you want, or not. The experiment will last for ten minutes, at which time I will return and collect this handout. This is (obviously) an informal setup; the only condition is that you stay in the room for the duration of the experiment. So, relax and enjoy.

Name one other activity you think Batman could make funny.

_____

The task shown in figure 6 meets the conditions I discussed above. It lends itself to a discussion of why or if the comic is funny and a discussion of what other activities would serve the same purpose, but it does not require that the speakers converse. I provided pens as part of the experiment; the subjects theoretically could have filled in their answers in writing and waited out the remainder of the ten minutes in silence.

One might argue that, if I were interested in truly spontaneously-arising conversation, I shouldn't have included the sentence "You may talk to the other person in the room, or not," as this forces subjects who would not otherwise consider discussing their answer to consider it as an option. However, I thought that this problem was less critical than not addressing what I believed would be the first thought the subjects would have upon reading the final question: "Can I discuss my answer with the other subject?" To stave off the possibility that some subjects might open the testing room door to ask me whether they could talk, I included it in the instructions.

Certainly the task is informal and unusual as the sole subject of an experiment. I crafted the paragraph of instructions to seem slightly offhand and rather amateurish, in order to cultivate a stress-free, unofficial atmosphere. I hoped that what seemed a like pointless experiment in the hands of an unimpressive Ling major would provoke people to not take themselves too seriously during the session, as this should shift their speech in the direction of the vernacular. At the very least, the subjects might have a very genuine conversation about why anyone would care enough about the webcomic to base their thesis on it. In regards to the effectiveness of the task as a means of producing vernacular speech, see the discussion of session 9 in section 6.5, below.

I believed that Batman, as a "pop culture icon", would be famous enough for any subject to participate in the experiment. But in a matter as subjective as humor, there could be no single correct response. I hoped the quirky humor of the comic would cause my subjects to forget about the recorder's presence in a laughing fit, or at least have them scratching their heads to

figure out what I could possibly be testing for. Whatever they thought, I hoped that they would be engaged.

From a couple of preliminary tests on subjects not under the experimental conditions outlined in section 3, below, I determined that the task took about 1.5-3 minutes to complete, depending on the reading speed and decisiveness of the subject. This left 7 minutes of spare time in which the two subjects might converse, guided only by their own whims.

3. Executing the experiment

3.1 Description of subjects

The subjects tested in my study were young men and women currently enrolled as undergraduates at Swarthmore College in Swarthmore, PA. The student body is made up of individuals aged 18-22 for the most part, with strong verbal skills and (usually) excellent educational backgrounds; they may be from very disparate areas of the country or the world. Although there are international students at the college who may struggle with English, by pure chance all of the subjects in my sessions were native speakers of English. I did not collect demographic information from the subjects as part of the experiment, but none of the subjects obviously fell outside the age-range of the average student: there were no subjects as old as 30 or as young as 14, for example.

Participation in the study was on a volunteer basis. I posted a signup sheet on a public bulletin board frequently passed by almost all students, promising two dollars and a candy bar as incentive to be in a 12-minute Linguistics thesis experiment. Subjects were asked pick the session of their choice from a list of open times, but informed that two subjects were necessary for each session to be carried out. The subjects were given no information about the experiment other than its duration and the fact that it was a Linguistics experiment. I chose to include the latter piece of information because it would normalize the difference between subjects who might know me (and therefore know my major) and those who did not know me. Even most friends and acquaintances of mine aware of my major did not know the purpose of my experiment, in that respect they were the

same as subjects who had never met or heard of me. At a school as small as Swarthmore (1500 undergraduates, no graduate students), opening up the experiment to my acquaintances was a necessary concession.

3.2 Method

3.21 The testing room

Before the arrival of the subjects, I organized the testing room to make it as identical in appearance as possible to every other session. The position of the chairs in which the subjects would sit and the recording device, the position and type of pens laid out for the subjects' use, the relative distance to the other chairs in the room, the absence of spare paper or other materials on the table or elsewhere in the room, the status of the computer (off), and the position of the window shades (down) – these were all pre-set conditions in the room. There were a number of posters on the wall from a class on movement and cognition, but these were immobile and remained the same throughout all sessions.

3.22 The proceedings

I recorded 12 sessions over the course of three days: Thursday, Sept. 22, Sunday, Sept. 25, and Monday, Sept 26. I asked subjects upon their arrival to leave their bags and other personal items outside the testing room. Once inside the room, I invited them to sit in two designated comfy chairs and gave them a short verbal briefing consisting of the following: they would be given an easy, short task to do, and ten minutes in which to complete it. I told the subjects that the task itself would take much less than ten minutes to do, but that I had chosen ten minutes arbitrarily as a time within which anyone could perform the task. For the sake of control, I put it, subjects were asked to remain for all ten minutes. I then informed them that the session would be recorded via an Ipod with attached Italk microphone; this device was plainly visible on the table throughout the experiment, as it was situated directly in front of the speakers.

I started the recorder, handed them the two-page handout (see figure 6, above) and delivered the line "I'll see you in ten minutes," before leaving.

The proceedings during the ten minutes of my absence were entirely up to the subjects' discretion. I returned at the ten-minute mark with a bag of candy and signature sheet. Before turning off the recorder[5] I asked the each of the subjects to state his or her name and then asked whether and to what extent the subjects knew each other before the session. After this the recorder was turned off; this terminated the experimental portion of the session. Subjects were invited to take two pieces of candy and to sign and date the signature sheet in order to receive two dollars for participation. If they wished (and every subject opted to stay) they could stay and hear my explanation of the experiment: that the task was actually extraneous to the true subject of the experiment, which was the length of utterances in spontaneous, two-person dialogue. Subjects then exited the testing room, claimed their personal effects, and left, while I pre-set the next session as described above.

3.23 The recording device

The Italk microphone that I used to record the experimental sessions plugs into any Mac Ipod. Users may start, stop and pause digital recordings, which are stored on the Ipod itself and catalogued by date and time. I transferred the digital sound files (.wav files) to my computer, a Mac Powerbook, for transcription, data extraction, and analysis.

It is traditional in most linguistic settings to use microphones attached to each speaker, feeding in to a single recording device. This is because it allows greater fidelity in the recording itself, and, if the recording device is capable of recording tracks, it allows one to isolate each speaker's voice for analysis.

---

[5] In one instance, I did not ask for the subjects' names and relationship before turning off the recorder, for no better reason than that I forgot. Instead, I asked about their relationship after I shut off the recorder. Their names I had in written form; I was recording names only as a means of matching audio files to written transcriptions.

Having no experience in recording and a limited choice of equipment to use, I opted for the Italk, an area mike that would pick up each speaker from a distance, for the following reasons. One, it was easy to manipulate, and so I could be sure that I would record sessions with little trouble or danger that I would "lose" sessions by incorrectly adjusting the microphone. Two, it allowed for digital extraction of the recorded material, which permitted me to manipulate the sound file as I needed for the purposes of transcription and data extraction (see below in sections 3.3 and 5, respectively). Three, I determined through sound testing that the Italk would pick up voice with sufficient fidelity for the experiment. The quality of the recording, though compromised by background noise or static to some extent, was high enough that I would be able to transcribe the speakers' words with surety. I therefore decided that the Italk, which was available to me, was the best option.

3.3 Transcription

At the end of each day of recording, I transferred the digital sound files (.wav format) from the Ipod to my Powerbook. Once I had collected all of the sessions' raw recording data, I listened to each session using one of two software music players – Itunes or a program called Sound Studio[6]. It seemed that the most objective way for me to analyze the recorded material would be to transcribe the sessions and create a permanent visual record of my analysis, rather than merely writing down the number of words/morphemes/etc. that I believed each speaker had spoken per turn. As it turned out, due to the difficult nature of qualifying what should I should be counting (see section 4, below) and sticking precisely to conventions for defining turns (see section 4.23, below), transcription was a crucial step in the process. It would have been impossible to produce a responsible and meaningful set of results if I had not transcribed the audio data before analyzing it.

---

[6] From the program literature: "Sound Studio is an application which records, edits, and applies effects and filters to audio." Like Praat or Audacity, Sound Studio allows one to look at a graphical representation of the waveform of digital sounds. I used demo version 2.2.4 for my purposes, which is available from Felt Tip software at www.felttip.com.

3.31 Conventions

The conventions I have used for transcription are simple and intended to be as clear as possible, while remaining within the formatting capabilities for writing in a normal word processor. The two subjects of any given session were arbitrarily given the labels A and B, respectively. I always transcribed my own words with the label E. A speaker's utterance, in its basic form, would begin with the speaker's label followed by a colon and single space, after which would appear the words that the speaker said. And example line of text from one of my sessions is given below as example 1. The citation practice I shall follow is parenthetical in the form (session number, approximate starting time of example from the beginning of the sound file, turn number(s) if necessary for reference). I have formatted the citation to be bold-faced to help separate the examples from the text.

Ex. 1: Sample transcription of a single utterance

B: like Batman riding a bicycle that would be funny **(11, 3:33)**

A sufficiently long turn could stretch onto two or more lines; subsequent lines within the same turn are indented one-half inch.

Short pauses between words (typically < .7s) are denoted by one or more spaces between words, depending on how halting I perceived the gap to be. Longer pauses are noted in [square brackets], as are many other kinds of editorial explanations of what was captured in the recording, such as *[giggles]*, *[inhales]*, and other non-speech sounds. I used the marker XXX to denote a word that I could not recover from the recording, for whatever reason (background noise, garbled speech, etc.) It was almost always possible to determine the number of words spoken in these situations, even if I could not be sure what the words were. When not possible, I made my best guess and inserted the appropriate number of XXX markers. Example 2 illustrates these conventions.

Ex. 2: A more complicated sample transcription of a single utterance

B: XXX s'like [2 tokens] the only way [pause ~1s] they've ever figured out to um [pause ~1.5s] um like notate dance y'know write it down **(2, 4:44)**

Note the editorial insertion *[2 tokens]* after the odd contraction *s'like*. The latter stood for *It's like* in the conversation, which I counted as two separate entities in calculating the length of the utterance. I have discussed my treatment of such contractions in section 4.15, below.

Turns by separate speakers are usually separated by a double return (i.e. skipping a line). However, speech is separated by concurrent sounds made by another speaker by a single return (i.e. on a contiguous line). Ellipses are used to denote lines that are broken up by these conventions that were unbroken in conversation, as shown in example 3:

Ex. 3: Sample of transcription of 2 turns, 1 by each speaker

A: I think that Batman could make something funny if something that Batman wouldn't usually do

B: I'll tell ya what he couldn't make funny is doing like five hours of homework…

A: [laughs]
B: …at …

B: …two o'clock in the morning **(6, 1:14.5)**

Note that speaker B's turn extends through A's interruption of laughter, which was concurrent with Speaker B's use of the word *at* and a slight pause, notated as two spaces before the ellipses.

I avoided punctuation, for the most part, in my transcription. However, I did make use of three different characters occasionally: the dash, the apostrophe, and the question mark. Unless the dash occurs within words normally written with a dash (like *uh-uh* or *uh-huh*), a dash indicates stuttering or interrupting—anything that would cause the speaker to abruptly stop in the middle of a word. An apostrophe appears in contractions, both of the common kind and more unusual ones arising from quick speech. Typically contractions present a problem or at least an issue for my analysis, thus it was important to mark them in the transcription (see sections 4.13-

4.15, below). Occasionally speakers would exhibit the rising inflection commonly associated with questions in English; I notated this behavior with a question mark. See section 4.231, below, for the significance of questions in my analysis. Below are examples of each of these punctuation marks as they appear in my transcriptions.

Ex. 4.1: Punctuation: the dash, for interruptions

B: it's really annoying but I feel like that kind of sou– [= 1 token] like not like vocal but just like **(8, 6:52)**

Ex. 4.2: Punctuation: the apostrophe, for contractions

B: I'm sure
A: You know [2 separate words, = 2 tokens]…

A: … if you do like sleep deprivation studies at Harvard they'll [= 1 token] pay you like a thousand dollars f'r a week **(3, 4:05)**

Ex. 4.3: Punctuation: the question mark, for canonical (rising inflection) questions

B: was this for that class?

A: yeah yeah the Cognition and Movement **(9, 3:57.5)**

Sometimes words that ran together were entered without spaces, e.g. *whaddyou* for *what do you*, or *gotta*, *sorta*, and *kinda* for *got to*, *sort of*, and *kind of*, respectively. This does not necessarily indicate that I counted the entire run-together phrase as only one entity. Typically, an editorial comment will indicate the number of tokens (counts) I assigned to it. My treatment of these words is discussed below in section 4.15.

3.32 Non-usage of accepted conventions

Certain studies involving turn-taking (Sacks et al. 1974, in particular) have had more precise conventions for transcription, especially for noting word-length, pronunciation, or overlapping speech. However, the first two items are not specifically relevant to my study (though word-length may be of interest to it; see 4.11 and 4.21, below), and the convention often used for the last item was not feasible. Sacks et al. (1974) use a system of brackets that mark off when one speaker began

to talk at the same time as another; these brackets extend across more than one line of text and are simply not possible to enter in a normal word processor. I feel that my conventions are adequate for the needs of my project.

3.33 A time-crunch problem, and a narrowing of focus

The glaring problem with my transcription strategy is not specific to my system, but rather must be intrinsic to transcription in general: the process was stupendously time consuming. In order to complete the transcription phase of the project within a reasonable timeframe, I was forced to narrow the focus: I chose the 6 clearest (i.e. most easily audible and understandable) sessions, and concentrated my efforts on transcribing them only. These sessions, in part by design and in part by luck, were distributed by sex to give me a full spectrum of combinations: male-male conversation, female-female conversation, and female-male conversation.

The conversations that I transcribed were chosen on no other criteria than those mentioned above (clarity and sex-distribution), except in the case of session 9. The subjects in session 9 had the distinction of being the only subjects I recorded who knew the purpose of the experiment. One of them was a senior linguistics major (and thus had heard me discuss my research as part of our thesis seminar), and the other was a theater student with whom I had shared Graham's playwriting exercise during my informal exploratory phase discussed in the introduction. I chose to transcribe session 9 rather than throw it out because I wanted to see to what extent knowledge of the purpose of the experiment affected the result. This would be a window into the extent to which my unusually opaque experiment might encourage the vernacular in the subjects. The discussion of session 9 appears below in section 6.5.

4. Defining the terms

Having transcribed the 6 sessions I intended to analyze, I needed only to extract the length of each utterance in each conversation to finally have results I could compare with Graham's template. However, the manner in which I should do this turned out to be no small consideration. What unit of

measure should one use to quantify length? Over what domain should one calculate these units, or, in other words, how should one define *utterance*?

4.1 A unit of measure for length

4.11 Time

The quick and easy answer to the question of how one should measure length is to use seconds of speaking. There is no more objective measure: no theoretical debates over what constitutes a unit of time, no lengthy considerations of what should count as the beginning and end of each chunk to be measured. Start the stopwatch when the speaker starts talking, and stop it when he or she finishes, plain and simple.

Sadly, the situation is not so clear-cut. First of all, time isn't perfect for the linguist as a unit of measure, and it certainly isn't perfect for the playwright. The playwright has little control over the time in which his lines will be delivered. It is already common practice to change or ignore stage directions in rehearsal; imagine how laughable it would seem to many actors and directors to encounter the stage direction *Spoken in 10.6 seconds, followed by a 2.3 second pause.*[7]

Linguistically, time as a unit of measure ignores many of the salient features of speech (particularly that it is composed of discrete words, not uniformly distributed content) and so would be unlikely to shed light on any issues of interest without further investigation. For example, in a strictly time-based study, differing rates of speech among the speakers would obscure the data: imagine that a talking robot is conversing with a regular woman, and that their

[7] There are some director/designers who have undertaken such exactly-timed projects, "betwixt and between theatre and the visual arts," with success, notably Robert Wilson (Abrams, 2003). However, this must be considered the exception rather than the rule. Most playwrights working within a conventional dramatic framework do not have enough artistic control to choreograph scenes so minutely; nor do most directors wish to waive their right to pace the scene according to their artistic vision.

dialogue is as I have given in figure 7 below. Imagine further that the robot leaves 1 second pauses between all of his words as part of his language programming.

Figure 7: Conversation between the slow-talking robot and Nancy, a regular person

Nancy: Hi Robot.

Robot: Hello Nancy.

Nancy: How are you this afternoon?

Robot: I am functioning well today.

Many linguists (and playwrights, too) might reasonably say that the Robot and Nancy have spoken about the same amount in the exchange given above. They have exchanged roughly equivalent greetings. Nancy queried Robot about one item, his health, and Robot responded about the same item only. Each of them included a single piece of information that was not strictly necessary to the query, namely the time of reference (Nancy's *this afternoon*, and Robot's *today*). In terms of both the information they have presented to the other speaker and the number of words they used to accomplish it, the speakers were equal.[8]

Consider, however, what the durations of their sentences would look like when compared: the Robot would appear to be dominating the conversation in virtue of pausing so long between each word. Clearly this would be counter-intuitive: it would not capture for the playwright the fact that the Robot is not a controlling force but a sluggish one, and it would not capture for the linguist that the Robot is no more "active" in the conversation that the human, it is merely less efficient.

Would the differences between speakers be so pronounced as to render the data useless? Probably not. As the clear similarities between the types of graphs I will present in sections 5.31 and 5.33, below, will make apparent, the length of time that a subject speaks is at least correlated

[8] It is true that the robot uses one more morpheme than Nancy. I will make the case in section 4.14 that morphemes are not a suitable unit of measure for the study of utterance length because they exaggerate differences among similar utterances.

with the amount they have to say, in real life. In fact, there are some advantages to using time as the unit of measure rather than, say words or sentences. Speakers use the time in which they speak their words to change tone and sometimes meaning: pauses for dramatic effect, slow and deliberate speech to make a point, drawn out words for emphasis – all these involve changing the rhythm or tempo of speech to affect a change. Some speakers have very different rhythms than others: they might have an unhurried drawl or a rapid-fire chatter, they may use a lot of time and/or words when few would do, or they can refuse to take up more than the bare minimum of conversational time for reasons of spite, fear, etc. If the point of Graham's exercise is to observe speech rhythms, as he said, it would be unfortunate to ignore time as a rhythmic factor.

So, if time is a factor important enough to include in a study but not informative enough to be the only unit of measure examined, it seems that the solution can only be to use time as a unit of length along with some other unit of measure that is more dependent on content. And, since time is more objective than any other unit of length, it can serve as a point of comparison for the analyses dependent on the other, contentful measure.

4.12 The contentful measure: sentences?

When Roger Brown discussed children's utterances (section 1.21, above), he was thinking of the sentences or proto-sentences that the children in his study were producing. For example, a child with an MLU of 2.0 on average produced two word "sentences" that probably involved the apposition of two nouns such as *me ball*, which might mean *give me the ball.* By the definition of MLU, it should be clear that it would be uncharacteristically sophisticated for this child to construct two of these two-word phrases in the same breath, since that would require 4 words (again, she averages 2 per utterance). Much more commonly the child might produce a three-word sentences such as *I want ball* for *I want the ball.* In that sense, this child is usually only capable of producing one two- or three-word "sentence" per utterance, and so it would have

been foolish for Brown to use sentences or proto-sentences as his unit of length when calculating MLU: every child's MLU would be equal to 1 for a very long time.

Adults, on the other hand, are fluent, and can string a number of sentences together very easily in the course of conversation. In theory, it should be possible to use the sentence as a unit of length to measure adult speakers' utterance lengths.

In practice, it is so difficult to define *sentence* that sentences as units for length would be impossible. Consider the following example:

Ex. 5: Run-on sentence

> B: Idunno [= 1 token] I think of like *The Incredibles* and like how alot of it's funny like he drives in the little tiny car and he like y'know it's just funny that he has to like sell insurance and stuff like that **(11, 2:58.5)**

We're told not to use run-on sentences when writing, but no one seems to mind in casual conversation. Should each of the clauses beginning with *and* be considered a separate sentence, or not? If the example counts as a single sentence, we have a bit of a problem with our unit of length: the sentence *He runs* would count as one unit, and the sentence given above (which is by no means the longest run-on sentence imaginable) would also count as one unit. That covers a lot of ground, and would make the utterance length statistic less meaningful because of its lack of precision. However, on what grounds should we chop up what is essentially one long conjunctive sentence?

Words not belonging to any clear sentence structure, such as *yeah*, *m-hmm*, *Wow!*, etc., also pose a problem. Conversation is littered with them; to leave them out would be to ignore what is actually present in the conversation, but to include them would ignore any definition of *sentence* as one usually conceives it, since these words often occur isolated from both a "subject" and a "predicate". Sentences pose too many problems as a unit of length, chiefly because of the difficulty in defining what a sentence is.

As the final piece of evidence against the use of sentences as a unit of length, consider example 6, below. Given the interruptious (a neologism) nature of the speech – sometimes called SELF-CORRECTING – it is difficult to consider how one could even decide where the sentences start and stop. The first sentence train-wrecks into the phrase *I I hope...*, and that phrase (sentence?) transforms into the question beginning with *how are you*, which seems to be a well-formed sentence in itself. The example does not appear too abnormal when spoken aloud, but, when written, demonstrates the difficulty of a sentence-by-sentence analysis.

Ex. 6: Speech with interruptions that defy sentence-by-sentence analysis

> B: yeah it's probably just to collect information on y'know what happens after we've started talking about I I hope that's what it is 'cause y'know how are you supposed to talk about Batman for ten minutes? **(6, 4:50)**

4.13 The contentful measure: words?

The sentence seemed to be too big and too imprecise a unit of length for the purposes of this project. Words are considerably smaller than sentences as a unit, and are not nearly as expandable as sentences, so they seem more precise. But, words as a unit of length also have their shortcomings.

In writing, words are separated from their surroundings by white space, making them discrete and easy to count. In speech, pauses – what one might consider the spoken equivalent of white space – do not always occur between words, making it considerably more difficult to count them. Furthermore, speakers' sometimes contract, delete, alter, compound, or interrupt their own or another's words. These adjustments poke holes in a strictly word-by-by analysis.

First of all, speakers insert EXPLETIVES, or filler words, into their speech to buy time during conversation. It is unclear whether *um*s and *uh*s should be considered words. Aside from that, there are certain common contractions that are conventional enough to be written as one word in even semi-formal writing, among which are *it's, cannot, can't, didn't, isn't, I'd, I'm, we're* etc. Perhaps one wouldn't use these in the driest kind of academic paper, but they appear in

editorials, articles, biographies, textbooks, and other publications that contain what is purported to be Standard English. These contractions display the fact that they are derived from two words originally, but are smushed into a single graphical entity in writing. Should this be taken as evidence that they count as only one word, in this day and age?

The problem seems small when discussed in terms of a handful of written contractions, but broadens considerably when one attempts to apply the same question to the numerous unconventional contractions arising out of quick everyday speech. I have already mentioned or included examples containing *gonna, kinda, sorta, s'like, y'know,* and *Idunno*. The first three are considered too informal for writing, but are still recognizable as common compounds. *S'like* is much more rare, and some people (including me, as I will discuss) balk at the idea that it should be considered a single word. The final two are ubiquitous as EXPLETIVES, or filler words, and are usually spoken so fast that it is hard to know whether they are contractions or whether it's just too hard to hear the full words. There are some more extreme examples, too. Consider example 7:

Ex. 7: Unconventional contractions in speech – a continuum

A: Well righnow [*right now*] we have like everyday everyday like things in our lives just like riding a bike…

A: …singing playing an instrument
B: Yeah like I don't think id [*it would*] be really funny…

B: …if like Batman went into space like

This example is useful because it illustrates both extremes of the CONTINUUM, or unbroken range of variation, along which contracted words may fall. Speaker A used the contraction [ɹajnaw] for the two words *right now*. The two words are both clearly visible in the contracted form, and in fact there is only one sound missing: [t]. Since Speaker A's word [ɹajnaw] is by no means a commonly accept contraction, and since the two words are contracted

so little, it seems reasonable that a word-by-word analysis should count this as two tokens, just as though the speaker said the two words *right now*.

When Speaker B responds, she utters the particle [ɪd], which I interpreted as a contracted form of *it would* based on the context. Here is a possible derivation: *it would* contracts to *it'd*, and the now-adjacent sounds [t] and [d] coalesce into a single sound through two linguistic processes, VOICING ASSIMILATION and DEGEMINATION.[9] This contraction is at the opposite end of the spectrum from *righnow*: both of the source words have largely been deleted in [ɪd] through an involved process that obscures which portion of each word is present in the final utterance. Is it the [ɪ] from *it* and the [d] from *would*, or is it essentially just the word [ɪt] with a sound change in the final consonant because of the proximity of the [d] in *it'd*?

A word-by-word analysis probably shouldn't care about these issues: the two source words for the contraction are clear, so the contraction counts as two words. Perhaps we might define a short list of common contractions to count as one word; Bruce Graham said that contractions counted as one word in his exercise, after all. Yet even though [ɪd] is not a conventionally accepted contraction, I am inclined to consider this extreme form of contraction as 1 word, sort of divorced from its original two source words, simply because it is so compact and so distant from the phrase *it would*: [ɪd] doesn't sound like [ɪt wʊ̄d] in conversation.

Linguists might object to the sentiment, but playwrights must concentrate on the minutiae of diction for the very practical reason that what "sounds right" is what they should write down. If the analysis is to be useful to the playwright, treating *id* and *it would* identically misses the point that they have a different effect when spoken aloud, and the sound of the lines as they are spoken aloud is the measure of all playwriting meant for performance.

---

[9] For the non-linguist: voicing assimilation would turn the [t] of *it'd* into a [d], and degemination would turn the resulting word [ɪdd] into [ɪd] by deleting one of two adjacent, equivalent sounds. This sort of derivation is entirely plausible in English (Crist p.c 2005).

Through similar reasoning we can come to grips with deletion, another problem for a word-by-word analysis:

Ex. 8 Deleted, but recoverable material

B: Used to watch Batman when I was little that was a good show

A: Oh really I never watched cartoons My parents never let me watch cartoons [pause ~1.5s] [very soft] but **(6, 5:04.5)**

Speaker B's utterance in example 8 illustrates a phenomenon that Donna Jo Napoli (1982) has described as INITIAL MATERIAL DELETION. Unlike Spanish or other Romance languages, English speaker are not "allowed" to drop the subjects from the beginnings of their sentences. However, they sometimes do, and they do so in an orderly manner. Napoli explains that speakers may optionally delete "unstressed (or lightly stressed) initial material" in informal speech, usually in brief utterances (99). However, the material that is deleted is recoverable from context: speaker A did not respond, "Who used to watch Batman?" because it was obvious that speaker B was referring to himself.

Now, speaker B's sentence could very well have begun *I used...*, since it was intended to be understood that way, and in fact it was understood that way. In a word-by-word analysis, should the implicit *I* at the beginning of his sentence be counted, or not? Well, the word wasn't spoken, so by the playwright's logic (sound when spoken = most important), it shouldn't count.

The question has more bite when part of the word remains present. The contraction *s'like*, mentioned above, arose from initial material deletion of the beginning of the phrase *it's like*. Sounds are dropped from words on all sides in rapid speech: I found many examples each of *'n* for *and*, *'cause* for *because*, and *–in'* for *–ing* as a suffix.[10] How much of the dictionary form of word need be present for that word to count? In a strict word-by-word analysis, counting things that are smaller than words seems like cheating. But anything that is spoken would be

[10] The last is not really a deletion so much as a sound change; it looks like a deletion because [n] corresponds to one letter in English orthography and [ŋ] to two.

important for the playwright to include in a line of dialogue; otherwise, how would the actor know to pronounce it? Linguists too would probably be disappointed not to consider sounds that clearly stand for words like *'n* as countable tokens, just because they aren't "real words." Is there a smaller unit of sound and meaning that could be used to as a unit of measure?

4.14 The contentful measure: morphemes?

The smallest unit of meaningful sounds is called a MORPHEME in linguistic literature. A speech sound or combination of sounds with associated meaning is considered a morpheme. So, the word *cat* is a morpheme, while the first sound of it, [k^h], is not a morpheme, because the latter has no associated meaning. Morphemes need not be full words, however: the plural marker *–s* is only one sound, but it has associated meaning (plurality), and so it is considered a morpheme. Similarly the prefix *re–* is a morpheme in virtue of having meaning, and so is the progressive suffix *–ing* on verbs.

Because morphemes are dependent on meaning, an analysis in which morphemes were the unit of measure of length would be able to accommodate partially deleted words like those mentioned in the previous section. Those partial words would still have sound and meaning, and so they would fit the requirements for counting as a morpheme.

Despite the fact that morphemes are an extremely useful entity in the study of linguistics, it is my opinion that morphemes are too fine-grain a unit for the study of turn-taking and utterance length as is required in my project. The reason for this is that counting morphemes will exaggerate differences between similar statements in the same way that only looking at time would exaggerate differences between similar statements. Consider the following hypothetical exchanges, givens as figures 8.1 and 8.2, below:

Figure 8.1: Hypothetical dialogue 1 showing pitfalls of a morpheme-specific analysis

A: What kinds of pets do you have?
B: Dogs and cats. [c.f. B: Dog and cat.]

Figure 8.2: Hypothetical dialogue 2 showing pitfalls of a morpheme-specific analysis

A: The propeller fell off.
B: That's no good.
A: Would you attach it?
B: Okay, I'm reattaching it.

Both of the dialogues above illustrate skewing effects morphemes might have on an analysis of utterance length or turn-taking. Remember that the point of the endeavor is to make the data reflect the back-and-forth sense of dialogue as much as possible, in terms of the amount of information exchanged and the amount of spoken material used to convey that information.

In Figure 8.1, the two possible responses for speaker B are roughly equivalent in the amount of information they supply to answer speaker A's question. They each make it clear that B has both dogs and cats as pets: in the first answer, B includes the information that he has more than one of each, and in the second answer, B includes the information that he has just one of each. In terms of the amount of material needed to convey the information, return to the notion of time as the most objective measure of that material. In rapid conversation, it takes almost no time at all to say the extra *–s* morphemes in *dogs and cats* as opposed to *dog and cat*, yet in a morpheme-driven analysis the former would be counted as 5 and the latter as 3 tokens. That's 66% more; surely the difference between these utterances is not so much as to warrant such a disparity in the analysis.

In Figure 8.2, again the speakers address the same issue with roughly equivalent material. I have underlined the sticky point for a morpheme-driven analysis: speaker A asks speaker B to *attach* something and speaker B responds that he is *reattaching* it. I set up the dialogue so that it would be obvious that these two words refer to exactly the same activity, yet *reattaching* would count as three times the amount of material as *attach* because it has three morphemes. Again, the former would not take three times as long to say as the latter.

These considerations point to the heart of the question: each of the contentful units discussed so far fails to capture what constitutes a significant contribution to the conversation. From a strict linguistic standpoint, each morpheme contributes information and is therefore significant to some degree. To a playwright, however, particularly the *–ing* in *reattaching* or the *–s* on *dogs* may not seem incredibly significant. Portions of words may be significant even if the word is not entirely present, such as *'cause* for *because*. But portions of words in conventional contractions may not seem significant compared to speaking both words. If *I can not* is to be counted as 3 tokens, *I can't* shouldn't also be counted as 3 tokens just because *n't* stands for *not* – the two sentences have an entirely different feel when spoken aloud, and the former definitely takes up more time and has a different stress pattern than the latter. It means something else when said, unlike *reattaching* v. *attaching* in figure 8.2, where the two words mean the same thing.

I have used the playwright as an example of a non-linguist who would object to counting morphemes, but the arguments hold true for linguists as well in certain contexts. I believe that the linguist interested in turn-taking as a broad-scope phenomenon should agree with the hypothetical playwright whose opinions I have discussed: morphemes are too small a unit of measurement; they do not provide the flexibility to tailor one's analysis to reflect the structure of vernacular discourse at the level of full conversations. If a speaker were to use a number of contractions over the course of a few turns in which they were telling a story, and if the story were told with a lot of gerundial (i.e. *–ing*) expressions and plurals, all of those excess morphemes would only get in the way of understanding the structure of the dialogue, for the same reasons as I have discussed above.

I therefore decided that, while words are too large a unit of measure of length, morphemes are too small and restrictive. Instead of picking the lesser evil, I decided to construct

a hybrid analysis that would minimize the discrepancy between what I believed was salient about turn-taking and what was as faithful as possible to the material spoken by the speakers.

4.15 My hybrid analysis

I constructed my analysis with an eye to what would be useful for the playwright. I believe that, for the purposes of broad-scope discourse analysis, the conventions I lay out for the playwright's sake have linguistic merit, in virtue of avoiding some of the pitfalls I mentioned above.

In my analysis, one "count" shall be referred to hereafter as a TOKEN. Essentially, I felt that using words as the unit of length was the proper level of specificity for the playwright's purposes, since the playwright works in a literary medium. However, because of the problems I listed for a word-based analysis in section 4.13, above, I opted to use my linguistic training to come up with a compromise analysis that would accommodate sounds not properly called "words." These include deleted but recoverable words, partially deleted words, expletives, contractions, and other aspects of informal speech.

4.151 Inflection and derivational morphemes

Bearing in mind what should constitute a "significant" addition to an utterance, I rejected the idea that I tokens be given for INFLECTION AND DERIVATIONAL MORPHEMES — those prefixes and suffixes that deal with tense (e.g. *–ed*), aspect (*–ing*), agreement (*–s*), turn one part of speech into another (*–ize*), or predictably alter the meaning of a root word (*mis–*, *re–*, *con–*, etc.). Thus *activate, reactivate,* and *deactivated* would all count as 1 token each; certainly this looks more like a word-based means of counting than a morpheme-based one, which would give those words 3, 4, and 5 tokens, respectively.

It is not the case that all BOUND MORPHEMES, or morphemes that may not appear in isolation, are stripped of their tokens. The bound morpheme *–surrect–*[11] would receive one token in the words *insurrection* and *resurrection*, because otherwise these words would receive no tokens at all. On the other hand, I would count *sailboat* as one token, not two, even though it contains two morphemes (free morphemes *sail* and *boat*) because in principle I am attempting to count words or partial-words that stand for whole ones, not morphemes.

4.152 Partial and full deletion

What does the phrase "partial-words that stand for whole ones" mean? Well, recall one of the problems I noted above for a word-based-analysis: initial material deletion. If a speaker says *Wish I had thought of that* instead of *I wish I had thought of that*, he doesn't say the first *I*. To assign a token to the invisible/inaudible *I* that is implied misses the point that the speaker simply didn't say it. But, if a speaker says *ts'interesting* instead of *it's interesting*, some of the word *it's* remains. In this case the speaker did say the word at least in part, and so I have decided that the initial sound [ts] from *ts'interesting*, like the initial [s] in *s'like*, is a partial-word that stands for a whole word. I have assigned tokens to these partial-words arising from initial material deletion, so *ts'interesting* and *s'like* would each count as two tokens.

By analogous reasoning, I have not counted any fully deleted words as tokens, even if the words are fully recoverable and not "supposed" to be left out, like the article *a* in the following sentence:

Ex. 9: Deleted material not counted as a token

A: I see him in like Sport Utility Vehicle **(11, 5:01.5).**

[11] Although this root was probably backformed from *–surrection*, nowadays, with *resurrect* as a word in the lexicon, the root must be *–surrect–*.

The speaker in example 9 neglected to say *a Sport Utility Vehicle*, and so I did not count the deleted *a* as a token, even though it is recoverable from context and was probably omitted as an (ungrammatical?) accident.

But, I have counted partially deleted words, through interruptions or fumbling speech, as tokens. Thus *that kind of sou-* in example 4.1 above equals 4 tokens, not 3, even though the word *sound* was not fully pronounced. It was a partial word that stood for an entire word. An extremely common means of producing partially deleted words was for a speaker to change his or her mind about what to say after beginning to speak, and so interrupt him- or herself (self-correct). Examples 10 illustrates this process twice:

Ex. 10: Partially-deleted word from change in conversation tack

A: …and just Bat—the cl— like the whole costume makes a big difference too **(11, 2:32)**

Speaker A was attempting to explain that Batman's clothes make him unusual and therefore funny in many mundane contexts, but was having trouble stringing the thought together. In the example, *Bat-* clearly stood for Batman, and *cl-* seemed to stand for *clothes*, but the speaker decided that *costume* was a more appropriate term, and so interrupted himself with the correction. I have treated partially deleted words from this kind of self-interruption as separate tokens no matter how clipped the interrupted word is, provided that some sound remains of the deleted word. A pause while the speaker considered what to say, for example, would not count as a token.

4.153 Stuttering – first sound repetition

One exception to the practice of assigning tokens to partially deleted words is that I did not count single-sound stutters at the beginning of words that clearly arose from the speaker being tongue-tied, as in the exchange below:

Ex. 11: Single-sound stutter at the beginning of a word (not counted as a token)

B: Wow very flashy…

A: yeah
B: … are you a L-…

B: …Linguistics major? **(9, 5:58)**

Speaker B's stutter *L- Linguistics* counts as 1 token, not two, because the partial word was merely a failed attempt to produce the first sound of the whole word that follows it. Since the speaker didn't intend to pronounce *Linguistics* twice and didn't come close to doing so, I decided not to count this type of DISFLUENCY, or speech error, as a separate token.

4.154 Stuttering – (near) full-word repetition

Stutters of more than the initial sound were counted as separate tokens, though. Speakers did not tend to produce stutters of the first consonant and vowel as much as stutters of the first consonant alone. If they managed to make it past the first consonant without stuttering, they often progressed fairly far into the word before they stopped themselves. The result was many full-word or nearly full-word repetitions. Examples of each kind of disfluency are underlined in 12.1 and 12.2, respectively. I treated both kinds as separate tokens, because the amount of time required for the speaker to stutter out nearly the whole word or the whole word was roughly equivalent to the time in which they could simply say a second word. Even if the speakers did not intend to repeat the word on which they stuttered, in effect they did, and so I treated the failed attempt as a separate token.

Ex. 12.1: Full word repetition (separate token)

A: I never got to watch cartoons never got to eat sugar cereals

B: <u>That's that's</u> [= 2 tokens] desperation right there sugary…

B: ... cereals like lucky charms
A: No you're never allowed **(6, 5:57)**

Ex. 12.2: Nearly full-word repetition (separate token)

A: and then but <u>that that's</u> [= 2 tokens] fun taking philosophy About Morality and then **(3, 9:22)**

In horror, Speaker B in the first example above (13.1) rapidly repeated *that's* twice when only one instance would have been necessary; I counted each as a token. (Incidentally, since I counted repetitions of words resulting from disfluency as separate tokens, it follows that I would treat intentional repetitions of words – separated by enough time to preclude their being stutters – as distinct tokens as well. I have done so.) In example 13.2, speaker A's *that* was rapidly followed by the word *that's* with greater stress; this indicated that the latter was a correction of the former. I treated this partial-word repetition as its own token.

4.155 Contractions - conventional

The disfluency in 12.2 serves as an excellent segue to a discussion of contractions. The picky reader may notice that I dubbed this example "nearly full-word repetition," but it does not technically involve a full word. Rather, it involves a contraction of two words, and the first half of the phrase *that that's* is just the first word of the two-word contraction. This discrepancy is reconciled by the fact that, as a matter of regard for the playwright's interests, I have decided to count all conventional contractions as single tokens. My reasoning for doing so is threefold. One, if I was reluctant to assign a separate token for the *–s* in *dogs* because it doesn't add significantly more information or time to the word, it seems silly to assign one for the *–s* in *that's* or *it's*, which takes about as much time and is largely redundant from context (speakers not infrequently left these *–s*'s out by accident). Two, these contractions are widely considered single units: Bruce Graham stipulates that contractions count as one word, and Microsoft Word's word count tool counts *I'm*, *can't*, *won't*, *we're*, *they're*, *I'd*, etc. as one word each. Three, the feel of these words when spoken aloud differs from their uncontracted forms because they are

conventionally contracted in informal speech. Again, this rationale matters only to the playwright and not the linguist, but I have adopted it anyway.

"Conventional contraction" refers not only to those contractions often seen in writing, such as the ones listed in the second reason above, but also to contractions commonly acknowledged but usually not written except in dialect, such as *kinda, sorta, woulda, shoulda, gotta, wouldja, couldja, whadja, gonna, dunno,* etc. I also lump a pair of common expletive phrases that are nearly always completely enjambed into this category: *Idunno* and *y'know*.

4.156 Contractions – unconventional

"Conventional contraction" does not include a number of other run-together phrases that appear in conversation fairly often, however, such as *whaddoes* for *what does*, *year're* for *year are*, *kid'll* for *kid will*, etc. (all found in my transcriptions). Rather than attempt to come up with some typology for what contractions are similar to the conventional contractions I count as only one token, I have merely treated all non-conventional contractions like partial deletions. That is to say, if any portion of a contracted word remained, I assigned a token to it. All of the unconventional contractions above would count as two tokens under this system. Certain contractions that involve sound change rather than simple contraction, such as *id* for *it would* or *thad* for *that would*, qualify as single-token utterances because the case can be made that none of *would* remains in these contractions, and they sound substantially different than their expanded forms.

4.157 Expletives and response words

In keeping with the maxim that what a speaker finds important enough to say -- consciously or unconsciously, accidentally or on purpose – should count as a token, I treat expletives that may or may not usually be considered "words" as tokens. These include *um*, *uh*, *like*, and *dunno, Idunno,* and *y'know* from before. I also treated what I would like to call RESPONSE WORDS as tokens, such as *yes, no, yeah, m-hmm, uh-huh, uh-oh, yup, nope* and other

similar words from conversation. Repetitions of these words have been treated as separate tokens, so that *yeah yeah* = 2 tokens, and *no no no* = 3 tokens.

There are some sounds that are not as recognizable as words as the response words listed above, but have a similar effect in conversation. I have treated most sounds making use of the vocal apparatus and intended to convey some emotional response to what has just been said as tokens. Thus, I have assigned tokens to certain types of responses such as *hmm*, *huh* (not the interrogative one), *eh*, *ugh*, and *agh*, which would almost certainly be considered just sounds and not words in most analyses. These words signal that the speaker is listening and that he has some response to what has been said, and since they use the same apparatus as words, I see no reason not to consider them as such just because they don't appear in any dictionary.

I draw the line, though, at sounds that do not make use of the vocal apparatus in the same manner as speech. I have not considered laughing, inhaling, exhaling, guttural grunting, snorting, or other noises as tokens.

4.158 Numbers and times

I imagine no one would dispute assigning one token to instances of any number from 0-20, or 30 or 40, and so on, since these numbers have unique names. However, since numbers like *twenty-one* are typically hyphenated in print and since the speakers' pronunciations of these types of numbers slurred together like compounds, I opted to treat numbers as a single token regardless of whether they needed a hyphen when written: *three* would be assigned one token, as would *thirteen* or *thirty-three*. I have further transcribed most numbers as numerals, so one would find *3, 13,* or *33* is the place of the examples above.[12]

---

[12] My treatment of numbers in less than rigorous. There may be an issue of to what extent large numbers should be treated as single tokens or divided into smaller units on the basis of mnemonic groupings. E.g., 6106901000 may be most easily remembered as a set of three numbers, 610-690-1000, just as phone numbers are grouped. Since there were no instances of any numeral greater than 99 in my sessions, I have not taken a stance on the issue.

Times are usually given as two numbers (hour and minute). I followed the convention that times of the form [hour]:[minute] should be counted as two tokens, while times of the form [hour] should be assigned one token. Thus *1:15* counts as 2 tokens and *4* (referring to 4:00) counts as 1 token. If a speaker used any of the conventional tags associated with times, such as *PM*, *AM*, or *o'clock*, I counted this as a separate token. No speakers gave a time in the form [duration] *after* [hour] (e.g. *quarter after 6*), [duration] *past* [hour] (*half past 6*), or [duration] *'till* [hour] (*ten 'till 6*), and so I have not come up with a convention for these phrases.

4.159 Titles and other multi-word conventional phrases

Despite the fact that they are, in some sense, multi-word compounds, I have not adopted the convention of treating titles – course titles, book titles, movie and play titles – as a single token. Although a speaker may have a single lexical entry for the movie *The Incredibles*, I believe that this lexical entry must be co-indexed somehow with the idea that the movie title consists of two words and not one big term. Consider the example of the course *Structure of Tuvan*, which I took last spring: most of the students in the course were Linguistics majors who knew that, to fulfill the requirements to graduate, they needed to take a course whose title was of the form *Structure of* [non-Indo-European language]. Thus every student, as far as I can remember, was aware that the course's title was *Structure of Tuvan*. However, when referring to the class, students used the shortened title *Tuvan*, not the full title listed in the course catalog (as evidenced by one of my former classmates referring to *Tuvan* in one of the recording sessions). If I were to assign one token to full title *Structure of Tuvan*, there would be no way to capture the fact that students obviously felt that this title was too long for casual conversation and then shortened it. *Tuvan* and *Structure of Tuvan* have very different feels when spoken aloud, and as I have argued before, my analysis should reflect it in order to be sufficiently responsive to the playwright's needs. I therefore have treated each word of a title as a separate token, and if the

title has been shortened, I assign a token to every remaining word or partial-word in accordance with the practices outline above in this section.

Certain other conventional phrases that seem compound-like in a way similar to titles have been treated identically. A case in point: at Swarthmore most courses are offered with either sets of two class times or sets of three class times. The courses with two class-meetings convene on Tuesdays and Thursdays; the courses with three class-meetings convene on Mondays, Wednesday, and Fridays. Thus, students attempting to differentiate one type of class from another may ask questions like the one in example 13, below:

Ex. 13 Multi-word conventional phrase

B: umm is this in your Monday/Wednesday/Friday?

A: no, on my Tuesday/Thursday **(6, 3:10)**

Perhaps if speaker B had spelled out *MWF* for *Monday/Wednesday/Friday*, I would have decided to count it as 1 token. Since he said the names of all three days in full and since speaker A said the names of both days in full, I have treated these expressions 3 and 2 tokens, respectively. I have decided to treat each individual word in similar convetional phrases as a separate token in order to avoid having to come up with a typology of what counts as a common phrase consisting of separate words versus what counts as a generally accepted multi-word compound.

4.16 Why not Roger Brown's analysis?

One of the few linguists ever to propose a codified counting system like the one I have developed above was Roger Brown, whose work is described in section 1.21 of the introduction. However, I have not followed Brown's system because, just as mine defers to my concept of what a playwright cares about, Brown's defers extremely to his notion of what appears or does not appear in the lexicon of a child language learner. Since I am concerned with adult speech, many of his conventions do not apply. For example, since Brown was concerned with cataloguing the grammatical complexity of the children he studied, he was intensly interested in

accurately capturing inflectional morphemes spoken by his subjects such as the plural *–s*, the possesive *'s*, verbal inflections like the third person singular *–s* and the past tense marker *–(e)d,* and the progressive *–ing.* As I detailed, I believe that attaching importance to these morphemes would be entirely orthogonal to the goals of my project.

Brown does include repetitions of words a separate tokens, and counts response words as tokens, but neglects to count expletives because they do not affirm any amount of grammatical complexity in children's speech. Since I am concerned with the surface from of speech and not the underlying grammatical complexity behind it, his rationale doesn't apply (Brown, 1973).

In general, it seemed foolish to use Brown's conventions for just that reason: he was measuring what children had learned, while I am performing a very "stupid" count that is primarily concerned with the nature of turn-taking and rhythm rather than language acquisition. Why use a tool designed for some other job? There seemed no good reason to follow Brown's precedent just because it was a precedent.

4.17 A note on transcription and the counting process

The analysis given in 4.15 may or may not be a complete list of every type of exception, contingency, or unusal circumstance one would encounter in attempting to perform an analysis of dialogue as I have done in this study. However, these were the considertions I felt were important enough to include and codify in this paper. In all other circumstances I have used my best judgement to count the number of words and partial-words that stand for whole ones that each speaker uttered as fairly and consistently as possible.

I hope one thing is clear from the copious detail in which I have examined how one should go about defining a unit for length of utterance and how one should go about counting those units: accurate transcription was vital. The presence or absence of *ts-* in *ts'intersting* makes the difference between one and two tokens for the utterance; this *ts-* might be only a tenth of a

second long. Only an exact transcription provides the solid basis for analyzing what might be hundreds of minute-but-significant phenomena in a single conversation.

4.2 Utterance as a domain of speech

4.21 Time

I shall rely on the arguments presented above in section 4.11 as evidence that the most objective means of studying speech is to chart how much each speaker said over time. It also captures some features of the rhythm of conversation that other analyses would miss. Just as above, however, I concede that knowing when speakers were speaking does not account for what they were saying, how fast they were saying it, and whether they were pausing between statements. So, I adopted time as one but not the only one of the ways in which I would define an utterance. Section 5, below, describes how I put this into practice.

4.22 Turns

Just as it would be too imprecise to use sentences as a unit of length (see 4.12, above), it would be too imprecise to define an utterance as a full sentence uttered by a speaker and then attempt to measure sentence length. However, it is not imprecise to define an utterance with an even larger unit than the sentence, namely the turn. Recall that "A Simplest Systematics" (section 1.5, above) observed that turn behavior is orderly even in informal speech: one speaker usually starts talking as the other speaker stops, sometimes with a small overlap or a small pause. If *utterance* is defined as one turn, the definition of utterance begins to sound like the time-based method for length that was so appealingly objective: the utterance begins when one speaker starts speaking and ends when, at around the same time, the second speaker begins to talk and the first speaker stops.

In the turn-as-utterance model the two conditions for defining the end of the utterance – that the other speaker start talking and that the first speaker stop talking at around the same time – are both easily and objectively indicated by the sound recording of each session. However,

there could be some objection to defining a turn (and therefore an utterance) only in turns of when another speaker joins in. Imagine that speaker A asked speaker B a question and that speaker B didn't respond at all. After five minutes of silence, speaker A finally says something else, such as "Why didn't you answer my question?" Should speaker A's turn include all five minutes of the silence, when speaker A clearly did not intend to monopolize the conversation for that time? I do not think it should.

Defining utterances in terms of turn, then, requires some judgments about how pauses relate to turn structure. This problem is not insurmountable; one simply needs to adopt a pause length beyond which one should consider any turn-in-progress to be terminated. The next time one of the speakers speaks then starts a new turn and a new utterance for the purpose of analysis. In my sessions I found that it was very rare for a speaker to pause even as long as 4 seconds between words, phrases, or sentences (things like them, at least) that specifically dealt with the same topic in the same way. This is not to say that every 4-second pause corresponded with changing the topic of conversation altogether; sometimes speakers would pause and then shift the focus within the same topic to a new area, or introduce new material. But, at least the focus was different. For the purposes of my analysis, I chose 3.99s as the maximum pause length within a single turn. Any pause of 4s or greater signaled the end of the current speaker's turn, and a new turn began when the conversation resumed, regardless of which speaker initiated the resumption.

4.23 Problems and issues

4.231 Questions

I noted above in section 3.31 that "canonical questions" involving rising inflection have been marked with question marks in the text. In defining utterances by turn length, I felt it necessary to have some provision for the current-speaker-selects-next device described by Sacks et al. (1974), in which the current speaker ends his turn by designating the next speaker to speak.

In a two-person dialogue, there would be only one other speaker, so any abdication of one's turn forces the other speaker to begin his or her turn. For the purposes of noting the time at which one speaker's turn begins and another speaker's ends, it seemed be necessary to identify points at which the current speaker used the current-speaker-selects-next technique to instantaneously end his own turn and start the other speaker's turn.

However, there did not seem to be any objective means of deciding when a speaker had used the current-speaker-selects-next technique except in the case of canonical questions, which always appeared to be addressed to the other speaker in order to prompt him or her to speak. I therefore marked these canonical questions in the transcription with question marks, and treated them as follows. When a speaker finished the last word of a canonical question, I immediately considered his or her turn terminated and immediately considered the other speaker's turn to be initiated. I followed this approach uniformly in both the time-based and turn-based approaches to utterance length detailed in section 5, below.

4.232 Interruptions and Continuations

Extremely often in conversation one speaker would be telling a story, anecdote, or other elongated utterance to the other speaker, and the speaker not talking would interrupt. Common types of interruptions included laughter, snorts or other non-speech sounds; response words or phrases such as *uh-huh, yeah, oh yeah,* or *oh right*; or contentful (multi-word) responses that may or may not have been intended to end the other speaker's turn. In some of these cases, particularly the shorter interruptions, after the interruption the original speaker would resume talking about the same or nearly the same topic. I call this a CONTINUATION. In other cases, particularly with the contentful responses but sometimes with short responses, after the interruption the original speaker's train of conversation would be derailed, and the original speaker would either resume talking about something else entirely or not resume talking at all.

These interruptions do not prove to be a problem for my time-based analysis of utterances as a domain, as my approach to organizing the data (see section 5, below) can accommodate many data points for one speaker in the same interval of time in which there is only a single data point for the other speaker. (I.e., one speaker may interrupt with as many short utterances as she likes during the other speaker's long utterance, and my time-based approach to data can still accurately display it.) However, for the a turn-based approach I had to make a choice as to what kinds of interruptions constituted the end of the interrupted speaker's turn and the beginning of the interrupting speakers turn, and what kinds of interruptions did not affect turn order.

Rather than dip into the nebulous realm of deciding which turns seemed to exhibit continuations after the interruption (based on subject change, etc.), I decided stick to the original definition of turn: a speaker's turn ends if he stops speaking at around the same time that the other speaker begins to speak. To define what counts as "beginning to speak" for the other speaker, I put my system for counting tokens to work: any interruption consisting of vocal material that was counted as one or more tokens constituted the end of the current speaker's turn and the beginning of the interrupting speaker's turn. When the interruption finished, that constituted the end of the interrupting speaker's turn. If the original speaker began speaking again after the interruption, this constituted a new turn. In each example below (14.1-14.3), I have notated the turns; the abbreviation A_1 denotes the first turn in the example for speaker A.

Ex. 14.1 Interruption not counted as its own turn: non-speech sounds

A: I XXX think I'm just gonna be a guinea pig… [turn A_1]

B: [snorts] [≠ token(s), thus A's turn continues]
A: …for long experiments instead of actually… [turn A_1 continues]

A: …having an actual job [turn A_1 concludes]
B: Can make a decent amount of… [turn B_1]

B: …money that way as a college student because you're in good sh- uh you're in health you're healthy [turn B_1 concludes]
(3, 5:25.5)

Ex. 14.2: Interruptions that do count as their own turns: response words

B: Yeah [inhales] oh Labanotation we totally learned about that in dance class [turn B_1]

A: uh-huh [= token(s), thus counts as turn A_1]

B: 'cause I had to take dance class [turn B_2]

A: right [= token(s), thus counts as turn A_2]
(2, 4:36)

Ex. 14.3: Interruptions that do count as their own turns: contentful interruptions

A: I think Batman would be at home in a car I don't think it would be funny if he drove a car even [turn A_1]

B: 'cause he drives a car [= token(s), thus counts as turn B_1]

A: even a normal car although I think a Sahara Jeep one o' the Jeeps with like umm like the survival windows and like the t- [= 1 token[13]] pull-off roof [turn A_2]
(11, 4:39.5)

4.233: Concurrent speech: buried responses and simultaneous responding

The definition of a turn, which I have equated with an utterance for the purposes of this analysis, relies on the idea that a speaker will generally stop speaking when his or her conversational partner begins to speak, allowing for a small overlap. For the most part, this assumption is born out in my transcriptions. However, once in a while the subjects would speak concurrently with one another beyond mere overlap at a transition between one speaker's turn and the other speaker's turn.

Sometimes one speaker would attempt to interrupt the flow of the other speaker's and fail, in that the other speaker would talk right over the interruption without any pause. Often these attempted interruptions were response words that would normally fit into small gaps in the conversation. When the subject attempted to respond to the speaker currently talking but misjudged when a small gap would occur, his or her response was often buried underneath the

[13] Since *t-* could not have come from *pull-off*, this must be an example of a partially deleted word arising from self-correction. It counts as a token given the guidelines described in section 4.152.

rest of the talking speaker's utterance, which would be delivered without hesitation. Such a BURIED RESPONSE is shown below as example 15. I numbered these buried responses with decimal points to indicate the fact that they fell between the speakers' other turns. The convention I adopted is that a turn of speaker A's numbered 3.1 ($A_{3.1}$) fell during speaker B's third turn (B_3), while a turn of speaker A's numbered 3.9 ($A_{3.9}$) fell during speaker B's fourth turn (B_4).[14] If there was more than one buried response during the same speaker's turn, I labeled the buried responses sequentially by tenths of a unit to preserve their order in time. (E.g. $B_{10.1}$ would occur before $B_{10.2}$)

Ex. 15: Buried responses (turn numbered as a decimal point)

A: uh-huh [A_{21}]

B: came up with a way and it's like [pause ~1.2s] see how ridiculous it is? [pause ~1s] it's like...[B_{21}]

B: ...those crazy... [B_{21} continues]
A: [drawn out] wow [$A_{21.1}$]

B: ...boxes like so complex [B_{21} concludes]

[pause ~3s]

A: yeah [pause ~1s] I would never be able to figure that out [A_{22}] **(2, 5:14.5, 21-21.1)**

The buried response type of concurrent speech poses fewer problems than what I shall call the SIMULTANEOUS RESPONSE type of concurrent speech, shown below in example 16. I was forced to come up with ad-hoc turn assignments in the face of two sets of simultaneous turns. In the second of these simultaneous turns, each speaker responded to what the other had said during the previous simultaneous turn. For the first simultaneous turn, I could use evidence from the

[14] I needed both conventions to allow me to maintain turn numbers corresponding with the order in which a speaker's turns came about. If speaker B's third turn (B_3) occurred before speaker A's third turn, during which B unsuccessfully interrupted, I would label B's buried response $B_{3.1}$ so that it would have a greater turn number than turn B_3, which would have occurred before it. If speaker B's third turn (B_3) were after speaker A's third turn, during which B unsuccessfully interrupted, I would label B's buried response $B_{2.9}$ so that it would have a smaller turn number than turn B_3, which would have occurred after it.

previous turns to help analyze it, but for the second simultaneous turn, I could not use evidence from what happened in the previous turn to analyze it, because the previous turn already had the unusual property of being concurrent speech with another speaker. This double-set of simultaneous responses occurred not once but twice within session 8 (6:48, …15-16 and 7:00, …18-19), but luckily did not appear in any other sessions.

Ex. 16 Simultaneous response (ad hoc turn assignment)

B: …and it's funny that you would need sound to … [B_{18} continues]

A: to convey that it's a picture [A_{18}]
B: …reinforce silence [B_{18} concludes]

A: yeah [A_{19}]
B: yeah [B_{19}]

[pause ~2.2 s] **(8, 7:00, …18-19)**

My analysis was as follows: In each of the instances in which this unusual behavior appeared, one of the speakers had been speaking a somewhat long utterance (*B_{18} continues*, above) to which the second speaker was responding in the first simultaneous turn (A_{18}). I treated the first simultaneous turn as a new turn for the responder (hence A_{18} where I placed it; the last thing A had said had been A_{17}), but a continuation of the previous turn for the other speaker (hence B_{18} *concludes* rather than B_{19} in that spot). In the second simultaneous turn, each speaker is responding to what the other speaker said in her last utterance, and so I counted it as a new turn for each speaker consisting of just the one word each (A_{19} and B_{19}). When speaker B resumed speaking again after the 2.2 second pause at the end of example 16, I treated it as the beginning of a new turn, B_{20}.

4.24 A note on turn assignment and transcription

The discussion in the paper is not necessarily exhaustive of the problems and issues associated with mapping turns to spoken dialogue; there are other, better sources for the systematic study of turn-taking (Sacks et al. 1974, and the research it touched off). In all cases

not addressed specifically in this section, I have used my best judgment to assign turns with turn numbers increasing over time such that, when one speaker begins speaking and another stops with either a small gap or small overlap of speech, it signals the end of the old speaker's turn and the beginning of the new speaker's turn.

It should be clear from the nature of turn assignment that an effective transcription was vital to the extraction of data (discussed in more detail in section 5, below). A speaker's turn might extend into the slight overlap where both speakers were talking; any words said during this overlap would be difficult to hear, but still count towards the length of the utterance associated with the turn about to end. It would have been impossible to accurately and repeatably note the beginnings and ends of turns and the tokens appearing therein without a detailed transcription.

5. Extracting the data

Now that the question of what *utterance length* means has been addressed, I may describe how I extracted my data from the sound recordings. I analyzed the data in four distinct manners, two based on time as the unit of length and two based on tokens (as discussed in section 4.15) as a unit of length. In each of these sets of two, one of the analyses charts utterance length against time (speaking time as utterance), and one of the analyses charts utterance length against turns (turns as discrete utterances).

5.1 Another concession: 25-turn dialogues

The extraction process described below would have been impracticably time-consuming to perform on the entirety of all six transcribed sessions. Already under pressure to begin writing as soon as possible, I decided to excerpt 1 DIALOGUE consisting of 25 turns per speaker from each session, and perform my analysis only on that portion of the entire conversation.

These 25-turn-per-speaker dialogues were intentionally constructed to be analogous to the 25-turn-per-speaker exercise created by Bruce Graham that served as the impetus for the project. Graham did not stipulate that the fifty lines exchanged by the characters in his exercise

should always be the first 25 turns of their conversation, and he specifically noted that it need not be the last 25 turns of the conversation: "End it on line 50 no matter where you are in the conflict" (Graham, 1995). I therefore semi-arbitrarily (this term to be explained two paragraphs below) selected the dialogues so that their starting points were spread out over the courses of the ten-minutes sessions: one dialogue included the first 25 turns, another the last 25 turns, the rest of them various selections from the middle. Since 25 turns of speech usually took anywhere from 2 minutes and 6 seconds to 3 minutes and 49 seconds, it was easy to spread out 7 dialogues so that they covered the ten-minute range of the recording sessions.

How did I arrive at 7 dialogues from 6 transcribed sessions? I noticed that the first half of session 9, which was the only session recorded in which the subjects knew the point of the experiment, did not seem to fit the general pattern of the data collected from the other sessions. In fact, the first half session 9 was one extremely long 25-turn dialogue, taking much longer than any of the other dialogues at 4 minutes and 47 seconds. However, the conversation of the subjects in session 9 seemed to grow more normal about halfway through the session, and so I decided to transcribe a second 25-turn dialogue from session 9 for the sake of comparison that began roughly where the first dialogue ended. That made the number of dialogues analyzed equal to 7, instead of 6. The dialogues from session 9 will be described in more detail in section 6.5, below.

The selection process for a dialogue was based on a few other factors than simply its placement within the ten-minute spectrum of the recording session, thus the process was not entirely arbitrary. First, although I did not always begin dialogues at the very beginning of the speakers' conversation, I did attempt to begin dialogues at what seemed to be initiations of new topics, or at least new tacks on old topics. This was to attempt to pick a dialogue that might begin at the start of something analogous to a new theatrical idea, in order to make the dialogues more comparable to a playwriting exercise. Typically the first lines of the portions of

conversation I excerpted as dialogues occurred after a pause, and introduced a new topic or piece of information into the conversation. Thus the dialogue I selected to reflect the last couple of minutes of the 10-minute recording session did not actually include the last turn of each speaker before I reentered the room to end the session. I chose to begin the dialogue in question at a new topic that happened to be 26 turns from the point at which I reentered. I believe that this still reflects the last two minutes of the session accurately enough for the purposes of my project.

One other concern biased my selection of portions of the conversations to be selected as dialogues for my analysis: pertinent data. I chose the dialogue from session 8 to include both instances of the simultaneous response phenomenon described in section 4.233, because it was a feature of conversation unique to that portion of that session, among all of the data I recorded. Similarly, there was a short exchange in session 6 in which the speakers alternated 1-token utterances back and forth for five turns – one of the very features I originally doubted about Graham's template. I intentional chose to include this data in the dialogue for that session.

I hope that the reader, rather than being discouraged by this bias in my selection, will attribute it to picking the most interesting and most pertinent data from a vast sea of data for analysis. These were my intentions in the face of having too little time to analyze all of my data (the unbiased procedure).

5.2 Extracting raw data from each dialogue

Referring to the transcription of each session to remain consistent, I examined the graphical waveform of each of the dialogue's sound files and extracted four types of data for each utterance: (1), the time at which the utterance began; (2), the time at which the utterance finished; (3), the length of the utterance in seconds; (4) the length of each utterance in tokens. For items (1) and (2), the unit was the number of seconds from the beginning of the dialogue. Because the dialogues were all situated differently relative to the ten-minutes of the whole recording session from which they were excerpted, the only fixed temporal reference point

relevant to any given dialogue was the point at which the dialogue began. Item (3) involved a simple subtraction of item (1) from item (2), because the temporal length of an utterance by definition is the time at which it is finished minus the time at which it begins. Item (4) was counted using the transcriptions and the conventions discussed above in section 4.15 and its many subsections as references.

5.3 Constructing the graphical representations of the data

I constructed four graphs for each dialogue, corresponding to the four analysis strategies briefly mentioned at the beginning of section 5.

5.31 Utterance Length (time) vs. Time – the objective analysis

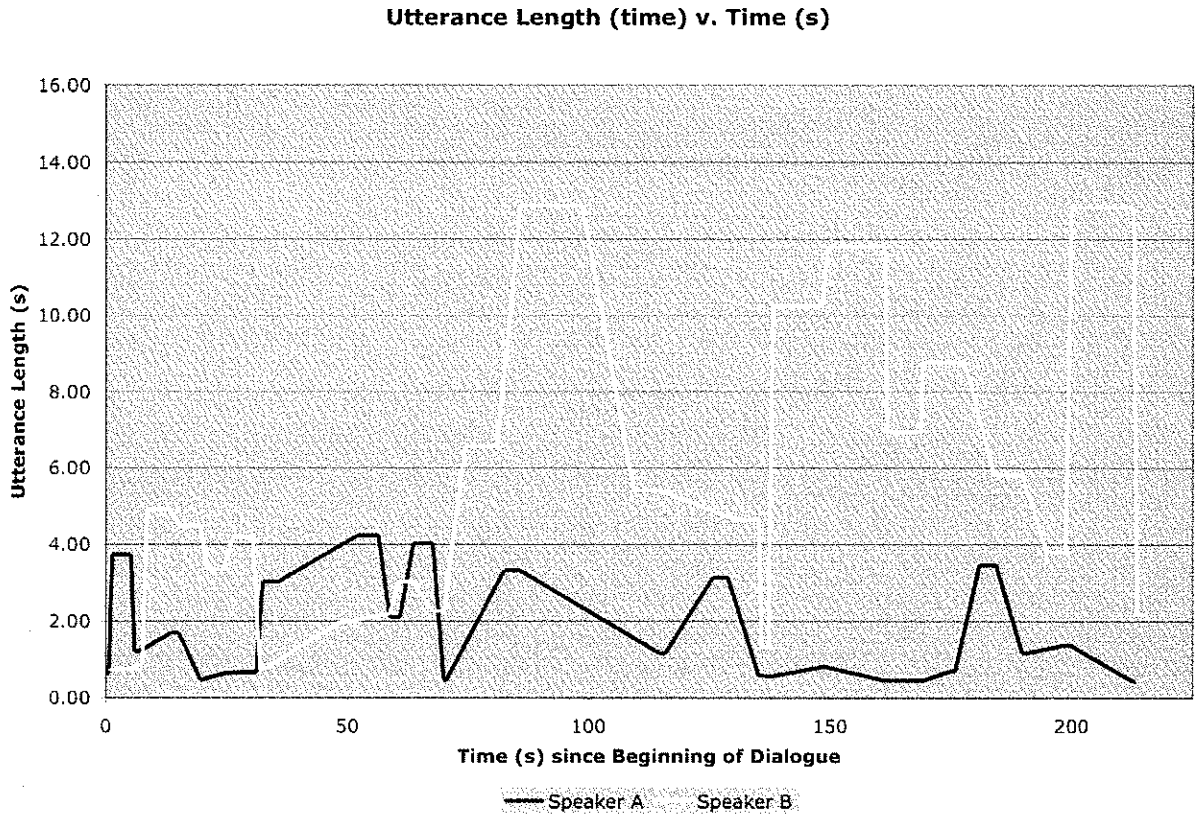
Recall that time is the most objective means of measuring length, and the most objective means of defining utterance (see sections 4.11 and 4.21). The first graph I constructed utilized time in both capacities, and so was intended to be the most objective graph possible to serve as a point of comparison for the other graphs. In order to construct the graph I performed an unusual statistical operation: I created a scatterplot in which each utterance was assigned two data points instead of one. The first data point for each utterance had as its X-value the time at which the utterance began (item 1 from section 5.2). The second data point for each utterance had as its X-value the time at which the utterance finished (item 2 from section 5.2). The Y-value for both data points per utterance was equal to the length of the utterance in seconds (item 3 from section 5.2). Hereafter I shall omit the phrase *from section 5.2*, but continue to identify the items used in each of the graphs.

Why would I do this? Well, in effect I was creating a graph of horizontal lines. These horizontal lines would be as long horizontally as the speaker had been talking for each turn, and they would be situated vertically so that longer utterances would appear to be taller on the graph, because their Y-values would be greater than the Y-values of shorter utterances. Those

experienced with statistics and graphs will notice that the Y-value of each horizontal line corresponds to the length of the horizontal line along the X-axis, although the scale is different.

The visual effect such a graph is the following: by interpreting the horizontal lines of the graph as time in which the speaker is talking, and slanted lines as times in which the speaker is not talking, one can actually look at a graphical representation of the course of the whole dialogue as it unfolded in time. Tall, wide plateaus mean that a speaker was in the middle of a big, long utterance; short, narrow plateaus denote short utterances; and long slanted lines mean that the speaker was silent for an extended period of time. A sample graph of this type is given on the next page as chart 1. It depicts the dialogue from session 2, which contained the first 25 turns of the subjects' conversation.

Chart 1: Session 2 Dialogue – Utterance Length (time) v. Time (s)



Note that the speakers' conversation was extremely LOCALLY IMBALANCED, meaning that at almost any given time, one speaker's utterances are considerably bigger than the other's. This dialogue also exhibits GLOBAL IMBALANCE, meaning that, over the course of the whole dialogue, one speaker has considerably greater utterance lengths than the other. While for the first 75s or so, the conversation appears to be globally balanced, after that point speaker B (in yellow) begins to employ considerably longer utterances than speaker A, giving the whole dialogue a lopsided look. Remember that, because using time as a unit or a domain does not account for rate of speech, it is possible that speaker B is not dominating the conversation in terms of information presented; she might just speaker much slower than speaker A.

5.32 Length of Utterance (time) v. Turn – a proof of concept for "fairness"

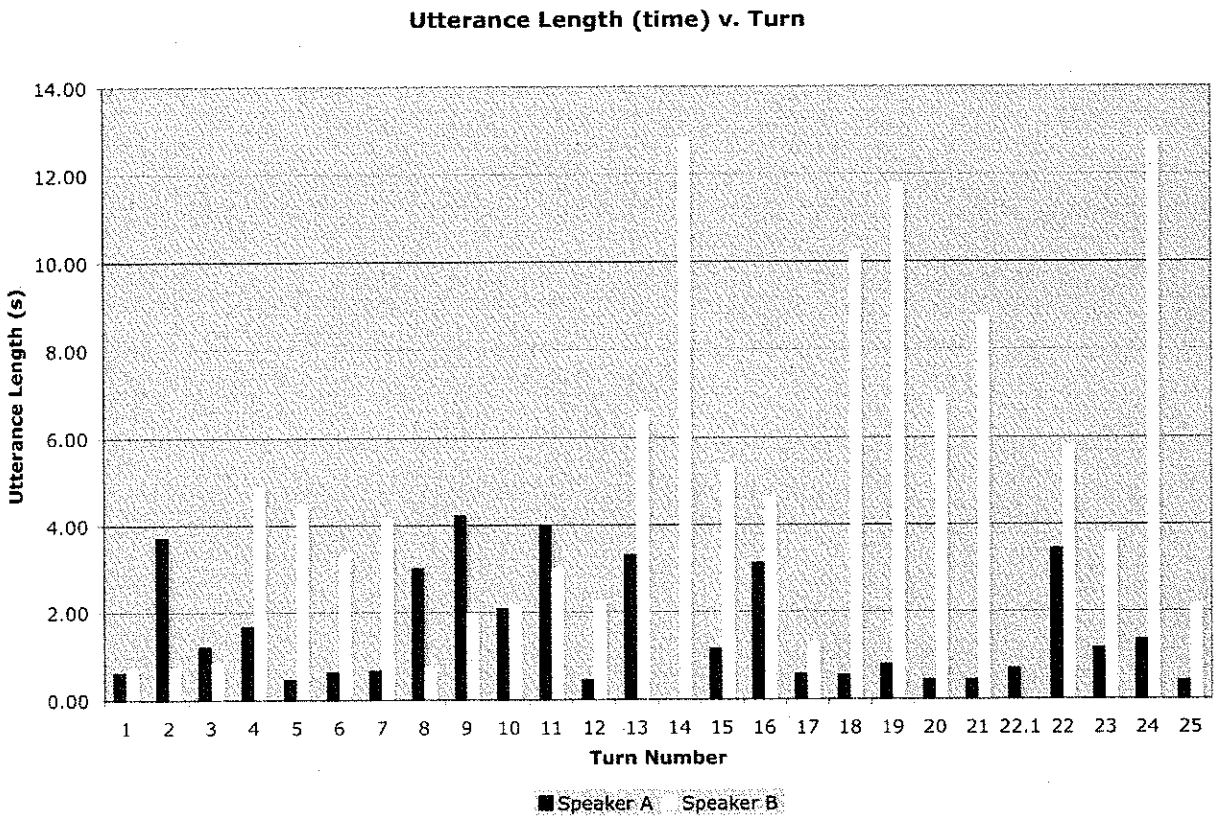
The second kind of graph, shown below as chart 2, is really just a simplified way of looking at the first type of graph. In this bar graph, the height of each bar (Y-value) is equal to

the length in seconds of each utterance (item 3). The X-value of each bar is the turn number I assigned to it according to the conventions described in section 4.22 and 4.23, above. Don't forget that, since turns are based on when speakers started and stopped talking, they are in some sense just variations of the time-based domain used for the X-axes on the graphs of the type described in section 5.31, which also depict when speakers started and stopped talking.

The bars have been paired for each turn to facilitate comparison of the speakers' utterance lengths. Do not be fooled into thinking that speaker A's utterance on a given turn always preceded speaker B's utterance. Speaker B's turn 4, for example, might precede speaker A's turn 4. I shall call the distinction of speaking a turn of one number before the other speaker speaks his or her turn of the same number the TURN INITIATIVE. Sometimes the turn initiative switches between speakers because one speaker abdicates his turn, i.e. remains silent. In my analysis, any pause of longer than 4 seconds constitutes abdicating one's turn, and so the speaker initiative was determined by who began to speak first after the pause.

Below in chart 2 is the session 2 dialogue graph of Utterance Length (time) v. Turn:

Chart 2: Session 2 Dialogue – Utterance Length (time) v. Turn



Note how similar this graph looks to the one based on time as a domain in the section above. This is a good sign: the more closely that the turn-based graph resembles the objective time-based graph, the more confidence one can have that turn-based analyses of the data are "fair," in that the represent the data as it appeared in the actual conversation. Of course, it looks similar in part because the Y-axis shows the same data as the previous graph in 5.31: utterance length in seconds. All that has been changed is the way that the data are grouped along the X-axis.

Certain of the bars in this graph may seem closer together or farther apart than the corresponding plateaus in the previous graph, but in general the same features are clearly visible: the local imbalance, the balance for the first part of the dialogue (until about turn 12), and the marked imbalance in the second half of the dialogue, outlining a large global imbalance in the dialogue.

On the whole, the X-axis does not appear to be too skewed, with one exception: the turn-based graph doesn't capture the fact that long utterances take up more time than short ones. Speaker B's long utterances in the second half of the dialogue stretch out the conversation so that one half of the turns take up about two-thirds of the time. (Compare the point at which speaker B's utterances grow to a much larger size than speaker A's in the two graphs, and you will see that it is about one-third of the way along in the time-domain graph, but one-half of the way along in the turn-based graph.)

Is one of these graphs better? Well, it depends upon whether you are concerned with the extent to which the pace of the conversation slows when the speeches of one character grow longer, or whether you are more interested in the number of long speeches the characters have compared to the number of short speeches. Sort of the difference between stage time and lines, in a way. They may be related, but they are not exactly the same thing: more lines does not necessarily equal more stage time, though it can very well contribute to more stage time. There is, however, a better way to measure the length of the speakers' lines in the sense that actors typically think of line length: as the number of tokens spoken.

5.33 Utterance Length (tokens) v. Time – the linguist's analysis

A third type of graph, introduced as chart 3, below, specifically addresses the number of tokens spoken by each speaker over the course of the dialogue. It is a scatter plot again, consisting of horizontal lines just like the graph-type described in section 5.31. Again, each utterance has two data points instead of one. The X-values for the two points, just like above, are the start and end times of the utterance, respectively. In this case, instead of the Y-value for each point being the length of the utterance in seconds, it is the number of tokens the speaker said during the utterance.

What will this graph tell us? It actually includes more information than the graph in 5.31. Just like above, the horizontal lines show us when in time the speaker is talking, and the slanted

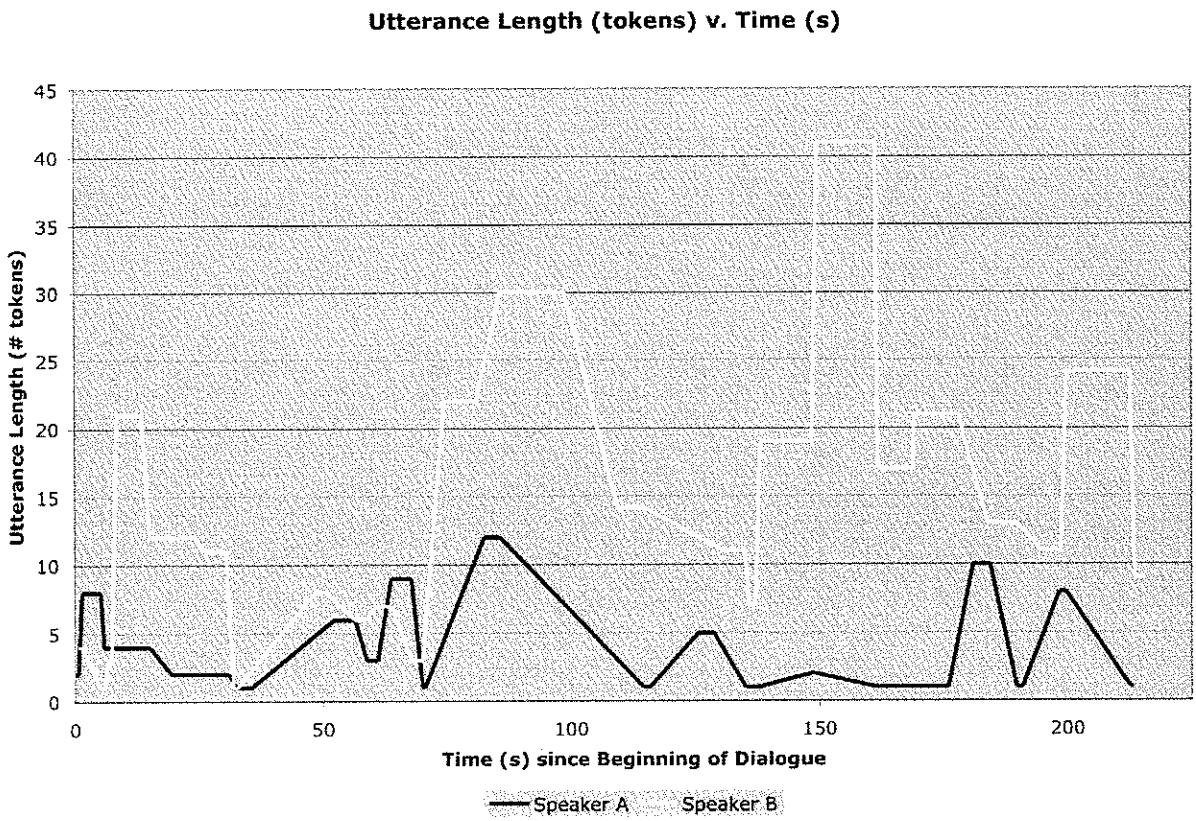
lines show us when he or she isn't. Unlike the graph in 5.31, the height of the horizontal lines tells us new information (the number of tokens used by the speaker per utterance) instead of repeating the same information (the length of time he or she was speaking). You see, the Y-axis in the type of graph introduced in section 5.31, although useful for visual understanding, is actually redundant, because it encodes on the Y-Axis the same information as the width of the plateaus along the X-axis. In the type of graph introduced in this section, however, the Y-axis is not redundant, because it measures length in a different way than the X-axis: in tokens instead of time. So, tall plateaus denote wordy utterances and wide plateaus denote utterances that took a long time. A short plateau includes only a few words, while a narrow plateau indicates a very brief utterance in time.

Of course, it seems to make sense that the temporal length of an utterance should be somewhat dependent on the number of tokens in it, so we should tend to expect that tall plateaus should also be wide, and short plateaus should also be narrow. But the most interesting thing about this type of graph is that there is the possibility of observing exceptional utterances: short, wide plateaus that indicate very slowly delivered but not verbose utterances, and tall, narrow plateaus that indicate very rapid gluts of words in a single utterance. For this reason, I propose that this type of graphical representation of conversation is the most useful for the linguist, who is presumably interested in not only what the speaker says but the rate at which he says it, how the conversation unfolds over time, and how both of those things might change over the course of a conversation. The example graph is given below as chart 3. Again, I shall continue to provide graphs of the session 2 dialogue for comparison. Note that the Y-axis has only counting numbers now, not decimals, because I did not count partial tokens.[15]

---

[15] As a means of keeping responsible numerical data, I noted the number of unrecoverable words per turn as a decimal. These turns' numerical entries had the form [# tokens].[# unrecoverable tokens]. Given the scale for the token counts (as high as 70), these decimals should be imperceptible; they were for internal data consistency only.

Chart 3: Session 2 Dialogue – Utterance Length (tokens) v. Time (s)



If we compare this to the graph in 5.31, we see that some of the plateaus have changed height, but all of them occupy the same horizontal space. This is because the X-values are the same in both graphs; only the Y-values have changed. One of the more interesting features of this graph is that the balance that seemed to exist between the speakers for the first 75 seconds or so of the dialogue does not show up on this graph. We can infer that in the period from about t = 30s to t = 75s speaker B must have been talking at a faster rate of tokens per second than speaker A, because while speaker A spoke for more time, the speakers use the same number of tokens. It is possible that speaker A was using larger words that take longer to say than speaker B's words did, or perhaps speaker B just speaks faster in general. Other than this difference between the graphs, they seem quite similar: each shows local imbalance and global imbalance across the entire dialogue. In section 4.11, above, I claimed that the length of a speaker's turn in seconds

was at least somewhat related to the amount they had to say in tokens. The similarities between charts 1 and 3 – representative of the similarities between both types of graphs throughout my results – provide evidence to show that this claim is reasonable.

Though not equivalent, once again it seems that counting utterances by tokens rather than time produces a similar enough result that it must be a "fair" way to look at the data. Particularly since I have proposed that this graph is the most useful for the linguist to consider for the analysis of speech, it's nice to know that it works reasonably well in practice in addition to being theoretically appealing. However, playwrights usually think of dialogue in terms of a series of back-and-forth lines composed of words; chart 3 still doesn't quite capture that.

5.34 Utterance Length (tokens) v. Turn – the playwright's analysis

At last we turn to the playwright's analysis: tokens per turn. Since a turn is defined by one speaker delivering what he has to say, and then stopping around the time that the other speaker begins to talk (barring some elongated pause), it is reasonably close in definition to a line of dramatic dialogue in a play, which is usually considered a unit of one character's speech. The line ends on the page when the next character says something or when the playwright writes in a pause, stage direction, or other interruption. The playwright has little control over the rates of his actors' speech, their decisions about when and where to pause, or even when they will speed up or slow down for emotional effect. Yes, the playwright can and does attempt to manipulate these things, but his tools are the words on the page: their forms, their feel, their tone. These are the things I attempted to be faithful to when constructing my hybrid analysis (section 4.15). Furthermore, the number of tokens per turn is my (hopefully more precise) analog to Graham's words per line in his exercise, and thus will serve as the means by which I can compare my real-life data to his hypothesized utterance lengths.

Chart 4, below, is a (bar) graph of this final type. The Y-axis is the same as that of the graph discussed in the previous section: the number of tokens per utterance. The X-axis is the

same as that of the type of graph explained in section 5.32: the number of the turn in which the utterance was spoken as defined by the conventions in sections 4.2 and 4.3, above. Again, this graph loses site of real time differences in speech rate and short pauses between portions of the same utterance, as well as tending to "flatten" out the slowing of long utterances. However, it provides a concise "play-like" map of utterance lengths over the course of a dialogue.

Chart 4: Session 2 Dialogue – Utterance Length (tokens) v. Turn

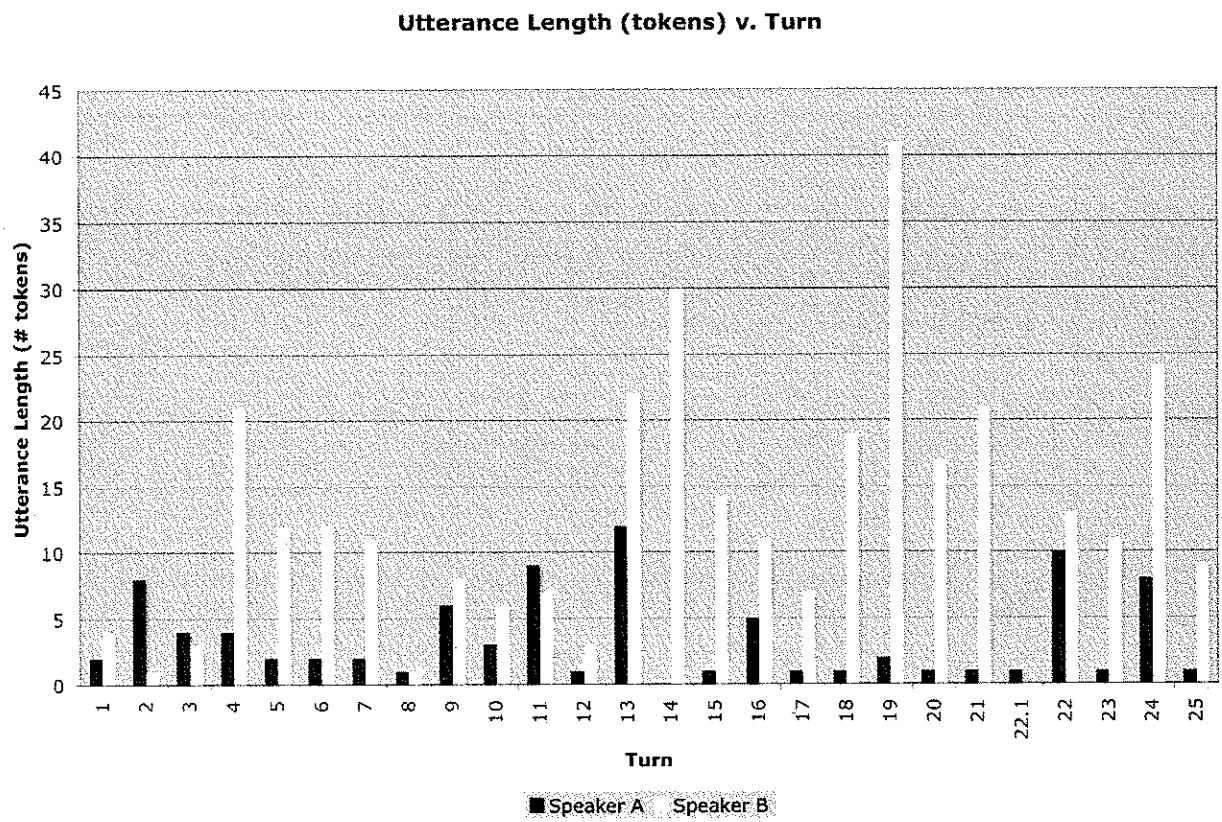


Chart 4 is to Chart 3 as Chart 2 was to Chart 1. That is to say, the Y-values in this chart are the same as in the chart from the previous section, Chart 3. However, the X-values are different from Chart 3; in fact, they are different in the same way that the X-values of Chart 2 were different from Chart 1's X-values: it can be seen that the point at which Speaker B really dominates the conversation seems to be at the halfway point by turn, but about 1/3 of the way through by time. Again, this is because the turn-based approach does not reflect that fact that Speaker B's long utterances later in the dialogue take longer to say that shorter utterances at the

beginning. The turn-based analysis treats every turn as taking an equal amount of time, in effect taking time out of the equation as a consideration of how the conversation progresses from beginning to end. Instead, it highlights the differences in the lengths of the utterances by eliminating the things that the playwright has no direct control over, such as delivery time.

To take a step back and get a hold on exactly how "fair" this representation of the data may be, compare Chart 4 to Chart 1, the most objective means of looking at the data. Other than the fact that Chart 4 is a bar graph (which is just for convenience, not out of necessity) and Chart 1 is a scatter plot, Chart 4 has been transformed along the X- and Y- axes. It has a different, more simplified way of distributing the order of utterances (turns instead of real time), and it has a different way of measuring length (tokens instead of time). Despite these differences, however, observe how the primary features of the graph still do not change. The local and global imbalances persist, and in fact it becomes clearer than ever that, in terms of words spoken, Speaker B really did dominate the conversation almost completely: speaker A spoke as many or more words as Speaker B on only two turns throughout the entire dialogue.

Any differences between the time/time approach and the token/turn based approach seem more often to be exaggerations than reversals. The latter seems to retain the significant features of the former, and may point up some features more pertinent to the playwright's interests. The fact that Speaker B always said more words than speaker A except for in two instances would be a very strong stylistic consideration, if a playwright were to construct a scene based on the numbers of tokens spoken by the subjects in the session 2 dialogue. I believe, for these reasons, that the graphs of the Chart 4 (tokens v. turns) type are a "fair" and relevant graphical representation for the analysis of conversation. Using them, it is finally possible to address the accuracy of Graham's template against real life conversations in a meaningful manner.

6. Results and analysis: the features of natural dialogue

As I have outlined in section 5, the analysis strategy that produced the fourth type of graph – Length of Utterance as measured in tokens vs. Turn – is the one best suited to evaluate Graham's template. I shall therefore limit this results section to observations arising from this strategy alone. There are, perhaps, many other observations to be made from the other strategies discussed in section 5.31-5.33, and graphical representations of the data structures produced by those strategies are available along with sound files and transcriptions, for those interested in further research. However, in the interest of addressing Graham's template as specifically as possible, I have deemed them beyond the scope of this paper.

6.1 Locally imbalanced speech

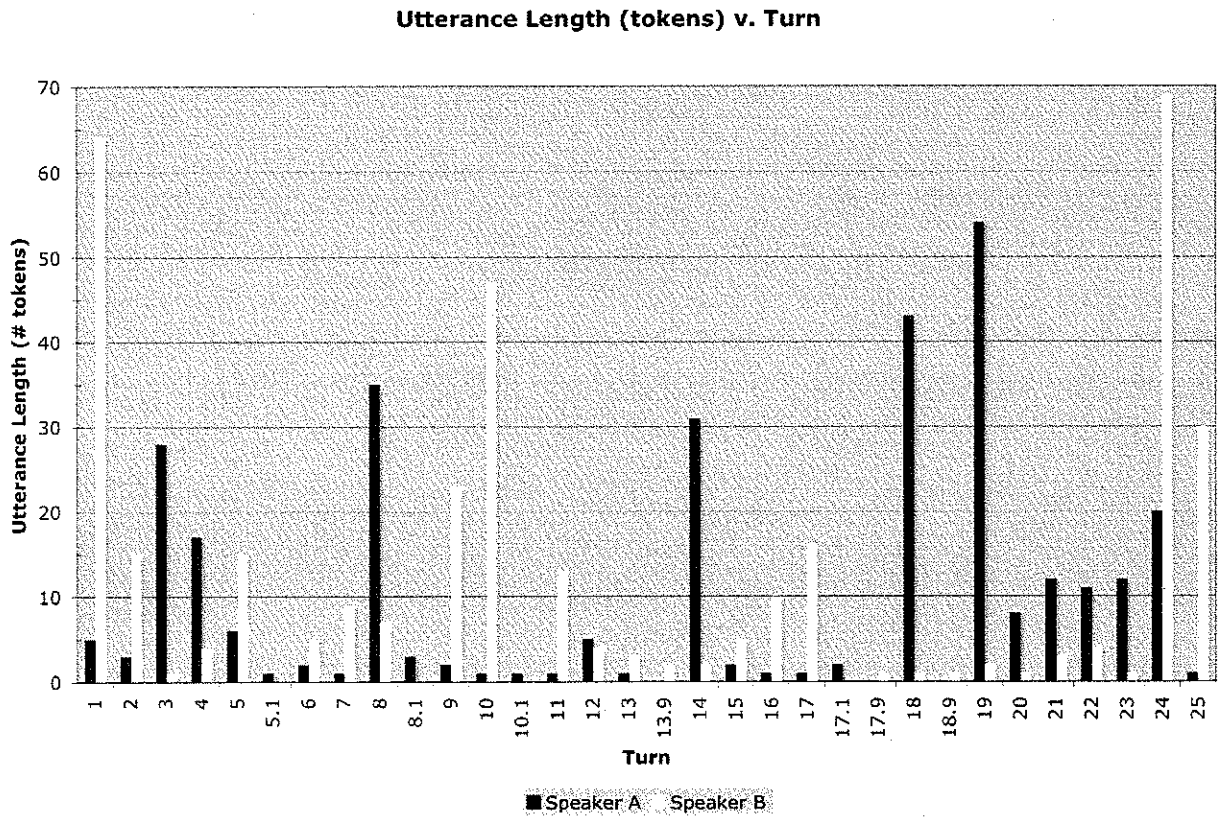
As the data from session 2 showed in the previous section, natural dialogue tends to be extremely imbalanced locally. This means that, given any few turns, one speaker tends to use far more words than the other. However, the distinction of being the speaker with longer utterances can switch back and forth between the speakers as the dialogue unfolds. I shall call the speaker whose utterances are longer for a given portion of a dialogue the DOMINANT SPEAKER, and I shall call the distinction of being the dominant speaker DOMINANCE.

The chart on the next page (chart 5) illustrates the number of tokens per turn used by both speakers in a dialogue representing the last 4 minutes or so of the third recording session. The speakers were two first-year males who had never met each other before the experiment. Perhaps because they were strangers, their turn-taking behavior was relatively orderly, and their speech was without question locally imbalanced. At almost no point is it unclear who the dominant speaker is,[16] although speaker B is only weakly dominant from turns 5-7 and neither

[16] Remember that decimal points denote speech that fell during the other speakers' uninterrupted turn, hence the apparent dominance by the speaker on any turn labeled with a decimal should be disregarded as an artifact of the notational system. The utterances marked with decimal points represent failed attempts by a speaker to begin his or her turn, and so I do not consider them in the discussion presented here.

speaker seems to assert dominance for turns 12-13. Otherwise, the bars of one speaker are very short when the other speakers' bars are long.

Chart 5: Session 3 Dialogue – Utterance Length (tokens) v. Turn



6.11 A pattern

It is not the case that the dialogues were locally imbalanced in a random fashion. Speaker dominance emerged in a patterned manner: a speaker would assume dominance with an extended utterance or series of utterances, the lengths of which would at first increase and then decrease, often finishing with a very short utterance. After this often-one-word final utterance, the other speaker might assume dominance with an extended utterance or series of utterances following the same pattern.

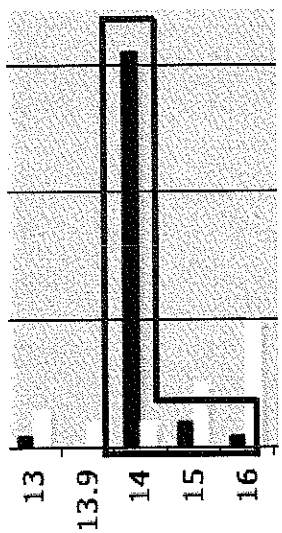
6.111 The BEAT: simplest form

Borrowing a theatrical term, I shall call this emergent pattern a BEAT. In its simplest form, the structure of a beat consists of one long utterance by a speaker, at least 20 tokens in

length but often much longer (60+ token utterances appeared in my dialogues), followed by one or more very short utterances, usually under 5 tokens and often just 1 token in length. I call the long utterance the BEATCENTER, as it is the most prominent feature of the beat, and I call the set of short utterances that follow the beatcenter the BEATENDER[17], since they seem to act as a signal that the beat is over and the other speaker may begin his or her new beat. Since silence serves to indicate that a speaker's beat is over as well or better than a 1-token beatender, I have analyzed silence on the turn following a beatcenter as a 0-token beatender.

Speaker A in chart 5 above exhibits two beats of this simplest type, one from turns 8-9 and one from turns 14-16 (detailed in figure 9, below). While one speaker's beat is in progress, the other speaker tends to minimize his or her utterance length with responses equal to or fewer than 10 tokens in length, and often between 1 and 3 tokens in length. An easy means by which to recognize a beat in its simplest form, as I have defined it above, is that if one were to draw lines around the bars that represent the utterances of the beat, they would form an L-shape, as illustrated in figure 9 below, a detail of chart 5.

Figure 9: The L-shape of a beat in its simplest form



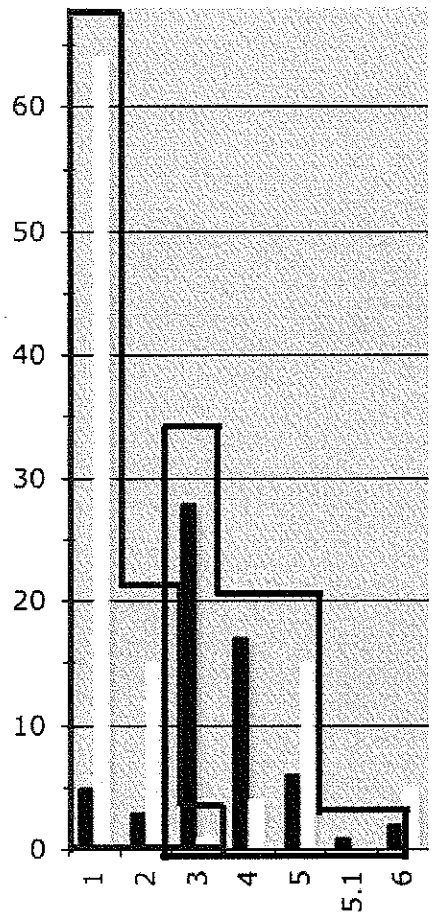
[17] These terms, like those I will employ to describe the other features of a beat, are designed to be transparent at the expense of brevity; hence, the term *beat* appears in all of them.

6.112 The POSTBEAT

Of course, examining chart 5 above it is clear that not every period of locally imbalanced speech looks exactly like the L-shaped beat shown in figure 9. Sometimes, after the extended utterance constituting the beatcenter, the speaker would continue with another long utterance or set of utterances. These utterances were longer than I have defined beatenders to be, but typically no more than half as long as the beatcenter that preceded them. I defined the term POSTBEAT, to capture these utterances in my analysis, as an utterance or set of utterances falling after the beatcenter and before the beatender, (each) varying in length from 5 tokens to about half as long (occasionally a little longer) as the beatcenter immediately preceding them.

As outlined below in figure 10, Speaker B's turns 1-3 in chart 5 constitute a beat containing a beatcenter, postbeat, and beatender, as do Speaker A's turns 3-6. Speaker A's utterances on turns 1and 2 are under 5 tokens in length; his turn 3 constitutes the beatcenter of his own beat as he assumes dominance. Because of this, it seems that the beatender of Speaker B's initial beat is omitted, in a sense. However, we might equally well say that the short response Speaker B utters on his turn 3 constitutes the beatender of his previous turn because, even though it overlaps with Speaker A's new beat, it clearly shows that he is no longer the dominant speaker. I have followed the latter strategy. By this reasoning, the beat begun by Speaker A on turn 3 extends until his turn 6; the postbeat of this turn includes two utterances (turns 4 and 5) instead of just one. In figure 10, Speaker B's initial beat is outlined in red and speaker A's subsequent beat is outlined in indigo.

Figure 10: Two beats each containing a beatcenter, post-beat, and beatender



The L-shape of a beat in its simplest form becomes a sort of staircase when a postbeat is present. Take notice of Speaker B's turn 5, which is longer than a typical response during the other speaker's turn. Perhaps because speaker A's utterance right before was only 6 tokens long, speaker B assumed the beatender was coming and made an attempt to start his own beat. The means by which he did so are explained in the next section.

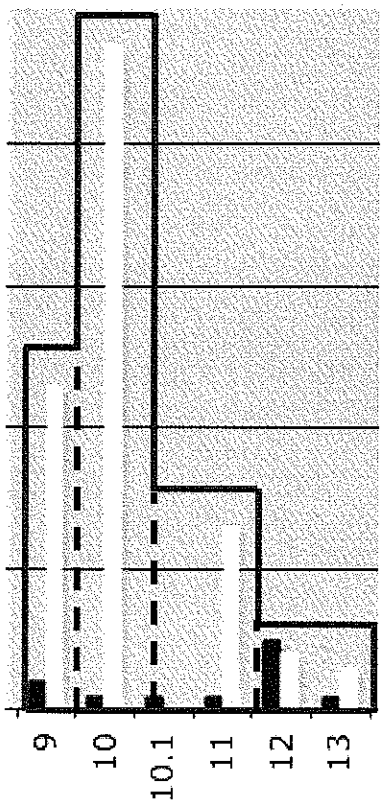
6.113 The PREBEAT

In addition to filling out the beat with utterances after the beatcenter, sometimes speakers would build up to a beatcenter with an/some introductory utterance(s). These utterances typically were long enough to signal that the speaker was attempting to assume dominance, but they were not as long as the beatcenter that followed them. I defined the term PREBEAT – meaning an utterance or set of utterances immediately preceding the beatcenter, the length of

(each of) which is between about 10 tokens and the length of the beatcenter – to capture these utterances in my analysis.

Referring to chart 5 again, Speaker B's turns 9-13 and Speaker A's turns 18-25 have examples of prebeats. Speaker B's turns 9-13 map very nicely onto the terms I have defined for beat structure: Turn 9 is the prebeat, turn 10 the beatcenter, turn 11 the postbeat, and turns 12-13 the beatender. This beat is outlined in red below in figure 11, with black, dashed lines dividing up the parts of the beat:

Figure 11: One fully expanded beat: prebeat, beatcenter, postbeat, beatender



The graphical representation of a fully-expanded single beat – including a prebeat, beatcenter, postbeat, and beatender – no longer looks like a L-shape. But, its shape is fairly distinctive nonetheless. Though this shape recurs throughout the data from each session, there are some portions of the data that do not fit the beat-structure I have described.

Speaker A's turns 18-25 are such a portion. We can see that turn 18 is a prebeat (a rather long one) to the beatcenter on turn 19, and turn 20 is pretty clearly a postbeat. After that, the

mapping breaks down. The feature of these turns that looks most unusual in light of my discussion so far is the 20-token turn 24, which occurs after four medium-length utterances (turns 20-23) that look like the postbeat of the beatcenter falling on turn 19. If we didn't look at turns 18-20, however, turns 21-23 might very well look like prebeats to turn 24, which is long enough to count as its own beatcenter. In fact, I have analyzed beat-structures of this form as 2 conjoined beats under the term RENEWED BEAT in the next section.

Before continuing, let me point out that a beat can include a prebeat without including a postbeat, but none of the beats in chart 5 display this possibility adequately enough to serve as an example. See chart 8 of the dialogue from session 6, below, for examples of this (turns 10-13 for Speaker A, and 21-23 for Speaker B).

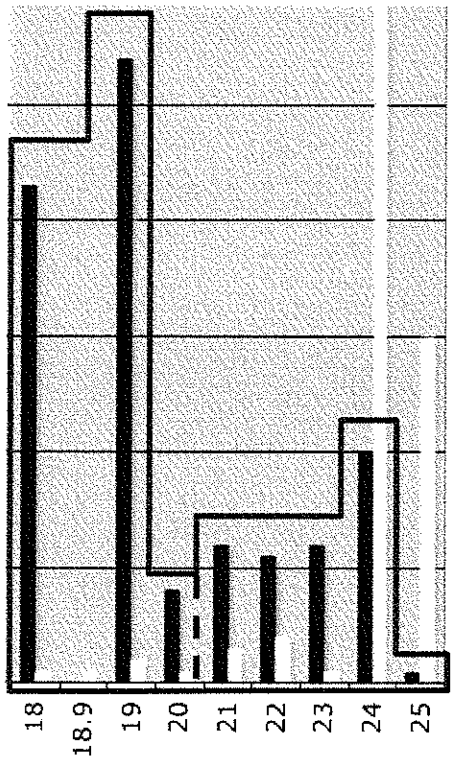
6.12 Other emergent structures

6.121 The RENEWED BEAT

Look again at turns 18-25 for Speaker A in chart 5, detailed on the next page in figure 12. As I said, turns 21-25 could be analyzed as their own beat: 21-23 prebeat, 24 beatcenter, 25 beatender. However, turns 18-20 appear to start a beat (18 prebeat, 19 beatcenter, 20 postbeat) that does not conclude before turns 21-25 begin. In light of this, it is unclear whether we should analyze turns 21-24 as a postbeat to turn 19, or whether we should consider them a new beat altogether. Well, I chose to do neither: to reconcile the seeming-independence of the second beat with its clear relationship to the first, I lump the entire structure, turns 18-25, under the term RENEWED BEAT. A renewed beat can be defined as any two contiguous beatcenters by a single dominant speaker connected by utterances that do not neatly map to the form {(postbeat) beatender (prebeat)}. Even if we analyze turn 20 as postbeat to the beatcenter on turn 19, and analyze turns 21-23 as prebeat to the beatcenter on 24 (the cleanest analysis I can think of for these turns), there is no beatender for the beat consisting of turns 18-20, and so turns 18-25 would fall under the definition of renewed beat.

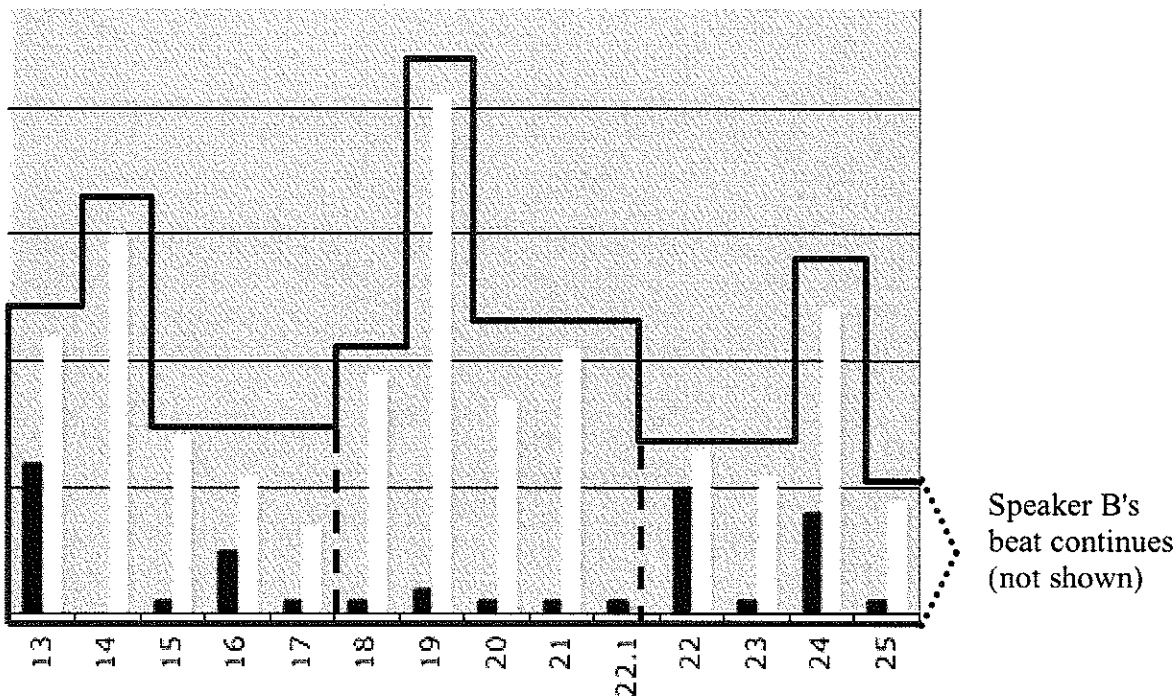
Figure 12 below details turns 18-25 from chart 5. The renewed beat is outlined in red, with a black, dashed line denoting the point at which the beat was renewed, i.e. the point at which it can begin to be analyzed as a second beat.

Figure 12: A renewed beat



For the dubious reader, unconvinced that the best way to analyze such structures is as two conjoined beats, I include below a detail of a doubly renewed beat from the session 2 dialogue, shown above as chart 4 in section 5.34. In the following example (figure 13) – illustrating turns 13-25 of chart 4 – we can clearly see a decline in Speaker B's utterance length in the postbeat (turns 15-17) before the beat is renewed with a new prebeat (turn 18) that leads to the second beatcenter (turn 19). However, speaker B isn't finished: the postbeat of the second beatcenter (turns 20-21) smoothly transforms into what looks like a prebeat (turns 22-23) to a third beatcenter (turn 24), which is followed by a portion of a postbeat (25). The 25 turns constituting the dialogue concluded, however, before the rest of this beat was uttered. The renewed beat is outlined in figure 13 below, with black, dashed lines indicating where each new renewal occurs.

Figure 13: A doubly renewed beat – proof of concept for this analysis' merit[18]



The jump in Speaker B's utterance length between turns 17 and 18 suggests that the first beat (13-17), which is petering out by turn 17, regains momentum on turn 18 to get to a second beatcenter at turn 19. More subtly, the drop in Speaker B's utterance length between turns 21 and 22 made me analyze turns 22 and 23 as prebeat to a new beatcenter at turn 24.

Without some sense of beat renewal, the structure of speaker B's utterances above would be just random noise. After all, the length of each utterance taken individually seems to pop up and down without much of a pattern. However, when the utterances are grouped into beat structures centered on the tallest peaks (longest utterances), a pattern consisting of three relatively well-defined shapes emerges – likes those we saw before. These beats are smashed into one another, and the last one is unfinished, but still the original beat shape of each is visible, providing a window into how such a complicated utterance length structure could evolve. Figure

[18] Note that the scale is different in figure 13 than in chart 4, above: each horizontal black line is equal to 10 tokens rather than 5 tokens. Figure 13's scale is comparable to that of chart 5; I have used this same scale for all of the data in section 6, for ease of comparison.

13 above, then, stands as a proof of concept for why my beat-based analysis is reasonable, and why it might be useful for the analysis of dialogue.

6.122 MACROBEAT STRUCTURE?

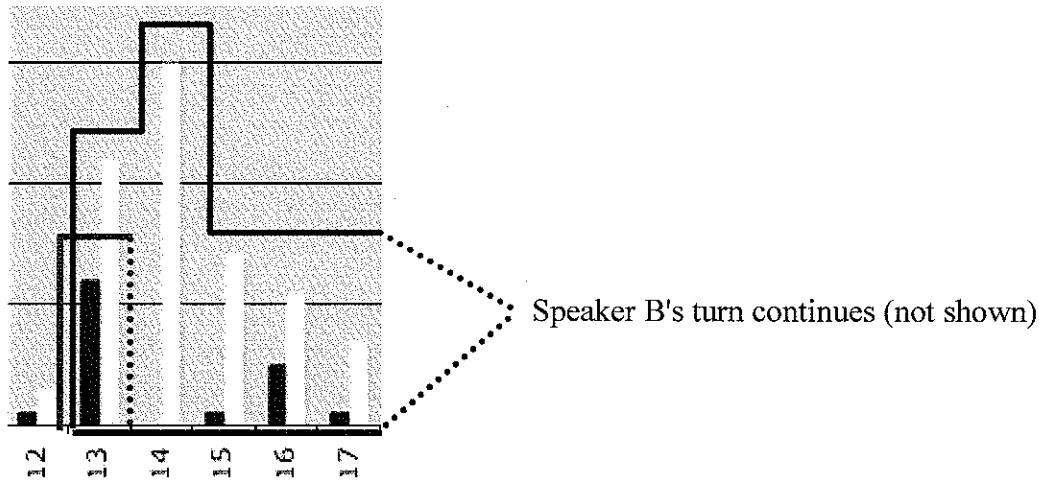
As one further point of interest, note the length of the beatcenters of each of Speaker B's three enjambed beats shown in figure 13. The first one is three quarters as long as the middle one, and the third one is about half as long as the middle one. The beatcenters of this renewed beat, taken alone, resemble the beat structure of a single beat: the first beatcenter vaguely resembles a prebeat to the second, tallest beatcenter, and the third beatcenter vaguely resembles a postbeat to the second, tallest beatcenter. It is possible that extended renewed beats – when one speaker remains dominant for the course of 2 or more beats without ever uttering a beatender – may exhibit some sort of MACROBEAT STRUCTURE that is identical to the beat structure I have developed: {(PREMACROBEAT) MACROBEATCENTER (POSTMACROBEAT) MACROBEATENDER}. Instead of utterance lengths forming the basis of the hierarchy, however, BEATCENTER LENGTH might fill that role. This topic recurs in the discussion of the session 11 dialogue in section 6.21, below.

6.123 Interrupted beats

Occasionally a speaker who seemed to be uttering a prebeat would never reach the beatcenter toward which they appeared to headed because the other speaker would preempt them. For Speaker A, the turn just before Speaker B's doubly renewed beat, discussed in the previous two sections, seemed to fit such a description. Shown below in figure 14, a detail of Chart 4, Speaker A's turn 13 is 12 tokens long, making it seem like a prebeat that was supposed to signal that Speaker A wished to assume dominance. However, speaker B did not seem to notice, and barreled ahead with her own extremely long period of dominance. Speaker A, as a well-behaved conversational partner, did not pursue her preempted prebeat, instead reducing her utterance lengths to fewer than 5 tokens for the next several turns. Speaker A's interrupted beat

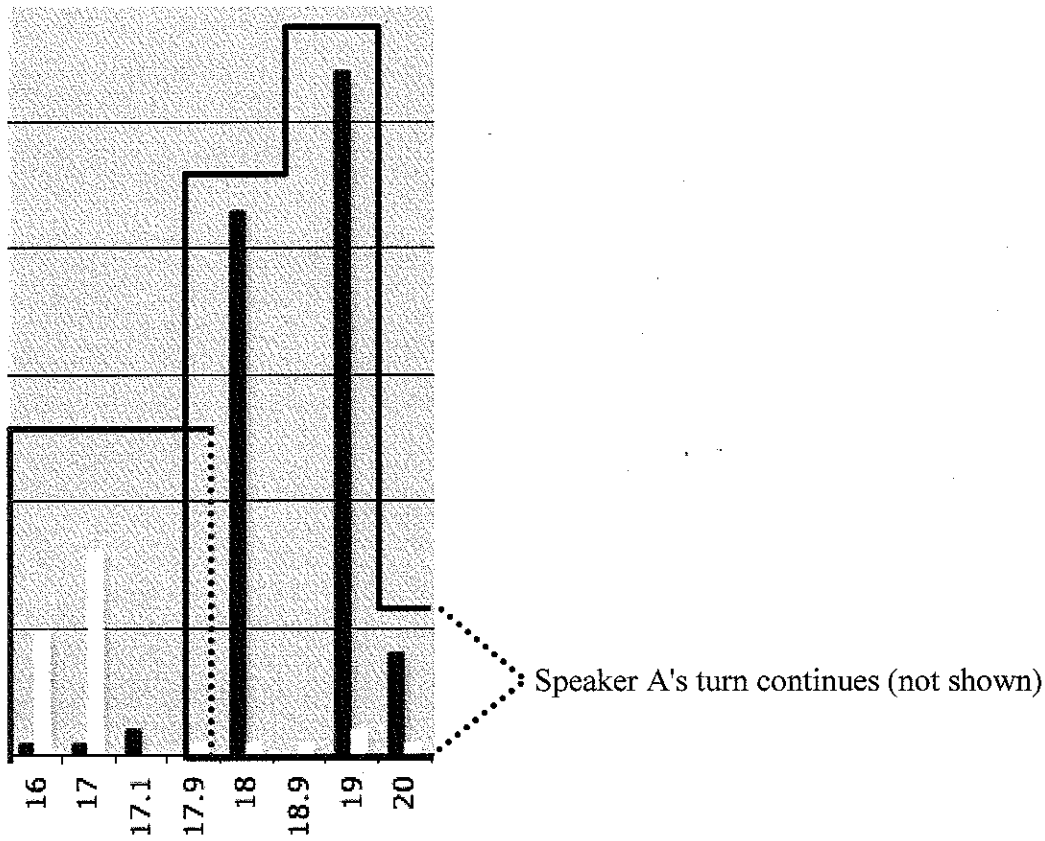
is outlined below in red, with a dotted line denoting the point of interruption. The visible portion of Speaker B's renewed beat is outlined in indigo.

Figure 14: An interrupted beat (Speaker B interrupts Speaker A)



Shown on the next page in figure 15, the same thing happens to speaker B in chart 5 on his turns 16-17. The two utterances, increasing in length, resemble a prebeat. However, the beatcenter never appears because speaker A assumes dominance with a very long utterance (40+ tokens) on his turn 18. In this case, again, Speaker B accepted the fact that Speaker A would assume dominance and decreased his utterances' lengths to one or two tokens in the following turns. In figure 15, Speaker B's interrupted beat is outlined in red and speaker A's beat in indigo. Dotted lines indicate interruption.

Figure 15: Another interrupted beat; smooth cession of dominance

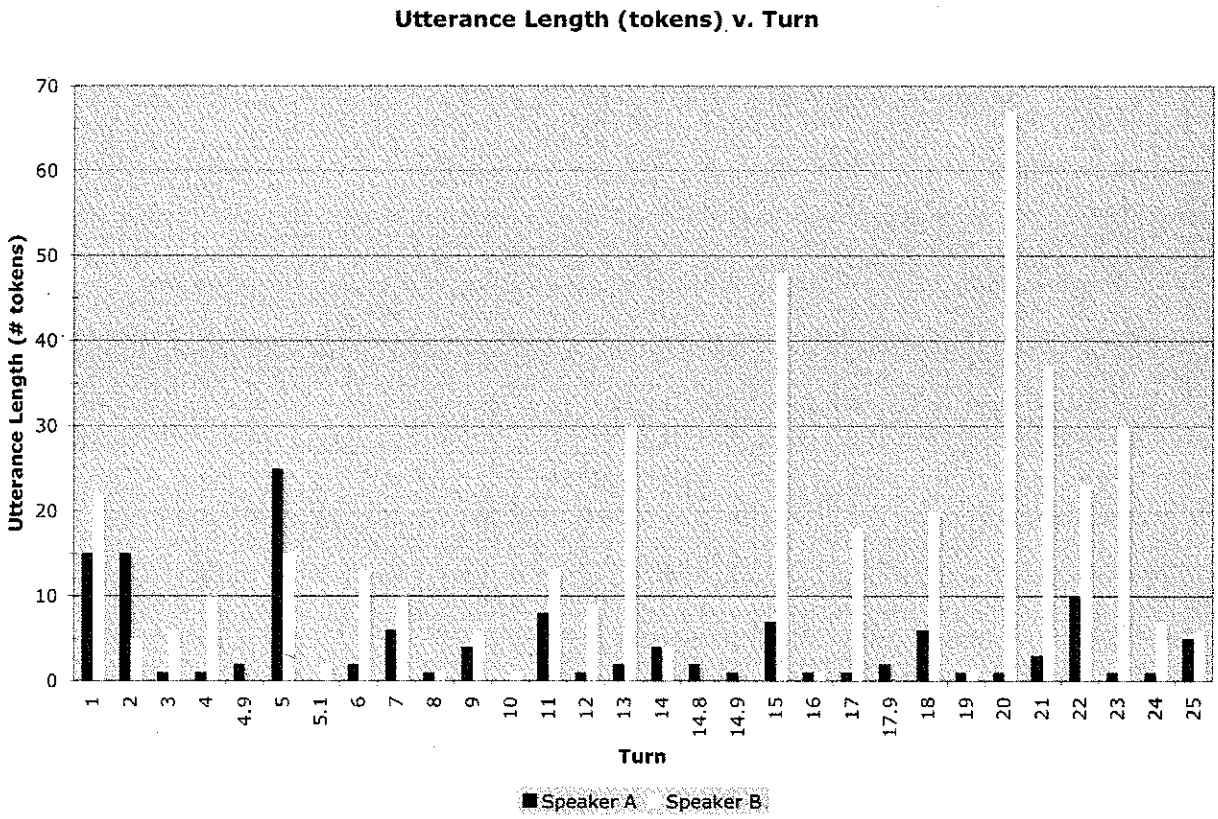


6.124 The NESTED BEAT vs. an interrupted beat

However, it is not always the case that an interrupted speaker would simply get out of the way of the interrupter. Sometimes, a non-dominant speaker would interrupt the dominant speaker's beat with a beat of his or her (the non-dominant speaker's) own. The interrupted speaker, however, instead of truncating his or her beat as Speaker B did in the previous example, would sometimes continue on unfazed. The resulting structure consisted of two beats – one by each speaker – one of which was partially or wholly nested inside the other. Session 8's dialogue provides an excellent example of a regular interrupted beat and an unusual interrupted beat that is very close to the description of nested beat, showing how closely related the two structures are.

The dialogue shown below as chart 6, representing minutes 5-8 of session 8, occurred between two female seniors who had been friends since freshman year. Perhaps because these women know each other so well, and so are less reliant on common courtesy to demonstrate mutual respect than strangers would be, their turn-taking was sometimes less than orderly. Particularly the simultaneous responses I discussed above in section 4.233 demonstrate their unusual willingness to talk at the same time as one another. Well before the point at which the simultaneous responses occurred, though, these two subjects exhibited an interesting dominance struggle, described below. Chart 6 appears below.

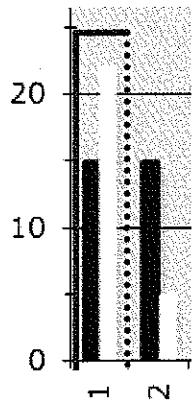
Chart 6: Session 8 Dialogue – Utterance Length (tokens) v. Turn



Consider the first 8 turns for both subjects in the dialogue above. Speaker B had the turn initiative, starting off the dialogue with an extended utterance that presumably was a beatcenter, but was phrased so as imply a question ("I wonder..."). Speaker A's response on turn 1, rather than being minimal, answered the implied question with another, direct question, which

prompted speaker B to answer that question in brief on turn 2, and speaker A to continue explaining on her turn 2. In effect, then, speaker A's response behavior interrupted Speaker B's beat. In this instance, Speaker B did indeed yields dominance to Speaker A on turn 2. (We shall see that she does not do so unequivocally in turns 3-8.) In figure 16, given below, Speaker B's interrupted beat is outlined in red; the blue bars indicate Speaker A's extended response – not really a beat because it does not have the structure defined above. In addition, I have included the transcription of these two turns for comparison.

Figure 16: Extended response that interrupts a beat, with transcription



B: I wonder if this is different with people who don't know each other like if they say different things to each other

A: probably…

A: …dunno [= 1 token]…
B: [giggles] [≠ a token]

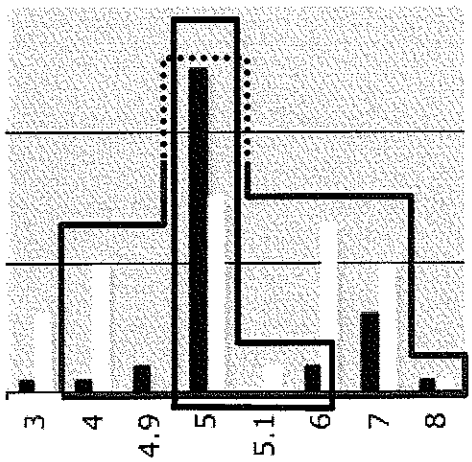
A: …we're not 'sposed to necessarily come up with same answer anyway are we?

B: No I don't think so

A: [quoting from the handout] You may talk to the other person in the room if you want or not **(8, 5:04, 1-2)**

This is an example of one speaker (Speaker B) unequivocally ceding dominance to the other speaker after being interrupted. However, in the following figure 17, when Speaker B next begins a new beat, Speaker A interrupts again with her own beat, and this time Speaker B does not retreat nearly as much. In fact, although Speaker B cuts her beatcenter (turn 5) short because Speaker A began talking, her next few utterances appear to be a continuation of the interrupted beat. In figure 17, below, I have outlined the structure of Speaker B's interrupted beat in red, with dotted lines indicating the presence of the beatcenter that I believe was preempted by Speaker A's beat, outlined in indigo:

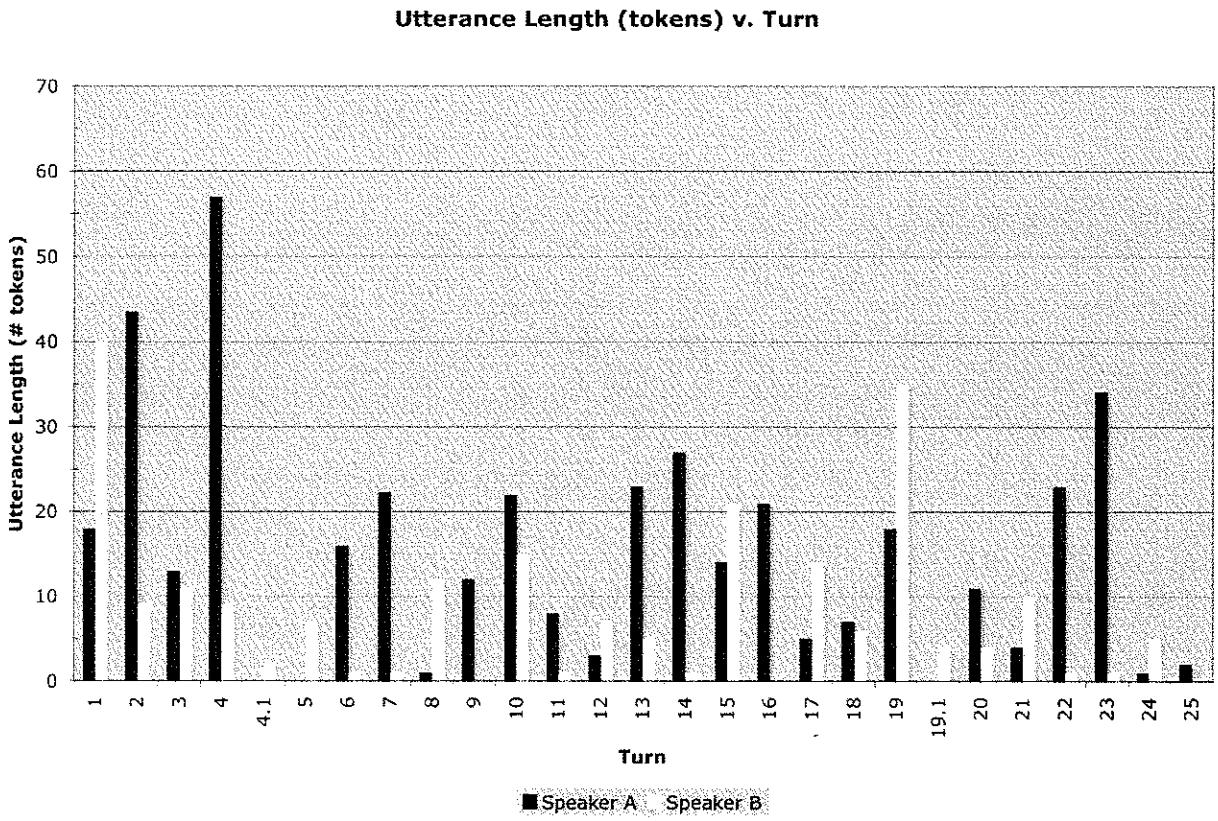
Figure 17: An interrupted beat that continues – sort of a nested beat



Since this beat appears to continue through the interruption of Speaker A's beat, it looks like Speaker A's beat is nested within Speaker B's more expanded beat. However, since Speaker B truncated her beatcenter, her beat also qualifies as an interrupted beat as well.

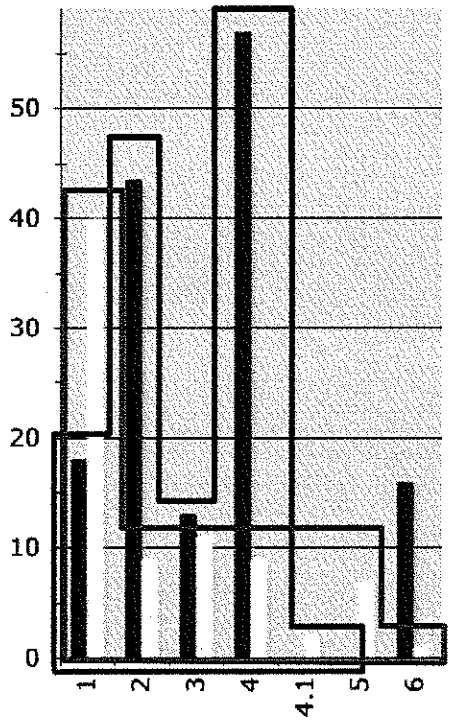
More complicated nested beats that involved no obvious interruptions occurred in the session 11 dialogue. I will discuss this dialogue in more detail in section 6.2 on MACROTURN-TAKING, below, but let me introduce it now. The session 11 dialogue's subjects were first years, one male and one female, who had not met each other before the experiment. Each of them was enthusiastic and very actively participated in the discussion, particularly speaker A. Perhaps because of the intense nature of their discussion, or perhaps because of Speaker A's loquaciousness (his utterances very rarely dipped under 5 tokens in length), their beats are an absolute mess, occurring on top of one another. Chart 7, on the next page, shows the bar graph of this dialogue, which represents about 3 minutes and 40 seconds of the middle of session 11.

Chart 7: Session 11 Dialogue – Utterance Length (tokens) v. Turn



The first 6 turns of this dialogue exhibit a drawn out pair of nested beats, shown in figure 18, below. Speaker A initiated the dialogue on turn 1 with a direct question (perhaps a feature of nested beats). Speaker B responded to the question on her turn one, beginning a beat, but Speaker A continued his beat to a beatcenter of more than 40 words. As Speaker A continued and then renewed his beat on turns 3 and 4, Speaker B's responses through turn 5 – a bit longer than the average responses by a non-dominant speaker – seemed more like a postbeat to her utterance on turn 1 than anything else. Speaker B's turn 6, at only 1 token, must be either a response or a beatender, however. Speaker A concluded his beat by not responding to speaker B's turn 5, after which there was a pause of more than four seconds; I have treated this as a silent beatender. In figure 18 below (a detail of chart 7), Speaker B's beat is outlined in red, speaker A's beat is outlined in indigo, and the two beats are right on top of one another.

Figure 18: NESTED BEATS



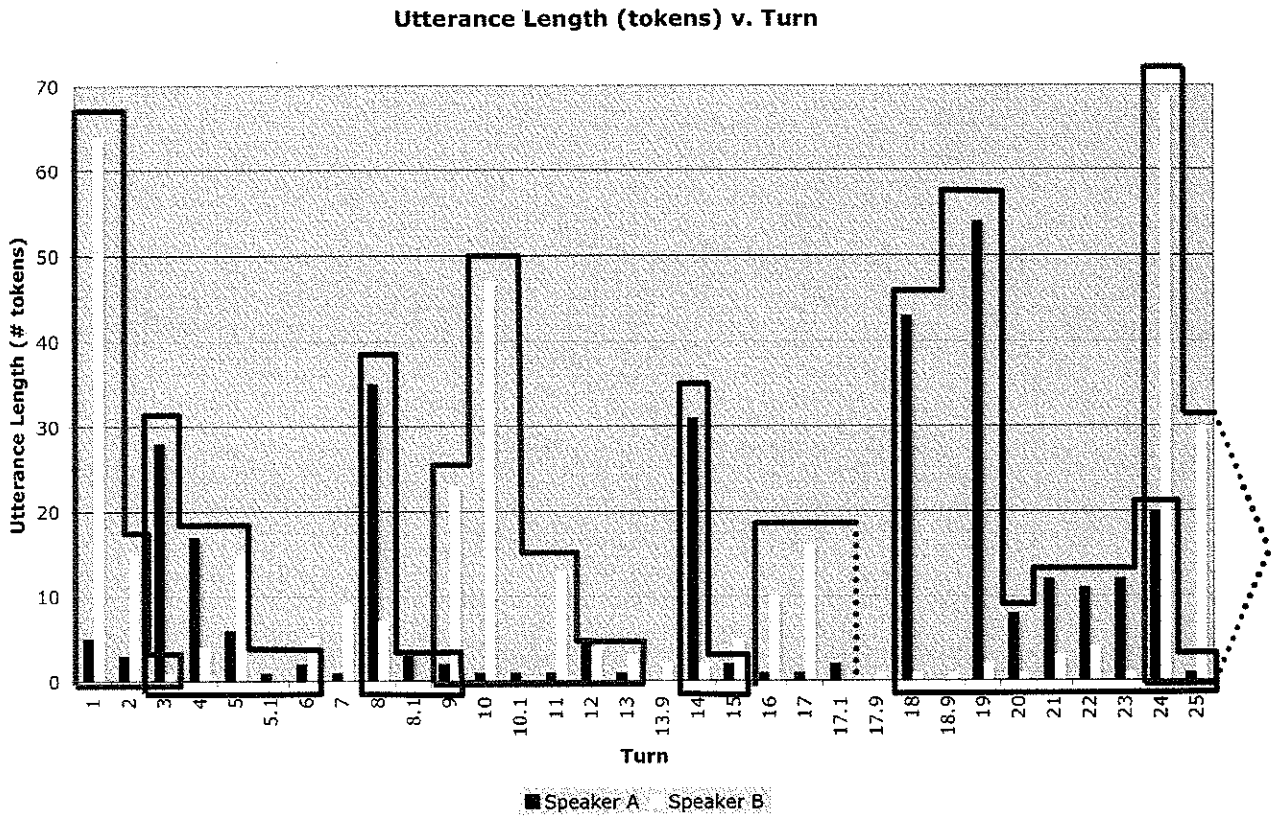
Again, without some means of abstracting from the utterance length data, it would be extremely difficult to ascribe any pattern to the six turns excerpted above. The beat analysis allows one to classify these turns as a special case derived from two more basic elements (beats), which in turn are derived from more basic elements (prebeats, beatcenters, etc.), instead of merely throwing in the towel and declaring the utterance lengths of each speaker to be random.

6.2 MACROTURN-TAKING

Far from being random, the lengths of the speakers' utterances, which I determined by relying on the notion of turns, can be organized into a MACROTURN-TAKING structure. Recall that I defined a single turn by one speaker as the period beginning when the speaker starts speaking and ending when the speakers stops speaking, provided it is just before, just after, or exactly at the time when the other speaker begins speaking. If we define a MACROTURN as beginning at the point at which a speaker assumes dominance in the conversation and ending when the speaker ceases to be dominant – which usually means one beat, renewed beat, etc. – then the utterance length data organizes itself into a relatively clear-cut turn-taking paradigm. Let us revisit chart 5,

illustrating the session 3 dialogue, since it was the cleanest example in my data. The chart is reproduced below as figure 19, with all of Speaker A's beats in indigo and all of Speaker B's beats in red. Dotted lines indicate interruptions, as before.

Figure 19: Orderly macroturn-taking in the session 3 dialogue



Taking for granted my analysis of speaker B's turns 16 and 17 as an interrupted beat (see section 6.123, above), the speakers in the session 3 dialogue have perfectly orderly, alternating beats – and hence, perfectly orderly, alternating macroturns – throughout almost the whole dialogue. One speaker takes a dominant beat, then the other speaker assumes dominance for a beat, then the first speaker reasserts dominance, etc. These macroturns are contiguous with only a slight gap or slight overlap, just like regular turns – remember that decimal-point bars represent buried turns that are concurrent with the preceding or following turn, and so the fact that they appear to be gaps is merely an artifact of my notational system.

The interrupted beat (Speaker B's turns 16-17) should not be considered a breakdown of the macroturn structure. When Speaker A interrupts with his renewed beat beginning at turn 18, Speaker B initiates the proper, well-behaved response to overlapping macroturns: he ends his macroturn by ceding dominance. The analogous process in regular speech turn-taking is quite common: when one speaker interrupts the other, the latter speaker initiates the proper, well-behaved response to overlapping turns, and ends his turn by ceasing to talk (see principle 9, section 1.5, above).

There is one feature of figure 19 that does not fit with the alternating turn model as I have described it up to this point. Speaker A has two contiguous beats during his turns 3-9. The key to assimilating this feature of the data into the analysis is the fact that turn 7 does not belong to any beat. Turn seven constitutes a LAPSE in the alternating beat-structure. In the case of speech-level turn structure, when a lapse in the conversation arises, the oft-initiated repair mechanism is for the speaker who just stopped talking – i.e. the speaker whose turn is in progress – to resume speaking, thereby continuing his or her turn. Comparably, for the beat-level macroturn structure, when a lapse in the beat-alternation arises, the repair mechanism is for the speaker whose beat just concluded – the speaker whose macroturn it is – to initiate a new beat, thereby continuing his macroturn. Thus, although a speaker's macroturn usually consists of a single beat, it can include more than one beat if the other speaker does not begin his or her own macroturn on the heels of the former's.

Speaker A had the turn initiative in the first portion of the dialogue shown in figure 19. A possible explanation of how this macroturn extension theory applies is as follows: Speaker A is winding down the beat he began on turn 3, uttering a 5-token sentence on his turn 5 as part of the postbeat. To this, Speaker B offers a 15-token response on his turn 5, which Speaker A understandably assumes is a prebeat because of it length. He therefore concludes his beat in the proper manner with a beatender on turn 6, believing his macroturn to be over. However, Speaker

B did not intend to begin his own macroturn – his extended response was misinterpreted by Speaker A – and so his turn 6 is not a beatcenter but rather another short response like those he uttered on his turns 3-4 during Speaker A's beat. Now there is a problem: Speaker A believes his macroturn to be over, and speaker B believes his macroturn has not yet begun. Turn 7 constitutes the lapse in which both speakers are in limbo, believing the other speaker should be dominant. Speaker A, the speaker whose macroturn it was before the lapse, repairs the situation by initiating a new beat and extending his macroturn. After this hiccup, the macroturn structure proceeds smoothly.

I am not arguing that speakers consciously consider macroturn order when structuring their conversations any more than I would argue that speakers recognize the features of their turn-taking mechanisms during conversation. I merely point out that these dialogues seem to organize themselves into a macroturn-structure that appears to have many of the same features as the utterance-level turn-structure, abstracted up to the beat level. If utterance-level turn taking is a socially acquired skill, my data suggests that the same skill is applied recursively to organize the conversation on multiple levels.

6.21 The spectrum of macroturn-taking

As the cleanest specimen from my data set, the session 3 dialogue fell at one end of the orderly-to-chaotic spectrum of macroturn-taking. Because of its alternating-beat structure, it is GLOBALLY BALANCED with respect to macroturns, meaning that each speaker had roughly the same number of macroturns. The dialogues from session 2 and 8 were slightly less clear-cut in that they each include one or more lapses and devolved into a one-sided macroturn instead of an alternating-beat structure. In other words, they were both GLOBALLY UNBALANCED with respect to macroturns. I have given the graphs for each of these sessions below, with the beats outlined as they have been above, as figures 20 and 21, respectively.

Recall that figure 20 represents the very beginning of the conversation between the two subjects, so the initial lapse (turns 1-3) can be explained as a period in which the speakers fished around to establish who would take the first macroturn. The second lapse (turns 9-12) presumably occurred because, when Speaker B ended her beat on turn 8, Speaker A was not willing or not ready to begin her own beat. Note that Speaker A, after the lapse persisted for a few turns, signaled that she did not intend to be the dominant speaker with a very short utterance (turn 12), but Speaker B's turn 12 was not long enough to constitute a pre-beat. Both speakers attempted to repair the situation on turn 13, Speaker A with a prebeat and Speaker B with a much longer prebeat (an utterance that could have been a beatcenter in other circumstances), and Speaker A resumed her unequivocal non-dominant status.

Figure 20: Session 2 dialogue – globally unbalanced macroturns; 2 lapses

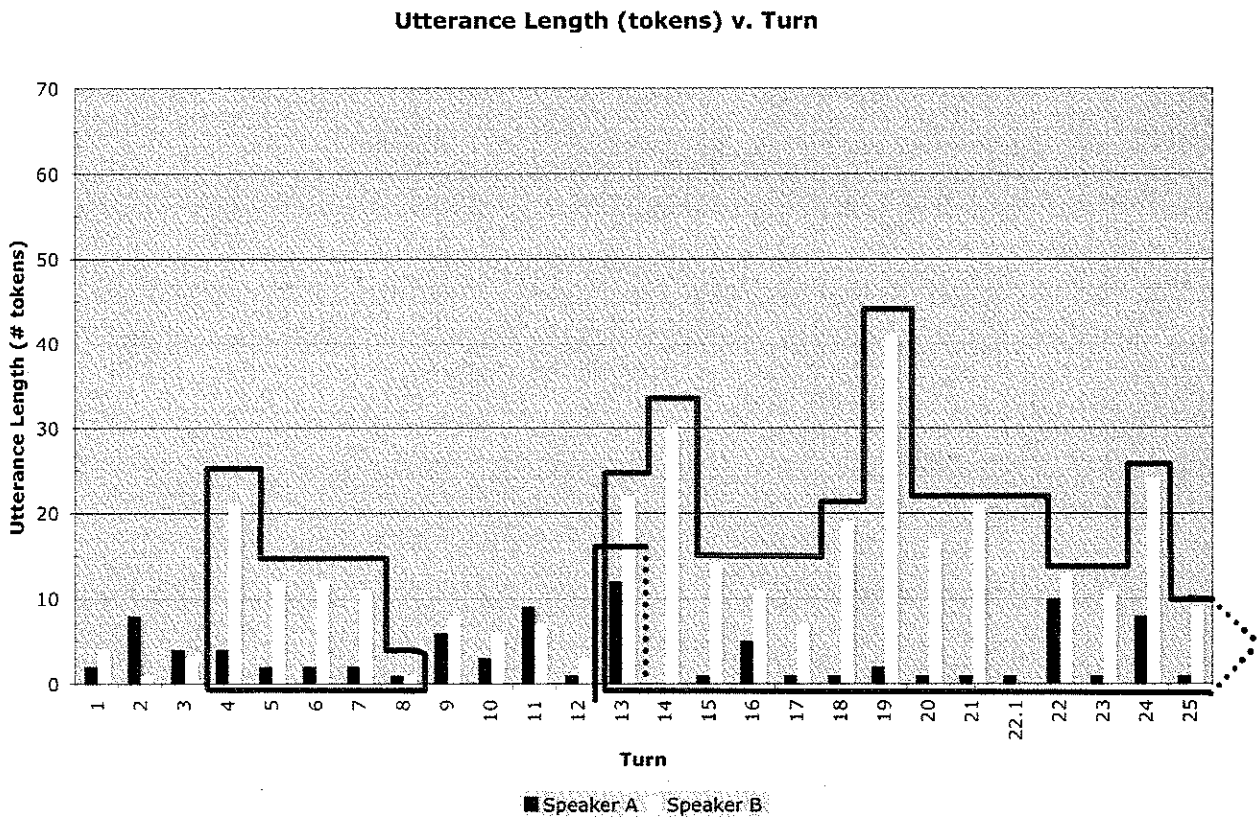
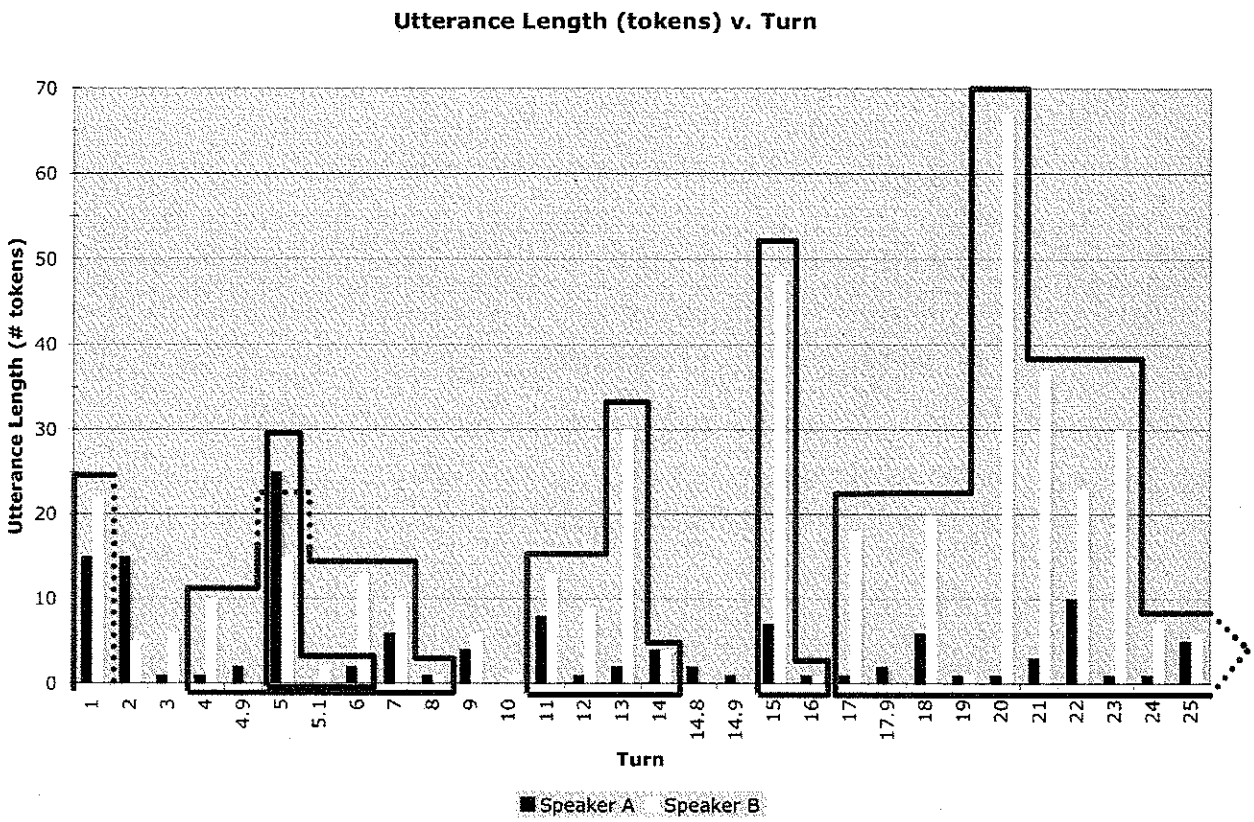


Figure 21: Session 8 dialogue – globally unbalanced macroturns, 1 nested beat



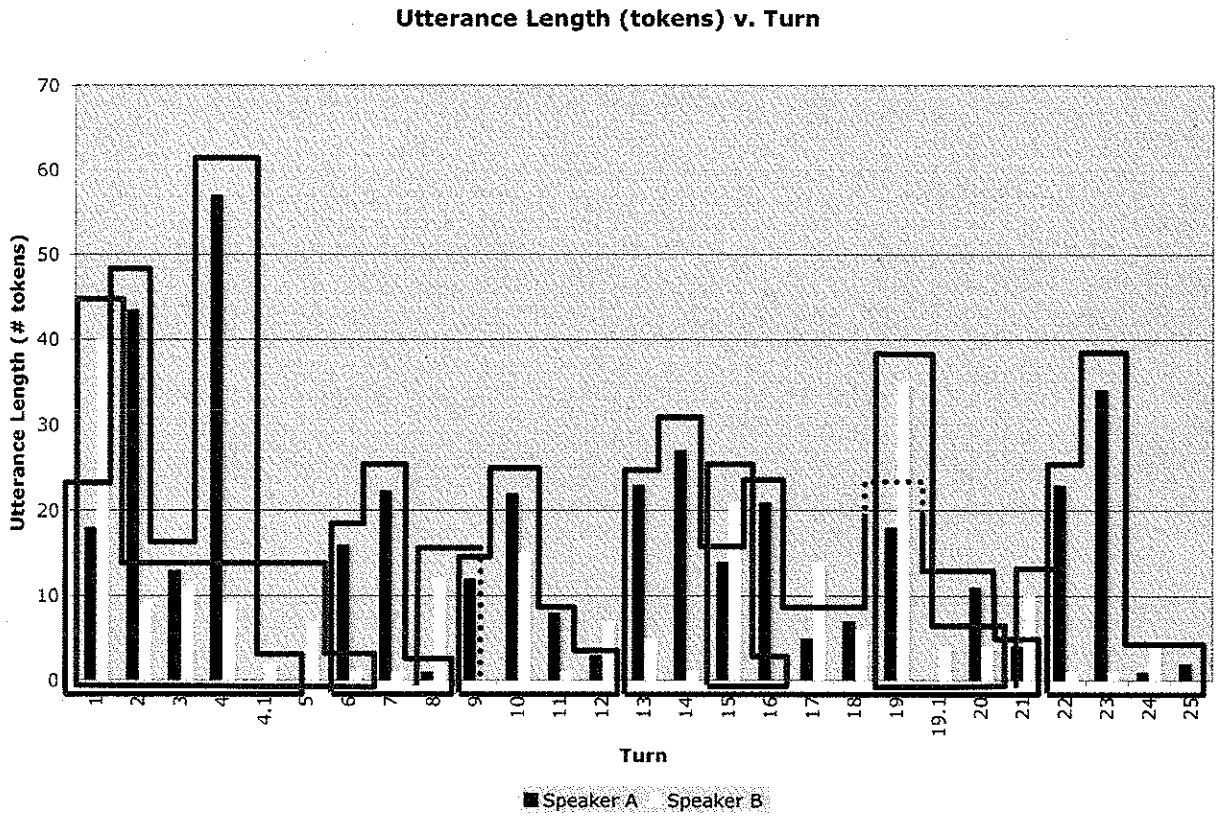
In Figure 21, despite the fact that speaker B monopolizes the conversation, the macroturn-taking in the dialogue follows the rules I have set out above with two modifications; one slight and one more substantial. The lapse in turn 2-3 can be explained as the period of confusion following the interrupted beat on turn 1. Speaker B repairs the lapse by continuing her macroturn with a new beat on turn 4. Speaker B repairs another lapse, turns 9-10, by again continuing her macroturn on turns 11-14. However, between the remaining beats she takes in the dialogue, there are no lapses at all. Speaker B simply continues her macroturn with new beats without giving Speaker A the option to assume dominance. To account for this, the rules for macroturns must allow a dominant speaker to extend his or her macroturn voluntarily, taking beat after beat. Of course, this is a natural addition to the rules for macroturns, since it corresponds exactly to the idea that utterance-level turns may be extended in length indefinitely by simply continuing to speak clause

after clause, as noted in example 6 in section 4.12 above. So, this is only a slight modification of the rules.

Turns 4-8 present a larger problem. I have already discussed the nested beat on Speaker A's turns 5-6 (section 6.124); it seems to me that I can only analyze this as a macroturn nested inside Speaker B's macroturn. As for how allowing overlapping macroturns reflects on macroturn structure in general, I must point to the fact that speakers do occasionally have concurrent turns at utterance-level, so it seems plausible that the same thing might happen at the beat-level. The problem will not go away so easily, however. As the next example shall make clear, macroturn-taking seems to tolerate more complexity than turn-taking.

Figure 22 below is a reproduction of chart 7, the session 11 dialogue, with my best attempt to outline the beats using the conventions I have developed above. As involved as my analysis is, admittedly it can only explain this dialogue with extreme difficulty.

Figure 22: Session 11 dialogue – globally unbalanced macroturns, chaotic



The major problem seems to be that Speaker A just won't shut up. That is to say, Speaker A continuously exercises his option to continue his macroturn by initiating new beats, regardless of whether Speaker B attempts to initiate a beat in response or not. Figure 22 shows that, if we analyze Speaker A's turn 5 as a 0-token beatender (as in section 6.124), the course of the dialogue for Speaker A is as follows: nested renewed beat, (turns 1-5), beat (6-8), beat (9-12), doubly-renewed-but-interrupted-on-the-second-renewal beat (13-21), beat (22-25). For speaker B, sneaking in beats despite the unbroken stream of beats from her conversation partner, the course of the dialogue is: nested beat (turns 1-6), interrupted beat (8), nested beat with a 0-token beatender (15-16), nested beat (19-20), interrupted beat (21)[19].

It does not take a wary reader to see that turns 13-21 represent a less-than-elegant analysis. Look at the hyphenated reference needed just to describe Speaker A's speech for that duration! On the other hand, if we were to look at the session 11 dialogue only through the lens of utterance length, if might be difficult to describe why exactly this dialogue is so much more erratic than the other dialogues cited in the paper. The speakers do not minimize their turn length during the other speaker's periods of long utterances nearly to the extent that the speakers in the other dialogues did. Speaker A in this dialogue has more utterances of length greater than or equal to 20 tokens than any of the speakers in the other dialogues, and he has fewer utterances under 5 tokens than any of them, as well. Speaker B has a number of utterances in the 10-20 token range that may or may not be contiguous with utterances longer than 20 tokens, which was not the case in the other dialogues. It would be extremely difficult to find a pattern in the utterance lengths alone, without abstraction, that accounted for both this dialogue and all the others. It is therefore a boon to my analysis that, even though it must stretch its limits, it not only

[19] Turn 17 for Speaker B is not an interrupted beat because Speaker A did not substantially increase his utterance length to preempt Speaker B's assumption of dominance. Speaker B ceded dominance without a strong signal from Speaker A, and so I have considered turn 17 an extended response rather than an interrupted beat.

accounts for the data in the session 11 dialogue but also leads to the observation that differentiates this dialogue from all the others: a single unbroken stream of beats from one speaker, and its effect on the other Speaker's macroturns.

A final note on figure 22 relates to the following question: if Speaker A in this dialogue continually exercises his option to extend his macroturn regardless of what Speaker B does, does that mean that his entire macroturn is one sort of humongous, unbroken macrobeat? It would explain why Speaker A is not concerned with ceding dominance, since he would plan (unconsciously or consciously) to conclude his entire macrobeat before doing so, just as he finishes each individual beat before ceding dominance. Of course, generalizing from a single instance of such a phenomenon is unsound, but it is promising that, if we examine the beatcenters of each beat, we see that the first one is about ¾ of the length of the second, longest beatcenter, and all of the beatcenters after the second are around half as long as or a little shorter than the second, longest beatcenter. This fits the {premacrobeat, macrobeatcenter, postmacrobeat...} form discussed above in section 6.122. The macrobeatender must have occurred after or on Speaker A's turn 25, since we do not see it's conclusion.

6.3 Local balance as a non-beat-based special effect

Although the session 11 dialogue was difficult to analyze with beat structure, a portion of the session 6 dialogue was clearly impossible to analyze under a beat-based system at all. The two subjects were first years, one male and one female, who knew each other fairly well as hallmates. Their conversation included a number of shared references, specifically relating to the Chinese language, which they were both studying here at Swarthmore. Chart 8, below, shows utterance length in tokens versus turns for the session 6 dialogue. Figure 23, directly below it, outlines the speakers' beats where possible, according to the conventions above.

Chart 8: Session 6 Dialogue – Utterance Length (tokens) v. Turn

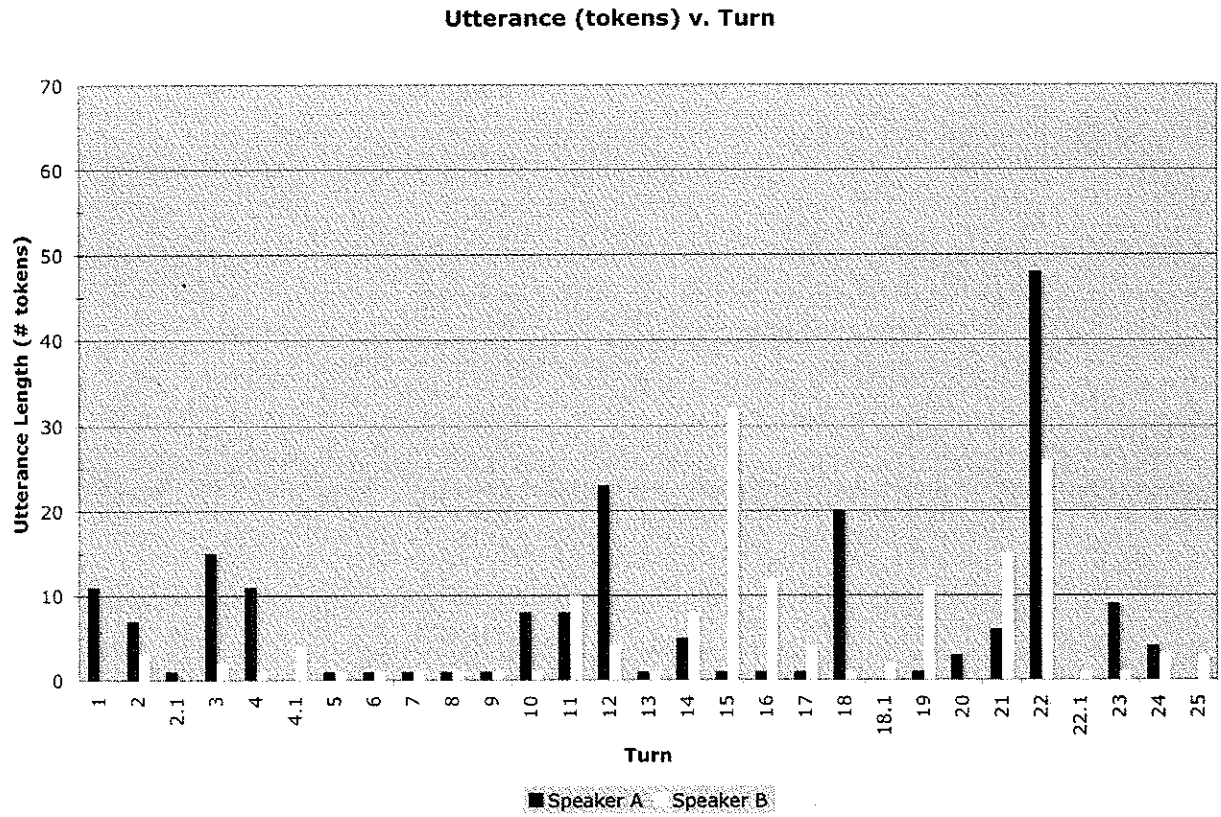
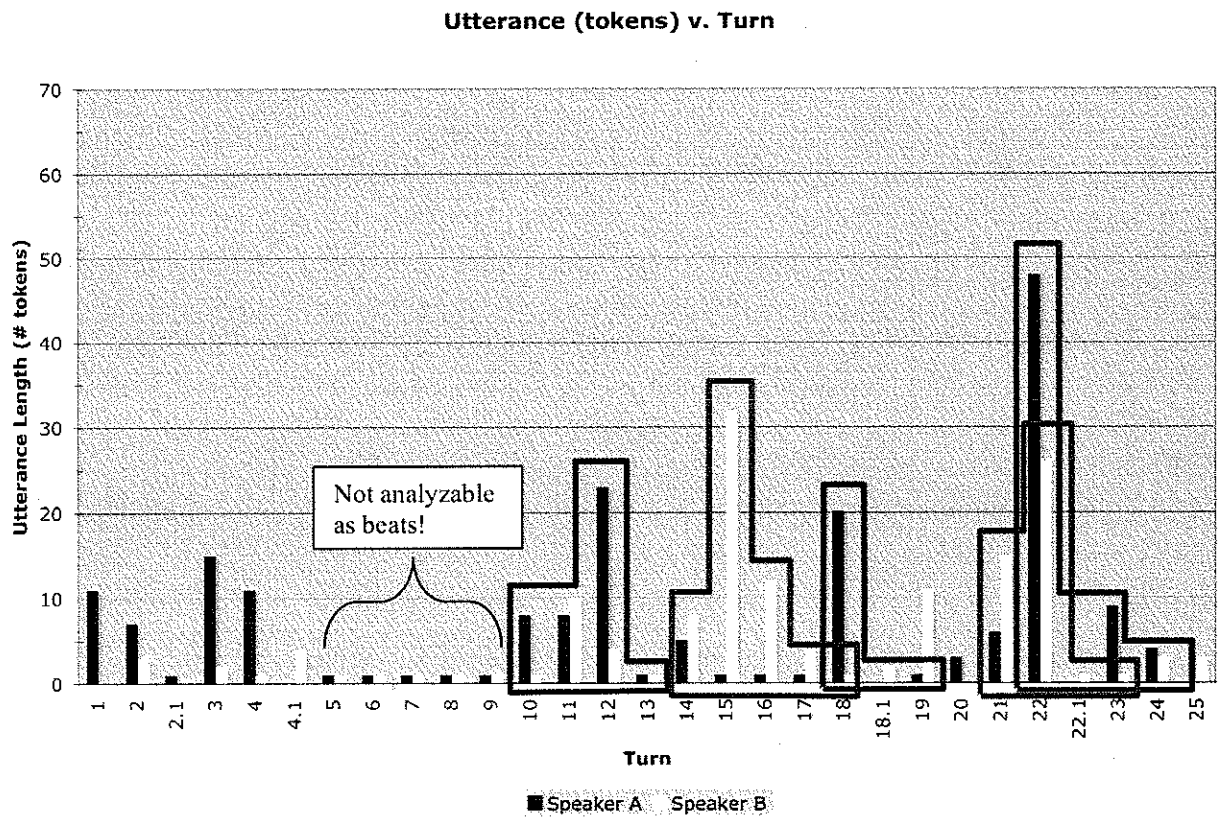


Figure 23: Beat analysis possible after turn 9, but not before



From turn 10 on, the dialogue is analyzable in terms of the beat structures I have described, and in fact the macroturn-taking is quite orderly. The only hitch comes when speaker B responds with an 11-token utterance to a beatcenter by speaker A, and speaker A mistakes it for a prebeat to a turn by Speaker B. There is a lapse on turn 20 as speaker A responds with a short utterance and Speaker B has nothing else to say. Speaker B recognizes the problem quickly, however, and initiates a beat on turn 21, preserving the alternating-beat structure[20].

As I have labeled in figure 23, however, turns 5-9 involve the speakers firing one-word utterances back and forth for five turns, one of the very features I objected to in Graham's playwriting exercise. Before I eat crow, let me give the transcription of the relevant data, as example 17:

Ex. 17: Back and forth single-word utterances

B: liù

A: lee-oh

B: qī

A: qī

B: bā

A: bā

B: jiǔ

A: jiǔ

B: shí
A: shí **(6, 2:45, 5-9)**

[20] The notation on turns 21-23 obscures the fact that Speaker A does not interrupt speaker A's beat here. Speaker B has the turn initiative on turns 21-23, and so utters the beatcenter on turn 22 before Speaker A begins a new beat on turn 22. Since Speaker B's beatender – a response to Speaker A's new beat – is the only portion of the beat that overlaps with Speaker A's new beat, Speaker B's beat is neither interrupted nor nested according to my conventions.

Speaker B was helping Speaker A review her Chinese numerals. The speakers uttered the numbers from one to five in unison, but then Speaker A mispronounced the Chinese word for the numeral *6*. From that point until they got to *shí* (10), Speaker B recited the word first, and then Speaker A repeated after him.

Repeating after someone as they count to ten does not qualify as normal, spontaneous speech. It is a highly special speech event, completely formulaic and non-spontaneous, that does not feature very often in conversation outside of the specific context of a language class or language teaching/learning. It is therefore not surprising that it should not follow the same rules as normal speech, and so I did not amend my analysis in order to capture it. Instead, I took it as evidence that locally balanced speech occurs during formulaic exchanges, as I originally hypothesized.

As for the macroturn lapse in turns 1-4, speaker A began the dialogue with a direct question, which I have suggested may somehow lead to turn lapses. Furthermore, Speaker A's turns 3 and 4 certainly resemble a prebeat. However, the prebeat was again a direct question, asking how to say *nine* in Chinese, which lead to the balanced exchange given in example 17. I believe the formulaic nature of turns 5-9 preempted the beat Speaker A seemed to be initiating.

6.4 Mean length of utterance (MLU)

Linguists who have studied utterance length have been primarily concerned with speakers' mean length of utterance (MLU). In light of my beat-driven analysis of dialogue, over what domain should we take the mean in order to obtain a meaningful (no pun intended) result?

I have discussed the balance or imbalance of dialogue between two speakers in two domains, the local and the global. It seems natural that both domains should be considered separately for the purposes of MLU.

6.41 Global MLU

Global MLU is the simpler and more objective way of looking at the data. I defined *utterance* in terms of turns for the purposes of my analysis. So, for each dialogue, I added up the total amount of tokens each speaker uttered and divided by the number of turns in which they uttered those tokens. Turns labeled with decimal points were counted as separate turns for the interrupting speaker (the speaker who spoke the tokens notated with the decimal turn number), but not counted as turns for the interrupted speaker. Non-decimal turns in which one speaker uttered no tokens were still counted as turns; the passage of a turn in silence can provide some information to the other speaker (e.g. my previous analysis, in which silence could serve as a 0-token beatender), and so I thought it important to count turns intentionally left unused.

Table 1, below, shows the Global MLUs for each subjects of each of the dialogues except for those of session 9, which will be treated separately in section 6.4.

Table 1: Global Mean Length of Utterance (tokens/turn) by Speaker and Session

	Session 2	Session 3	Session 6	Session 8	Session 11
Speaker A	3.42	10.69	7.27	4.45	16.24
Speaker B	13.14	12.79	5.36	15.58	8.23

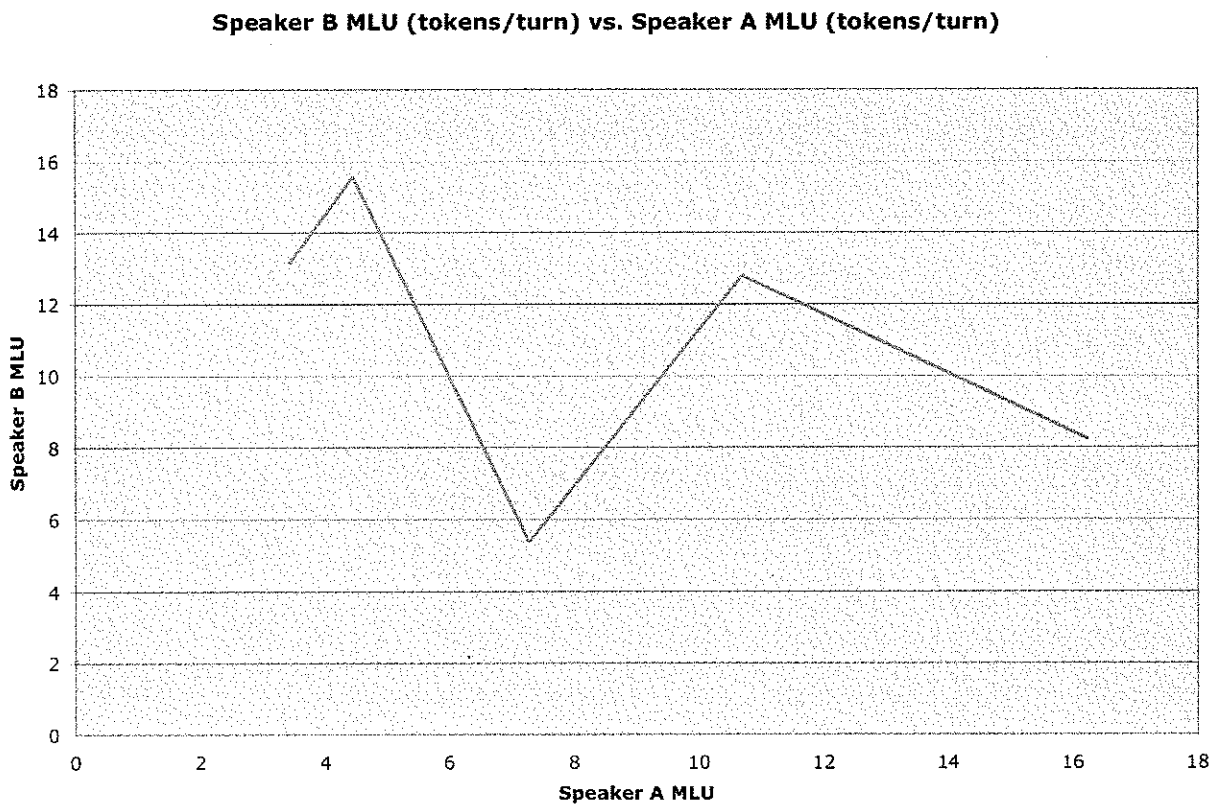
There is certainly a lot of variation among the speakers globally, with MLUs from 3.42 tokens/turn to 16.24 tokens/turn. The mean global MLU across all sessions was 9.72 tokens/turn.

The MLU numbers line up with the balanced/unbalanced labels with respect to macroturns: the dialogues from Sessions 2, 8, and 11, in which one speaker's beat or beats took up far more of the 25 turns than the other speaker's, were globally unbalanced with respect to MLU as well. Sessions 3 and 6, which displayed relatively orderly, alternating-beat macroturn-taking (where macroturn-taking took place), were globally balanced with respect to MLU, or at least more balanced than the other sections. The correlation makes sense: macroturns were

defined by beats, which were in turn defined by long utterances. If one speaker utilized more long utterances, he or she tended to monopolize the macroturns in my analysis of the dialogue. The significantly greater MLU of one speaker (longer average utterance) therefore caused an imbalance in macroturns in favor of that speaker.

One might expect a negative correlation between MLU of one speaker and the MLU of the other, since beat-structure involves on speaker intentionally limiting his or her MLU while the other speaker is dominant. There did not seem to be any correlation between the MLU of one speaker in a dialogue and the MLU of the other speaker. Figure 24, below, is a graph of Speaker B's MLU vs. Speaker A's MLU for each dialogue. The fact that it zigzags so erratically instead of progressing generally upward or generally downward is evidence that there is no simple negative correlation between one speaker's global MLU and the other's.

Figure 24: Speaker B MLU vs. Speaker A MLU – not a simple negative correlation



6.42 Local MLU

When considering MLU in the local domain, it is again important to consider what local domain(s) are important enough that MLU data from them will be meaningful? For each speaker, I have already implicitly divided each dialogue into two types of local domains with my beat analysis: times at which the speaker's utterances are part of a beat (IN-BEAT utterances), and times in which the speaker's utterances are not part of a beat (OUT-OF-BEAT utterances). In the latter case, my analysis treats the speaker as responding to the other speaker's beat. I decided to adopt these two local domains, in-beat and out-of-beat, and the ones that will yield meaningful MLU data: from the former, we see the average number of tokens the speaker uses when assuming dominance to communicate something, and from the latter, we see the number of tokens the speaker uses when listening and responding to the other speaker's communication.

6.421 IN-BEAT MLU

I calculated IN-BEAT MLU using the beat-structure analyses I provided above. For each dialogue, I analyzed a portion of the utterances of each speaker as being part of a beat, interrupted beat, etc.; I shall call these IN-BEAT UTTERANCES. I then totaled the number of tokens spoken by a speaker during his or her in-beat utterances and divided by the number of turns in which that speaker uttered those tokens. I applied the same conventions for decimal-numbered turns and silent turns as in the previous section. Table 2, below, shows the in-beat MLU of each speaker in each dialogue except for the session 9 dialogues.

Table 2: In-Beat Mean Length of Utterance (tokens/turn) by Speaker and Session

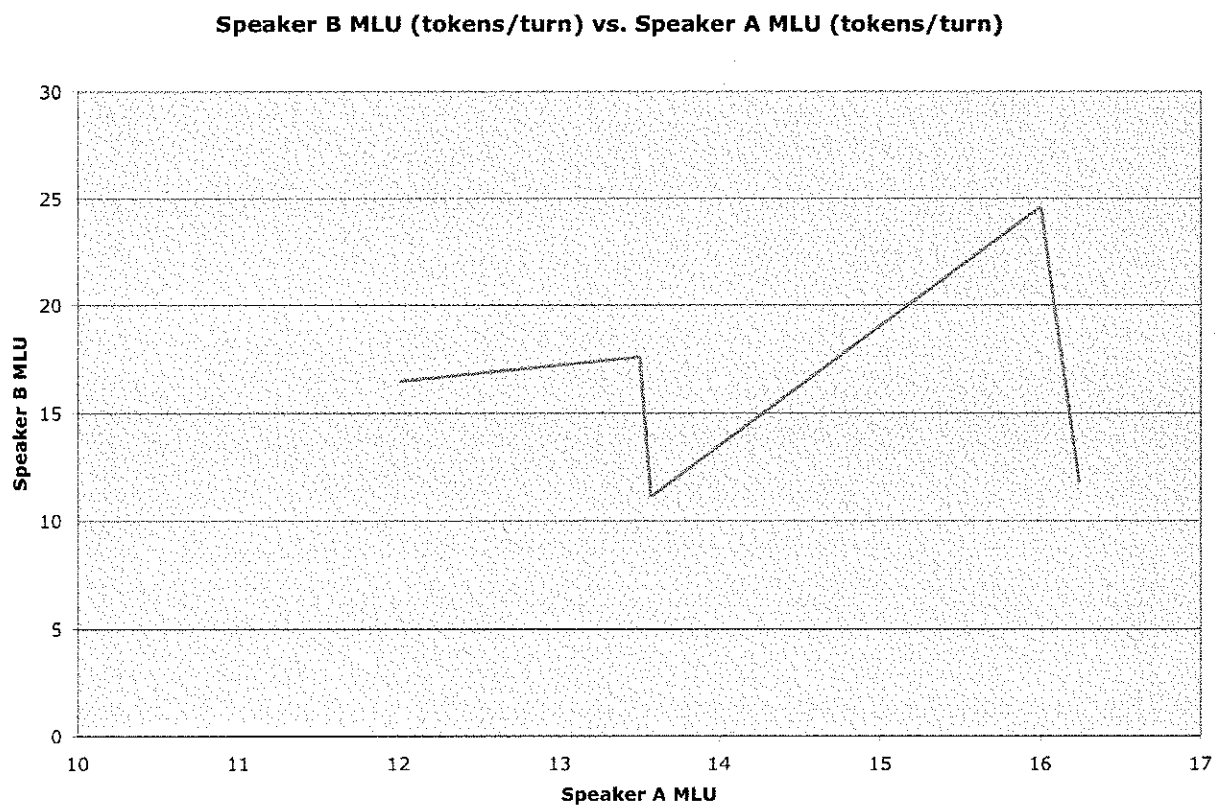
	Session 2	Session 3	Session 6	Session 8	Session 11
Speaker A	12.00	16.00	13.57	13.50	16.24
Speaker B	16.47	24.60	11.11	17.59	11.77

Not surprisingly, these figures are much greater than the global MLU figures. A beat forms around a beatcenter that exceeds 20 tokens in length, and often includes pre- or postbeats that are

longer than responses. Because of this, the in-beat MLU tends to be high. There is still considerable variation among speakers, with in-beat MLUs ranging from 11.11 tokens/in-beat-turn to 24.6 tokens/in-beat-turn. The mean MLU across all sessions was 15.29 tokens/in-beat-turn.

Since the in-beat utterances for the two speakers in a dialogue were, ideally, on different turns (because the speakers exhibited alternating-beat behavior), there should be no correlation between Speaker A's in-beat MLU and Speaker B's MLU. Of course, we saw in some of the sessions that speaker's beats could overlap, and so this ideal assumption could be flawed. Figure 25 below shows Speaker A's in-beat MLU plotted against Speaker B's in-beat MLU. Again, the zigzagging lines shows that there is not much of a correlation between the two figures:

Figure 25: Speaker B in-beat MLU vs. Speaker A in-beat MLU – no correlation



6.422 Out-of-beat MLU

I calculated OUT-OF-BEAT MLU through the converse method to in-beat MLU discussed in the previous section: any utterance not analyzed as part of a beat, interrupted beat, etc. was considered an out of beat utterance. I counted the number of tokens uttered by each speaker on these out-of-beat utterances, and divided by the number of out-of-beat turns in which the speaker uttered those tokens. Again, I applied the same conventions for decimal-numbered turns and silent turns as in the previous section. Table 3, below, shows the out-of-beat MLU of each speaker in each dialogue except for the session 9 dialogues. In the session 11 dialogue, Speaker A uttered no out-of-beat utterances; thus this statistic is not applicable for him.

Table 3: Out-of-beat Mean Length of Utterance (tokens/turns) by Speaker and Session

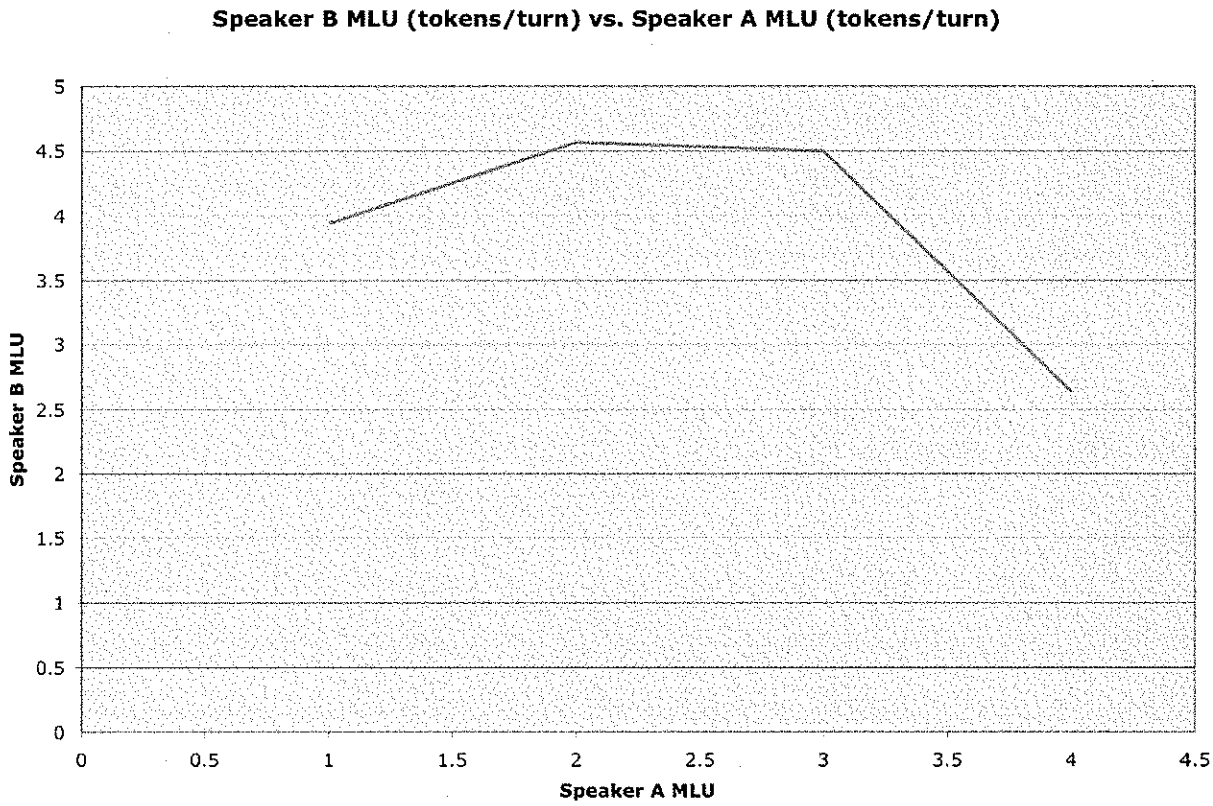
	Session 2	Session 3	Session 6	Session 8	Session 11
Speaker A	3.08	2.00	3.94	3.78	N/A
Speaker B	4.57	3.94	2.63	4.50	4.38

Unlike the global and in-beat MLU data, the out-of-beat MLU figures fall within a relatively small range, from 2 tokens/out-of-beat-turn to 4.57 tokens/out-of-beat-turn. This supports the idea that there is some learned social behavior that encourages a non-dominant speaker to use relatively few words. If this were not the case, ten- and twelve-word responses should bring these numbers, much lower than the in-beat MLUs, up. The average out-of-beat MLU across all sessions was 3.65 tokens/out-of-beat turn.

Just as with in-beat utterances, the out-of-beat utterances of one speaker should not take place at the same time as the other, and so there should be no connection and no correlation between them. During turn lapses, both speakers utter out-of-beat tokens at the same time, however. In fact, as figure 26 illustrates, below, the relationship between the out-of-beat MLU of the two speakers does not produce as erratic a pattern as the other MLU data in the previous two sections. However, since there are only four data points instead of five, and since five data points is not much to base a statistical finding on, I am reluctant to make any claims about the

out-of-beat MLU data. I point it out as an area that might yield interesting results with further research. Figure 26 below shows Speaker A's out-of-beat MLU plotted against Speaker B's out-of-beat MLU:

Figure 26: Speaker B out-of-beat MLU vs. Speaker A out-of-beat MLU



6.5 The unusual session: number 9

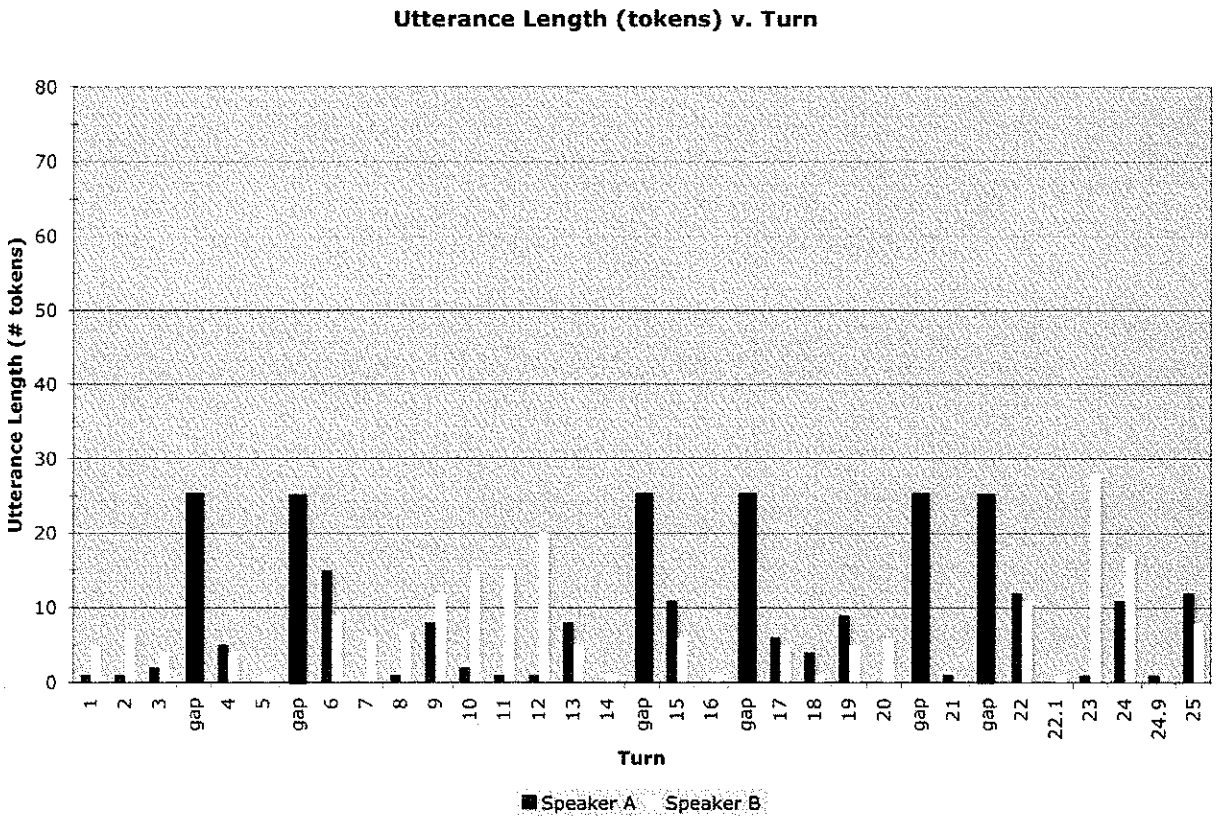
Earlier in section 5.1 I mentioned that the one session in which the subjects knew the purpose of the experiment did not seem to fit in with the rest of my data.

6.51 Dialogue 9.1

The first dialogue I analyzed from session 9, which I shall call dialogue 9.1, lasted from the very beginning of the subjects' conversation (about 1 min, 15s into the session) to about the 6-minute mark of the session. The conversation was frequently interrupted by long, silent gaps. These periods of extended silence – lasting, in order, for about 20 seconds, 30s, 30s, 8s, 30s, and 45s – were a feature unique to session 9. Only the 8-second pause was comparable to the shorter

gaps in conversation I found frequently in the other sessions. Chart 9 displays this dialogue according to the conventions above, except I have inserted black bars of a uniform length to indicate the extended silences. Note how sparse the conversation is: this dialogue is much shorter on the page and less varied than the other sessions' graphs.

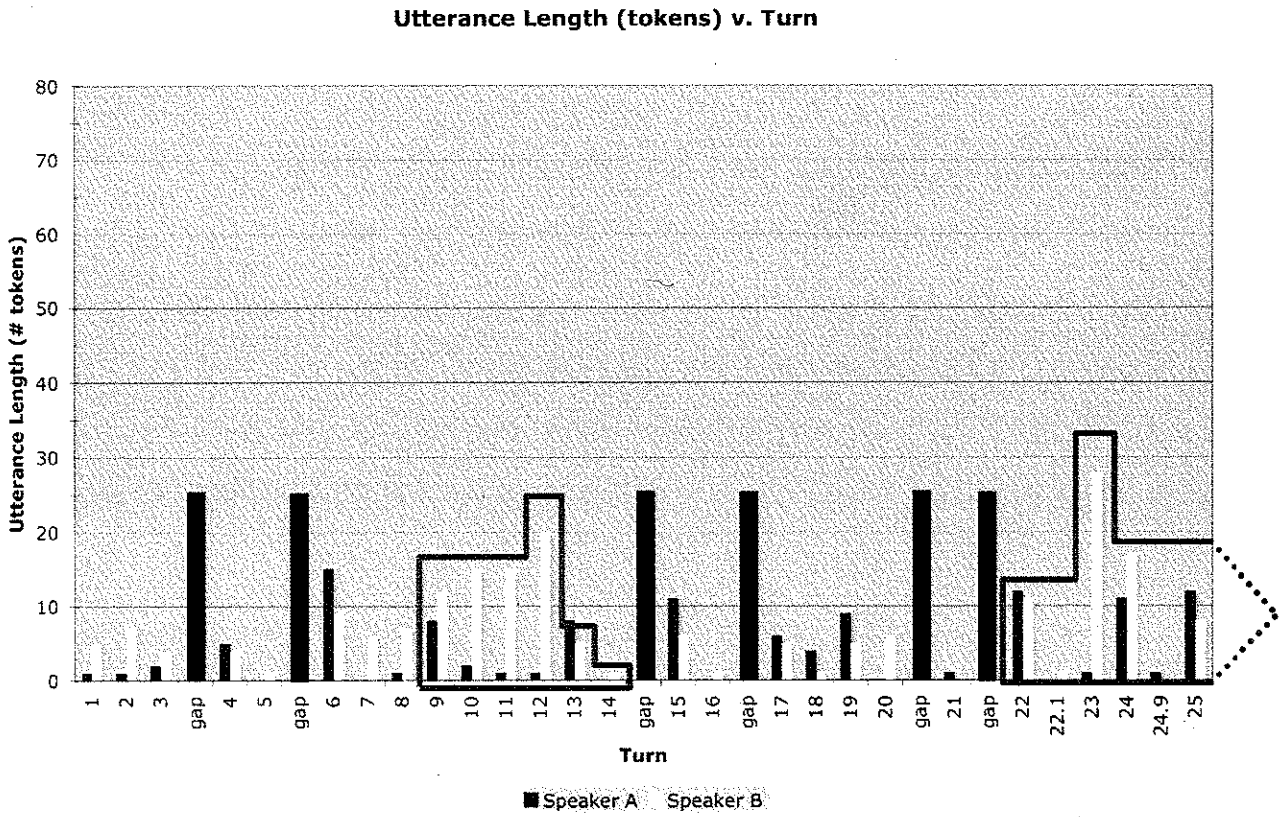
Chart 9: Session 9 Dialogue 9.1 – Utterance Length (tokens) v. Turn



I have included the gaps in the conversation in the graphical representation to help explain the other unusual feature of this dialogue, namely that it has very few beats. Figure 27, on the next page, outlines the two beats visible in Speaker B's speech in red. The rest of the conversation appears to consist of lapses macroturn structure. But, when the extended periods of silence are factored in as in chart 9, these lapses seem less unusual. Turns 1-8, for example, would look like an 8-turn lapse in otherwise continuous conversation if one were unaware of the long pauses before and after turns 4 and 5. With the pauses present, however, a story emerges: the subjects fished around for something to talk about (which would jumpstart the macroturn-

taking system), and, finding nothing after three turns, gave up. 20 seconds later, awkwardness forced them to try again, but with no success – another long pause. On the third try, however, they managed to find a topic of interest, which happened to be a discussion of what the experiment was about, revealing that each of them was aware of the point of the experiment. The subjects struggle to keep conversation going, however, and the second half of the dialogue ended up looking much like the first half, with lapses and gaps and, finally, another beat.

Figure 27: 2 beats, lots of lapses in dialogue 9.1



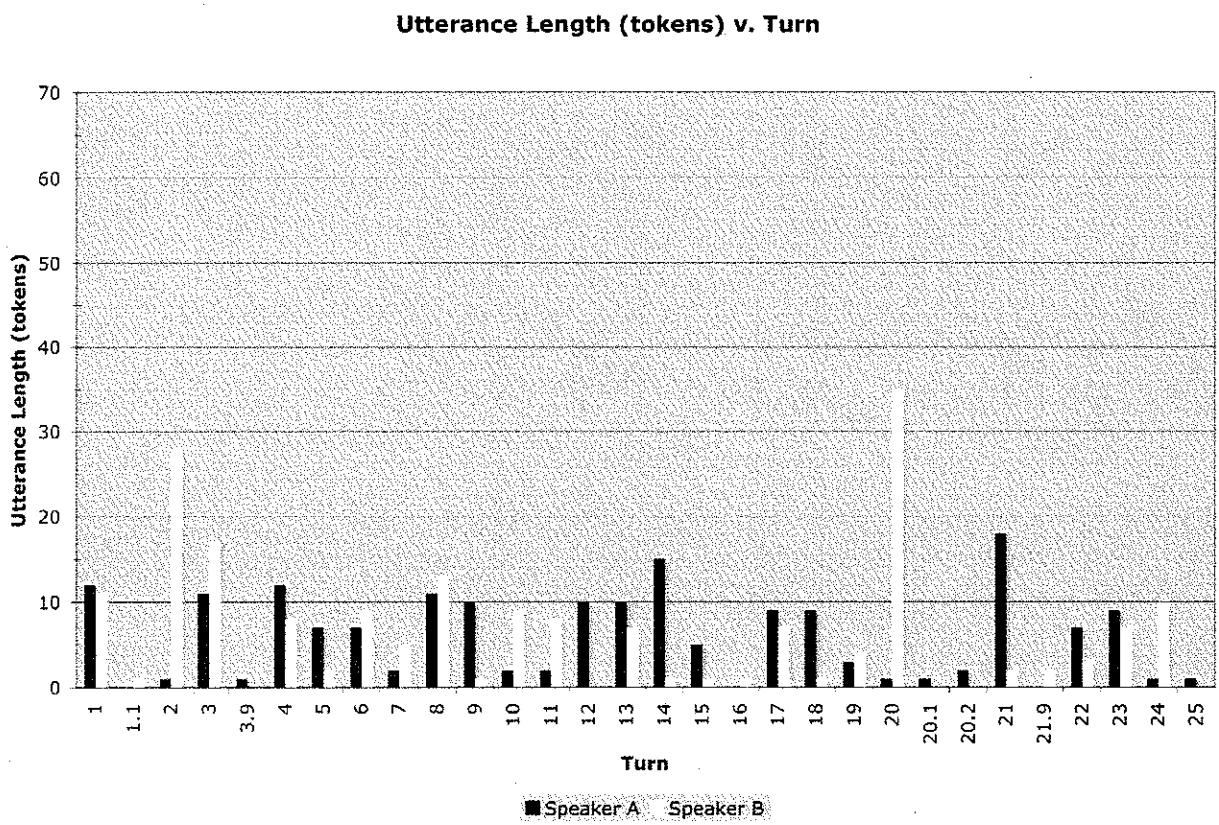
Part of the problem seemed to be that Speaker A was reluctant to assume dominance in the same manner as the subjects in the other sessions. She never produces an utterance longer than 20 tokens. This problem persisted throughout the whole session.

6.52 Dialogue 9.2

However, as session 9 continued, the subjects' conversation ceased to be interrupted by long gaps. Thinking, because of the absence of gaps, that the second half of session 9 sounded

more "normal" than the awkward, silence-ridden first half, I analyzed a second dialogue from session 9, which I shall call dialogue 9.2, lasting from about 5 min, 30s into the session (turn 22 in session 9.1, above) until about 8 min, 30s into the session. Chart 10, below, shows the data.

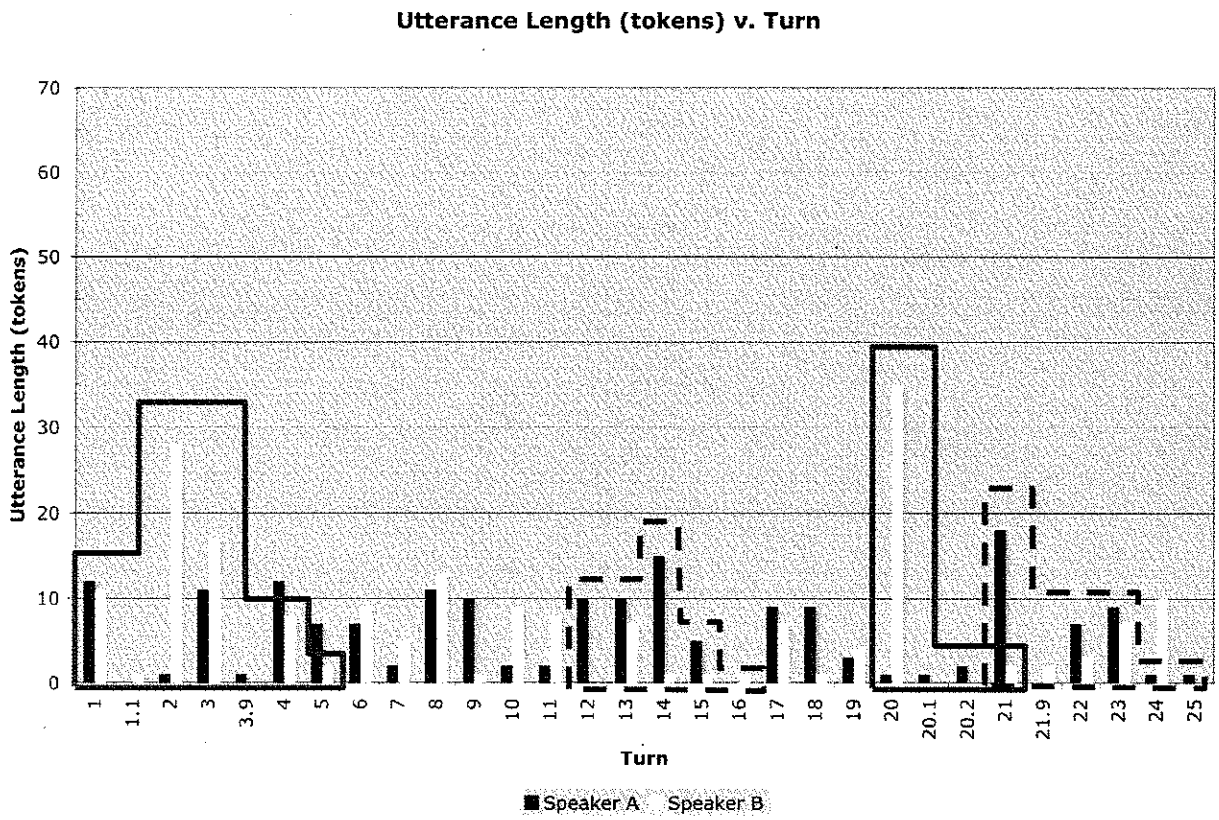
Chart 10: Session 9 Dialogue 9.2 – Utterance Length (tokens) v. Turn



Although there are no extended silent gaps in session 9.2, it still doesn't follow the patterns set but the other sessions' data. Speaker A still refused to unequivocally assume dominance with an utterance of 20 tokens or more, instead producing a slough of sentences around 10 tokens in length. Speaker B had a couple of intermittent beats, but between them again was a long lapse. Figure 28, below, outlines speaker B's beats in red lines as well as some possible shortened beats for speaker A in dashed indigo lines. By "shortened beats" I mean to say that, since speaker A does not ever produce any beatcenters 20 tokens or more in length, perhaps she has a different (shorter) sense of how long a beatcenter should be. Her turns 12-16 and 21-25 could be considered beats with shorter beatcenters, perhaps. But, regardless of

whether I analyze these turns as beats for Speaker A, there are still extended lapses without a beat (turns 6-11 and 17-19). I believe the macroturn lapses found in dialogue 9.2 arose because Speaker A's behavior was confusing to Speaker B: all of those around-10-token utterances seemed like prebeats but weren't. They prompted Speaker B to attempt to cede dominance to Speaker A, who had no desire to assume dominance.

Figure 28: Speaker A's medium utterances disrupt macroturn-taking in dialogue 9.2



Although dialogue 9.2 is more similar to the other dialogues than dialogue 9.1 is, it is still different. Extended lapses set it apart.

6.53 Significance of session 9

The failure of the session 9 dialogues to match the dialogues from the other sessions in form bodes well for my project. Speakers in the other dialogues could and often did guess that the task given to them was too trivial or random to be the point of the thesis, but even if they assumed their conversation would factor into the purpose of my project, they could not know

what aspect of it would be analyzed. Oppositely, the speakers in session 9 were the only subjects tested who knew that the point of the experiment was to test for utterance length in casual, two-person dialogue, rather than for something directly related to the task they were given. In a sense, the speakers in most of the sessions could infer, if they so chose, that they might be supposed to talk, while the speaker of session 9 knew that they must talk if their data were to be useful to the project. Thus, the subjects of session 9 were more conscious of their speech than the other subjects, and certainly more conscious of their turn-taking, because they knew that the course of their conversation would be under scrutiny.

If vernacular speech is the most natural, least self-conscious speech possible, then the more conscious the speakers are of their own speaking, the further removed their speech will be from the vernacular. Since the session 9 speakers were more conscious of their speech than the other subjects tested, their conversation is set apart as the least vernacular. In theory, it is therefore not too shocking that their dialogues should look somewhat different than the other sessions' subjects' dialogues.

The fact that the theoretical discrepancy between dialogues from session 9 and those from the other sessions was born out in practice leads me to believe that my experiment – designed to set the subjects at ease, distract from the real point of the project, encourage conversation, etc. (see section 2.4, above) – succeeded in eliciting vernacular speech, to the extent possible, from the subjects in the sessions other than number 9. Awareness of what aspect of conversation I would be studying implied a command to talk, rather like the experimental methodology that I rejected consisting sitting two subjects down and shouting, "Converse!" I did believe such an experimental strategy could yield vernacular speech for study, and, from the looks of session 9, this hypothesis was correct.

The second half of session 9, from which dialogue 9.2 is excerpted, sounded more "normal" to me than the obviously unusual first half, which makes up dialogue 9.1. It is

heartening to know that the normal-sounding dialogue more closely resembles the other sessions in form than the unusual-sounding dialogue, as this suggests that the other sessions' data was normal, that is to say, natural.

Furthermore, the fact that the awkward, self-conscious speech of dialogue 9.1 began to normalize as the session continued (session 9.2) may speak to the fact that, regardless of circumstance, the form into which conversation is organized should emerge sooner or later. After all, the point of studying vernacular speech is to see how people talk by default, when there is no disturbance, social impediment, or other factor that cause them to alter their speech. After a time, unless some reminder appears to keep the speakers vigilant, it seems to me that speakers should tend towards the vernacular because that's just how they talk. Evidence for this could be that session 9.1 – in which the speakers had just begun talking, with the point of the experiment in their minds –differed markedly from the other sessions, but session 9.2 – when they had been sitting together for 6 minutes, and sort of cobbled together some conversational momentum – differed to a lesser extent. Perhaps if the two women from session 9 had been kept together for a couple of hours and recorded, they would have eventually reached a point where their conversation resembled all of the other subjects', regardless of knowing the point of the experiment.

7. Conclusion

7.1 Answering the question

The impetus for the project was the question, "Does Bruce Graham's template for utterance length throughout a dialogue accurately reflect natural (i.e. real-life, nonliterary) speech?" Having gone through a considerable amount of data from real-life two-person dialogues comparable to Graham's playwriting exercise, what is the answer to the question?

7.11 No

Graham's template for the number of words per turn is reprinted below.

Figure 1 (reprinted): Utterance length template from Graham's exercise (1995)

Lines 1-20: 2-4 words	Lines 37-38: 20 or more words
Lines 21-30: 1-2 words	Lines 39-45: 4-6 words
Lines 31-36: 1 word	Lines 46-50: 1-2 words

In short, the answer is no. Graham's template lays out a LOCALLY and GLOBALLY BALANCED dialogue, and the data I collected shows conversations to be almost uniformly LOCALLY IMBALANCED, and are only LOCALLY BALANCED for short periods by accident or as a special effect (e.g. repeating after someone). Conversations can be GLOBALLY BALANCED, but they may also be GLOBALLY IMBALANCED.

As for the number of words per line, Graham's template does not reflect conversation in general because it proposes ABSOLUTE limits on utterance length. The speakers must use a certain number of words on the first turn, a certain number of words on the second turn, etc. Real conversation, on the other hand, made use of RELATIVE limits on length: the length of any given utterance relied on the lengths of the surrounding utterance by both speakers and on the speakers' intentions. It was not the case in my data that Speaker A's 1st utterance in every session was around 8 tokens long, and his or her second turn was around 20 tokens long, and his third turn was again around 8 tokens long, etc. From session to session, any given turn could be made up of an utterance of any given length.

Rather than being fixed by turn-position in the dialogue, utterance length reflected the speakers' interplay at a MACROTURN level that provides guidelines for improvisation, not rigid token limits. Utterances were organized within a loose structure of BEATS consisting minimally of a BEATCENTER (20 or more tokens) followed by a BEATENDER (less than 5 tokens, including 0). Speakers could fill optionally out a beat with a PREBEAT – an utterance or set of utterances before the beatcenter, (each) longer than about ten tokens and short than the beatcenter – or a POSTBEAT – an utterance or set of utterances after the beatcenter and before the beatender, (each) longer than 5 tokens and shorter than about half the length of the beatcenter.

The constraints on utterances' lengths evolve over time, based on the lengths of the utterances of each speaker as the conversation unfolds and on the speakers' intentions. When determining how long to speak on any utterance, a speaker might consider whose macroturn it was, whose it had been, at what point in the current macroturn the utterance in question fell, and how they intended to affect the macroturn structure. Speaker A, for example, could utter 1 token in her first turn, but she could equally well utter 50 tokens. If Speaker A produced a 50-token utterance on her first turn, Speaker B would be likely to recognize that Speaker A was initiating a beat, and that it was therefore her macroturn. Unless he wished (consciously or unconsciously) to interrupt her turn, he would probably produce an utterance of fewer than 5 tokens on his first turn. Speaker A would receive this signal that her turn was still in effect and could opt to prolong her beat with further extended utterances (a postbeat), or end her turn with a short utterance (a beatender). Depending on her choice, speaker B would assess the situation and decide whether to continue offering short responses (usually under 5 tokens) or initiate his own beat (with a prebeat or beatcenter). As the conversation continued to unfold, each speaker would use his or her own utterance lengths as a signifier of the macroturn structure while relying on similar signals from the other speaker. The result is a set of evolving constraints on utterance length that depend on context and intension.

Globally, the MEAN LENGTH OF UTTERANCE (MLU) of the speakers in Graham's template is around 4 tokens per turn, which falls within (though at the bottom of) the spectrum of the conversations studied in the course of this experiment. Thus, Graham's template cannot really be attacked for having the speakers talk too little, per se. However, the two speakers who exhibited global MLUs this low (Speaker A in Sessions 2 and speaker A in session 8) both found themselves on the responding end of the other speaker's macroturn for most of the conversation. It seems that if one speaker's MLU is as low as 4 tokens per turn, the other speaker's MLU should be high. Graham's template misses this point, as both speakers have the same low MLU.

The real problem with Graham's template is that there are no beats for most of it. The speakers carry on what is essentially a 36-turn-long lapse, both keeping their utterances at an OUT-OF-BEAT response level (under 5 tokens) until turn 37. Neither speaker asserts dominance, and neither speaker utters anything even remotely as long as a beatcenter. None of my sessions – even Session 9, a halting, awkward failed conversation – exhibited such a long lapse. Furthermore, at turn 37, both speakers initiate a beat at the same time. I would guess that if a speaker were reluctant enough to assume dominance to the extent that he or she waited 36 turns to do it, he or she would be happy to let the other speaker take the floor for an extra few turns and establish a conventional macroturn structure at last.

7.12 Did Graham lie?

If Graham's template is supposed to be based on actual conversations he studied, but does not appear to reflect actual conversation, is it necessary to conclude that he lied? Actually, the answer is no: there are a number of reasons why his template could differ from the form of the conversations observed in this study.

First an foremost, it is not certain whether the template was indeed based on real conversations (see note 2). Even if it were, Graham probably didn't carry out an experiment similar to the one presented here. Instead, he probably listened to dialogue informally, as I did as a precursor to my experiment. Because of the difficulty of accurately capturing spontaneous conversation without recording and transcribing it, it is entirely possible that Graham honestly could have based his exercise on real conversations and still failed to capture them exactly.

There are, in fact, certain features of Graham's template that indicated that he did base the template on real conversations – his perception of them, at least. The feature of the template that is most convincing in this respect occurs on lines 37-50. Here Graham outlines for the speaker something closely resembling a beat: a line of 20 or more words to begin, corresponding to my beatcenter of 20 or more tokens; a set of medium-length utterances of 4-6 words, corresponding

to my postbeat (more than 5 tokens, less than half the size of the beatcenter); and finally a short set of utterances of 1-2 words, corresponding to my beatender (less than 5 tokens). This portion of Graham's template would form the stair-step shape commonly found within my figures, if represented graphically.

I believe that this similarity was no accident. Despite the fact that the two speakers' beats occur simultaneously on lines 37-50, it seems that Graham deliberately outlined beats for his speakers in this portion of the outline, which requires a certain familiarity with the form of conversation. The flaw is Graham's perception was merely that, when on of the speakers initiated a beat and began to talk in much longer utterances, he seemed to believe that the other speaker would immediately begin to respond with utterances of comparable length. In actuality, speakers tended to waits a few turns until the other speaker's beat had finished before beginning their own period of lengthy utterances. Graham just needed to push the second speaker's 20 or more word lines back a few turns, and the template would be much more accurate for lines 37-50.

Of course, lines 1-36 are too short and too balanced to reflect conversation. Perhaps Graham believed as many people do that, since people often speak in incomplete sentences when speaking informally – a language intuition I have heard from friends and acquaintances – they must speak in short, clipped utterances. In actuality, they do speak in short, clipped utterances that may be incomplete sentences when they are responding to another speaker, but tend to speak in complete sentences, run-on sentences, or series of either of these when they take a beat.

Finally, as a playwright, Graham may have based his template in part on his observations of conversation and in part not, because he simply might not care about accurately reflecting reality. Graham self-avowedly concerned himself with the rhythms of lines as a theatrical device, not as a linguistic undertaking. He may have used rhythms found in real speech to add believability to his exercise, but may well have intentionally deviated somewhat from the form

of natural dialogue for rhythmic and dramatic effect. There is no particular reason that literary dialogue must exactly parallel natural dialogue, no more than a painting of a tree must look exactly like the tree acting as the model. Artists may imitate life when it suits their purpose, and not imitate it when it doesn't.

7.2 Revisiting my hypotheses

Reprinted below is figure 4, detailing my hypotheses based on the informal observations I undertook out before carrying out my experiment.

Figure 4 (reprinted): Hypotheses on the form of spontaneous, two-person dialogue

1) Utterance length will not, for the most part, be balanced:
   - 1a) – For almost any given portion of the conversation, one speaker will employ considerably more words per turn than the other
   - 1b) – The distinction of having longer utterances in the short term will abruptly switch from one speaker to the other at points throughout the discourse
2) If utterance length is roughly equal between the two speakers over a given period, then the nature of the conversation will either be formulaic speech, or periods of emotional intensity for both speakers (e.g. name calling).
3) The MLU of each speaker over the course of an entire dialogue should be larger than 4 and smaller than 19, probably between 7-12.

These hypotheses turned out to reasonably accurate. Although I had not considered local balance versus global balance at that time, my description in bullets 1a and 1b point to the fact that I believed conversation would be imbalanced in the short term, i.e. locally imbalanced. This was true throughout all of my sessions. However, I was partially incorrect with my claim in 1b, because DOMINANCE (the term I used to define the distinction referred to in that bullet) does not necessarily switch between speakers, nor does it always switch abruptly.

The macroturn-taking paradigm of beats that emerged from the data does indeed meet the description in 1a and 1b some of the time. Session 3, which I consider the most orderly of the dialogues I analyzed, followed this pattern. But, when the speakers do not alternate beats – because of renewed beats, nested beats, or one-sided talking in general – there does not

necessarily have to be a switch in which speaker is dominant, even over the course of the whole dialogue. Furthermore, dominance does not always switch cleanly and abruptly. Speakers can and do speak over one another, take overlapping beats, or LAPSE into a series of utterances in which neither speaker is clearly dominant.

I seemed to be on the right track with my second hypothesis, although again it did not capture all of the data. When utterance length was relatively locally balanced to the greatest extent (the back-and-forth one-liners in session 6), the nature of the speech was indeed as formulaic as possible: one speaker repeated after the other, and both produced counting numbers by rote. However, utterance length was locally balanced to some extent in the lapses occurring infrequently, but not excessively so, in the alternating-beat structure of the dialogues. Furthermore, in the unusual Session 9, the speakers produced a more locally balanced dialogue than in any other session. Perhaps self-conscious or awkward dialogue tends to produce locally balanced speech. Regardless, the number of means by which locally-balanced dialogue may arise is greater than I initially assumed in hypothesis 2.

My third hypothesis, based on the idea that speakers in spontaneous dialogue tended to speak with longer utterances than children (i.e. MLU > 4.00) and shorter utterances than callers to a customer service line (i.e. MLU < 19.00), is still up in the air. Certain of the speakers MLUs of near or under 4.00 tokens per turn, but these speakers spent all 25 of turns of the dialogue I analyzed responding to the other speaker's much longer utterances. It is possible that, if I had been able to analyze all ten minutes of each session, these speakers would have eventually taken extended macroturns themselves, and raised their MLU. It is also possible that some speakers simply speaker about 3.5 tokens per turn. The same applies to maximum MLU. It probably is not as high as 19.0 tokens per turn, since the greatest MLU I found was only 16.24, and that was from a speaker who did not seem to relinquish his turn throughout the whole dialogue. However, the actual maximum falls victim to the same problem of sample size: I did not have time to

examine the whole course of any conversation to see whether MLUs would converge on some specific range. Assuming that my data accurately represents all conversation, the average MLU was 9.72, within the range I set out in hypothesis 3.

7.3 Constructing the template

As the capstone to this project, the final task is to construct an utterance-length-by-turn template à la Bruce Graham that accurately reflects spontaneous dialogue. Based on my research, I have constructed one below in figure 29 that – although it by no means encompasses all the possibilities – represents one conceivable course a dialogue could take through 25 turns.

Figure 29: One possible conversation template – tokens by turn

Turn Numbers	Speaker A (tokens)	Speaker B (tokens)
1	20 or more	2-7
2-3	5-12	1-2
4	0-2	10-20
5	1-3	20 or more
6	1	1
7-8	1-5	10-20
9	3-10	20 or more
10-11	1-5	1-2
12	15-25	0-5
13	40 or more	5-12
14	8-15	20-25
15-16	8-15	0-2
17-19	0-2	10-20
20	10-12	20 or more
21-22	0-5	0-5
23	5-8	10-20
24	1-2	20 or more
25	1-3	1-3

I will briefly run through the course of the dialogue, to clarify the template. Turns 1-4 represent a beat for Speaker A consisting of a beatcenter, postbeat, and beatender. Speaker B

begins a beat on turn 4 with a prebeat, continuing with a beatcenter on turn 5 and beatender on turn 6. Speaker A, however, does not assume dominance on his turn 7, so Speaker B continues his macroturn with a new beat initiated by the prebeats on turn 7-8, and continuing through beatender on turns 10-11. On his turn 12, Speaker A assumes dominance with a slightly long prebeat that still remains about half as long as the beatcenter on turn 13. While Speaker A fills out her beat with a long postbeat (turns 14-16), speaker B interjects a nested beat in its simplest form on turns 14-16. Speaker A's beatender is then concurrent with another prebeat by Speaker B (turn 17-19 for both speakers). After Speaker B's beatcenter on turn 20, however, Speaker A's extended response on her own turn 20 leads Speaker B to end his beat on turn 21. There is a brief lapse because of the confusion (turns 21-22). On turn 23, both Speakers attempt to repair the lapse, but Speaker B's utterance is clearly a prebeat, and so Speaker A minimizes her utterance lengths through turn 24-25 as Speaker B finishes his beat with a beatcenter (24) and beatender (25).

Most of the decisions I made in constructing the template above were arbitrary. Particularly the response lengths of out-of-beat speakers, about which my analysis explains very little, were under 10 tokens for the most part but otherwise without pattern. However, the general progress of the dialogue and the means by which the speakers form beats and take macroturns reflects the analysis I have presented above. Some playwrights searching for a golden key to natural dialogue may be disappointed at the fact that I present only a general pattern, not a specific solution. However, linguists, language enthusiasts, and even many playwrights may rejoice that richness of the mechanism is, in the end, impossible to distill into a single conversational pattern. Just as speakers have considerable free will in determining what to say, they have the same amount of leeway in deciding how to say it.

Bibliography

Abrams, Joshua. 2003. Show people: downtown directors and the play of time, curated by Norman Frisch, Exit Art, New York City, 11 May-17 August 2002. An exhibit review. *Theatre Journal* 55.1. 161-164.

Brown, Roger. 1973. *A first language: the early stages.* Cambridge, MA: Harvard University Press.

Crist, Sean. 2005. Personal communication.

de Villiers Peter A. and Jill G. de Villiers. 1972. Early judgments of semantic and syntactic acceptability by children. *Journal of Psycholinguistic Research* 1. 299-310.

Dybkjær, Hans, Niels Ole Bernsen and Laila Dybkjær. 1993. Wizard-of-Oz and the trade-off between naturalness and recogniser constraints. Proceedings of EUROSPEECH '93, Berlin. 947-50.

Gorin, Allen. 2003. Semantic information processing of spoken language. Proceedings of the 8th international conference on Intelligent user interfaces. Miami, Florida.

Graham, Bruce. 2005. A bio written for inclusion in programs for his plays. Available through Dramatists Play Service, Inc. online at http://www.dramatists.com/text/authorbios.asp.

Graham, Bruce. 1995. Playwriting exercise distributed at an authors workshop circa 1995, given to me by Donna Jo Napoli (p.c) in September 2005.

Labov, William. 1984. Field methods of the Project on Linguistic Change and Variation. Language in use: readings in sociolinguistics, ed. by John Baugh and Joel Sherzer. New Jersey: Prentice-Hall. 28-53.

Labov, William. 1973. *Sociolinguistic Patterns.* Philadelphia: University of Pennsylvania Press.

Linguistics – social dialectology. 2005. Encyclopædia Britannica. Encyclopædia Britannica Online: http://search.eb.com/eb/article-35141.

Napoli, Donna Jo. 1982. Initial material deletion in English. *Glossa* 16.1. 85-111.

Napoli, Donna Jo. 2005. Personal communication.

Prieto, P. 2004. The search for phonological targets in the tonal space: evidence from five sentence-types in Peninsular Spanish. *Laboratory Approaches to Spanish Phonology*, ed. by Timothy Face. Mouton de Gruyter: The Hague. 29-59.

Sacks, Harvey et al. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50.4.1. 696-735.

Seabrook, John. 2005. Talking the tawk. *The New Yorker*. Available online: http://www.newyorker.com/printables/talk/051114ta_talk_seabrook.

Wang, Michelle Q. and Julia Hirschberg, 1991. Predicting intonational boundaries automatically from text: the ATIS domain. Proceedings of Speech and Natural Language (workshop). Pacific Grove, California, February 1991. 378-383.