Can You See Whose Speech is Overlapping?

Charles F. Meyer Robert A. Morris Ed Blachman

Recently in linguistics there has developed an increased interest in the analysis of computer corpora — samples of speech and writing distributed in machine-readable form. Computer corpora are typically annotated with markup to indicate such phenomena as paragraph boundaries and titles in written texts and pauses and speaker turns in spoken texts. As computer corpora become more common in linguistics, linguists need to concern themselves not just with developing standards for the markup they use but with ensuring that this markup is presented to the user in as readable a format as possible. In our discussion, we focus on a common characteristic of speech that any annotation system must deal with — overlapping speech — and describe software that we have developed that not only accurately marks the boundaries of overlaps but presents them to the user in a very readable format.

First we discuss the types of overlapping speech that any markup system will have to describe and then we critique two types of current systems for marking overlaps: those that stress readability and those that emphasize descriptive adequacy. We describe the problems inherent in each of these systems and conclude by discussing a system we have developed which is based on sophisticated document processing software. This system presents speech overlaps in vertical columns and balances the necessity of accurately describing the boundaries of overlaps with the need of the user to be presented this information in as readable a manner as possible.

Can You See Whose Speech is Overlapping?

Throughout the paper, we use sidenotes instead of footnotes. We do this in furtherance of our central point that more attention to the two dimensional nature of typography can address certain points of interest which arise in the written representation of spoken language

The study of English usage has a long tradition, dating back to the 19th and early 20th century grammars of Sweet (1891-98), Poutsma (1926-29), Curme (1947), and Jespersen (1909-49). These grammarians based their analysis of English on (primarily) literary texts of English and their lifelong work on these texts yielded much valuable information about the structure of English. Because these early grammars described only English and were heavily influenced by earlier grammars of Greek and Latin, they fell out of favor among those interested in studying non-Indo-European languages and in developing theories of human language. These interests gave rise to a range of new theoretical models of language, including structural grammars (such as Bloomfield, 1933) capable of providing linguistic descriptions of Native American languages and generative grammars intended to provide a theoretical framework for the study of language universals rather than the structure of individual languages (see, for instance, Chomsky, 1957 and 1975).

In recent years, however, some linguists have moved away from purely theoretical descriptions of language based on sentences they often invented themselves to studies based on the actual usage of language in texts such as those that the early English grammarians studied. These linguists recognized that if they wanted to obtain an adequate description of English usage, they had to not only investigate various types of English (e.g., spontaneous conversation, press reportage, technical documents) but base their investigations on a large enough sample to insure that an accurate description of English usage is obtained. To help the linguist carry out such studies, the computer has proven an invaluable tool for storing and analyzing large samples of language in the form of "computer corpora."

See Edwards and Lampert (1993:263f) for a comprehensive overview of the various corpora currently available or under development

The earliest computer corpus of English, the Standard Corpus of Present-Day English (Kučera and Francis, 1967), was compiled in 1961 at Brown University and contained 2,000 word samples of various types of edited written American English (e.g., journalism, fiction, belle lettres and government documents). This corpus spawned two parallel corpora: The Lancaster-Oslo-Bergen (LOB) Corpus of edited written British English (Garside, Leech and Sampson, 1987) and the Kolhapur Corpus of edited written Indian English (Shastri, 1988). In addition, the London-Lund Corpus of spoken British English was developed (Quirk and Svartvik, 1980) to allow the study of spoken English and to enable comparisons with written English.

Since the compilation of these corpora, there has been an explosion in the development of computer corpora. One of the projects of the Linguistic Data Consortium (Church and Liberman, 1991 and Walker, 1992) is to collect, computerize and grammatically analyze 100 million words of American English. The British National Corpus (Quirk, 1992) will be of similar composition except that it will contain 100 million words of spoken and written British English. The International Corpus of English (ICE) (Greenbaum, 1992) will consist of written and spoken samples of English from various regions of the world: Great Britain, the United States, India, Nigeria, Australia and New Zealand, to mention but a few of the countries whose English will be represented in the project.

The proliferation of corpora and hence computerized documents has made it increasingly important that a standard for the encoding of electronic documents be developed. In the Brown Corpus, for instance, a hodgepodge of symbols were used to distinguish periods ending sentences from those ending abbreviations; to mark section headings in scholarly articles and distinguish them from the text of the article and so forth. To establish such a standard, the Text Encoding Initiative (TEI) has begun an effort to develop standards for the mark-up of electronic documents (Sperberg-McQueen and Burnhard, 1994).

Our interest in the mark-up of electronic texts stems from our involvement in the American component of ICE. Because the TEI project is ongoing, ICE had to establish its own standards so that individual teams could begin marking individual corpora. The hope is to make ICE markup TEI-conformant once the TEI project has an approved system of markup. In the meantime, we have begun developing an editor (described in Morris, Blachman and Meyer, 1994) that will enable a user to interactively insert ICE markup into a document. In creating this software program, we have grappled with the normal problems that developing software raises. But we have found particularly problematic the development of a way of representing overlapping speech in an electronic document.

Overlapping speech is a phenomenon that occurs in speech, particularly spontaneous conversation, any time that two or more individuals speak simultaneously. For instance, in the fictitious conversation to follow, speaker B's first words overlap with speaker A's (overlapping segments are set off with slashes):

```
Speaker A: I heard that my son's history teacher may win
a teaching award /he/
Speaker B:      /You've/ got to be kidding
```

Overlapping speech is of interest to linguists because it gives a clue to how speech is organized. As Sacks, Schegloff and Jefferson (1974) note, one way for an individual to gain the floor is by overlapping his or her speech with the person who is currently speaking. Overlapping one's speech with another's is a technique for gaining the floor not just in languages such as English but in other languages as well (see, for instance, Hafez's 1991 study of turn-taking in Egyptian Arabic).

Encoding overlapping speech in an electronic document is problematic because a document is essentially a linear text, with each piece of the text occurring separately. When two pieces of the text overlap, however, the linear nature of the text is disrupted, and representing this disruption in an electronic document has created in current markup systems a tension between the need to adequately describe where overlaps begin and end and the desire to present this information to the user in a readable and visually revealing manner. As we will demonstrate, none of the systems for annotating overlapping speech is able to satisfactorily reconcile this tension: most that comprehensively mark overlaps create a visually difficult text to read; those that strive for readability tend to be descriptively inadequate.

To demonstrate how we developed a markup scheme that resolves the difficulties described above, we first outline the types of overlapping speech that any markup system will have to account for, and then detail how these instances of overlaps are inadequately treated by current markup schemes. We conclude by describing our markup scheme and the manner in which it overcomes the inadequacies of existent markup systems.

The types of overlapping speech

Even though the samples of spoken English we have been working with contained many different types of overlapping speech, we have determined that there are two basic kinds of overlaps: "paired" overlaps, in which single stretches of speech overlap in the turns of two speakers; and "multiple" overlaps, in which more than a single overlap occurs in a speaker turn or more than two speakers overlap. We found it necessary to distinguish these two types of overlaps because while single overlaps require merely that the sections that overlap be marked, multiple overlaps require not only that the sections be marked but that some additional system of marking be developed (such as numbering) so that it is clear precisely which segments go together.

See endnote 1 for an explanation of the identification symbols and numbers following examples we have taken directly from corpora.

The examples below contain instances of paired overlaps. In the first example, the segment of speech ending Speaker A's utterance, *trash,* overlaps with the segment beginning Speaker B's turn, *your.*

```
A:    so everyone has a minimum so that takes care of your your maintenance and your snow and your /trash/
B:    /your/ maintenance and all that stuff
(ICE-USA-S1A-002)
```

In the next example, single segments of speech also overlap, except that the first segment, *I'm,* occurs in the middle of Speaker B's turn; and the segment of Speaker E's turn that overlaps with *I'm, no,* is uttered a few seconds prior to the segment of speech (*I asked Dad....*) that follows it.

```
D:    do we have dinner reservations for a certain time tonight
B:    I don't know it's my birthday /I'm/ not worrying about it
E:    /no/ I asked Dad he's not yet so
(ICE-USA-S1A-003)
```

The examples below illustrate the two kinds of multiple overlaps. In cases like these, it becomes necessary to number the overlaps to insure that there are no ambiguities concerning which segments overlap.

In the first example, Speakers A and B overlap on two separate occasions: *mean* overlaps with *no*, and a few words later in the same turn, *that at all* overlaps with *when you go into*.

```
      A:    it's too oh it seems like there's like fat deposits
in there you know what I /1 mean/
      B:    /1 no/ it's not like /2 that at all/
      A:    /2 when you go into/ Portugal you know they don't
even it's not even like it's refined it comes right out of
the cow they just put it through a little strainer over the
sink and then here you go it's warm you know no refrigeration
it's like when you take it it's like a big lump of cream
(ICE-USA-S1A-004)
```

In the next example, three different speakers overlap on separate occasions. The first syllable of a word in the middle of Speaker A's turn (*re*) overlaps with an expression (*uhm-uhm*) uttered by Speaker B. Following this overlap, the last word in Speaker A's turn (*perhaps*) overlaps with the first words in Speaker C's turn (*this is*).

```
      A:    but this is a good illustration of how studying some-
thing forces you to /1 re/conceptualize as you just said what
how language works because language does not work necessarily by
genre but by parameter /2 perhaps/
      B:    /1 uhm-uhm/
      C:    /2 this is/ more of a scientific methods question
what happens when you get half way through a research project
and you realize that your original conceptualization was all
wrong you have to go back and start all over
      A:    yes that's why you have to you know how long we've
been planning this project
      B:    I think a very long time indeed
      A:    Yes
      (ICE-USA-S1A-001)
```

Because overlapping speech has been the subject of much linguistic research, it has become necessary to develop systems of annotation for marking the instances of speech that overlap. These systems have been employed in two different contexts: written documents (such as scholarly books and articles) in which linguistic research on overlapping speech has been reported and computer corpora (such as the London-Lund Corpus) which are available for general linguistic research. The systems that have been developed are of three types: 1) those that emphasize "readability"; 2) those that strive for "descriptive adequacy"; 3) and those that attempt to strike a balance between readability and descriptive adequacy.

Systems stressing readability

The earliest systems for annotating speech overlaps were developed by linguists engaged in conversation analysis. Although these systems vary in detail, all are clearly biased towards readability at the expense of descriptive adequacy.

The systems employed by conversation analysts are quite readable because they rely heavily on vertical alignment to depict overlaps, a visual device that satisfies a number of Edwards' (1992: 131f.) criteria for readable markup. For instance, in the example below (cited in Schenkein, 1978:xii), the segments *a lot* and *I see* are uttered simultaneously, a fact that is conveyed by the vertical alignment and matching brackets. This method of annotation is readable, Edwards (1992) argues, for three reasons. First, "related events" (in this case, segments of speech that overlap) are "placed spatially closer to each other than those which are less closely related" (Edwards, 1992:131). In addition, the brackets are "iconic"— they provide a visual cue as to the simultaneity of the utterances — and symbols that are iconic help "to expedite recovery of their meaning during rapid reading" (Edwards, 1992:131). Finally, the markup is economical and therefore avoids cluttering up the text with distracting information.

```
Tom:    I used to smoke a lot more than this
                            [     ]
Bob:                         I see
```

Unfortunately, while vertical alignment and bracketing quite satisfactorily mark paired overlaps, they are inadequate for annotating multiple overlaps. We attempted to use the markup conventions above to annotate a series of multiple overlaps in an example cited earlier in the paper and the results were anything but satisfactory:

```
A:   it's too oh it seems like there's like fat deposits in
     there you know what I mean
                         [       ]
B:                         no      it's not like that at all
                                                  [          ]
```

```
A:                                                     when you go
into Portugal you know they don't even it's not even like it's
refined it comes right out of the cow they just put it through
a little strainer over the sink and then here you go it's warm
you know no refrigeration it's like when you take it it's like
a big lump of cream
(ICE-USA-S1A-003)
```

The problem with the above example is not its readability but its accuracy: in the last turn, *into*, which begins the second line of the turn, is actually part of the sequence (*when you go*) that over laps with the last part of Speaker B's turn. But because *into* extended too far to the right in the first line of the turn, margin settings forced it onto the next line. While the margins could be reset for this example, the most important point is that vertical alignment cannot be consistently maintained in the system above because there is nowhere to place the right bracket if the overlaps extend beyond a line.

Importability of alignment often results from the use of the space "character" for formatting. With sophisticated document processing software, or even with desktop publishing programs, the interword space is often manipulated by the software to make typographic improvements.

Yet another problem with vertical alignment is its fragility. Du Bois (1991:89) notes that the alignment can be lost "if a user changes the margins, tab settings or fonts...". This is particularly problematic when documents are transferred from one system to another. For instance, when we ported a draft of this paper from Word Perfect 4.2 to Interleaf, the example on page 117 was rendered as:

```
Tom:     I used to smoke a lot more than this
                              [      ]
Bob:                            I see
```

Without manual intervention, the segments *lot more* and *I see* are no longer vertically aligned either with each other or with the braces intended to show overlap.

Systems emphasizing descriptive accuracy

The systems used by conversation analysts predate the widespread availability of low-cost computers and have focused on readability for printed text. However, many of the systems used to annotate overlaps in computer corpora have been biased towards descriptive adequacy and have shown little concern for readability. These systems surround the regions of overlap with special tokens.

This convention is not unique to linguistic scholarship, but is used in other disciplines where unambiguous description is required.

For example, concurrency must also be represented in computer languages for programming parallel computers (systems with more than one processor unit operating simultaneously on the data). That technology always permits but does not require the overlap of computation. In fact, for such computers, the actual order of overlap is never specified. Representation of such languages must only specify which instructions are permitted to overlap and iconicity is not an issue. Special tokens marking the concurrency boundaries provide a solution. A typical example is

```
cobegin
      instruction1;
      instruction2;
      instruction3;
coend
```

indicating that instruction1, instruction2 and instruction3 may be carried out in any order and with any degree of overlap.

A similar token–based strategy is employed in the markup of corpora, except that the order of the overlaps is conveyed by the linear order of the text. In the London-Lund Corpus, paired overlaps are indicated by stars placed around the segments that overlap. In the example below, *I* in Speaker A's turn overlaps with *and* beginning Speaker B's turn.

```
A:        yes that I think you told me *I*
B:        *and* none of them have been what you might call
        very successful in this world (LLC S.1.13 3-5)
```

In instances of multiple overlaps, stars are placed around one pair of overlapping segments, and plus signs around the other:

```
B: yes but they +would have been very few+
 *hardly any*
A +<<murmur>>+ *anyway*  (LLC S.1.13 98-100)
```

In the International Corpus of English (ICE), separate markup is used to mark the start of an entire string of overlapping speech (<{_>) and its end (<{/>). In addition, each separate segment of overlapping speech is marked by a tag that begins the segment (<[_>) and ends it (<[/>). In the example below, the entire string begins with *thought* and ends with *his*. Within this string are two individual segments that overlap: *thought* and *and what's his*.

```
<$B><#/> yea that's right it's Anthony's birthday
<$D><#/> have a little birthday party
<$C><#/> today's more important
<$B><#/> just because I'm a hundred and one
```

```
<$D><#/> I <{_><[_>thought<[/>
<$B><#/> <[_>and what's his<[/><{/>
(ICE-USA-S1A-003)
```

For multiple overlaps, the same markup is used except that each separate string is numbered to insure that each segment in the string is unambiguously matched:

```
<$D><#/> now honey
<$C><#/> come on nightcrawlers cost money
<{_1><[_1>babe<[/>
<$D><#/> <[_1>when you<[/1><{/1> spend all the money you're
going to have to fork <{_2><[_2>over your birthday
cash<[6/2>
<$C><#/> <[_2>nightcrawlers cost money<[/2><{/2>
(ICE-USA-S1A-003)
```

After this article went to press, TEI released version P3 (Sperberg-McQueen and Burnhard, 1994).

The Text Encoding Initiative (TEI) is based on the Standard Generalized Markup Language (SGML), whose markup is by definition non-iconic. The markup for spoken texts (TEI P2, Chapter 34) provides a mechanism whereby timelines can have points specified with absolute or relative time, or no time at all (merely order). These points are assembled into one or more alignment maps and references can be made from those timeline locations to marked points in the text or vice versa or in both directions.

Thus, TEI would mark an earlier example (see page 116)

```
<align> <loc ID=P1> <loc ID=P2> <loc ID=P3><loc ID=P4>

<u who=A>it's too, oh it seems like there's like fat
deposits in there you know what <ptr target=P1> I mean <ptr
target=P2>

<u who=B> <ptr target=P1> No <ptr target=P2> it's not like
<ptr target=P3> that at all. <ptr target=P4>

<u who=A> <ptr target=P3> when you go into <ptr target=P4>
Portugal they don't even it's not even like it's refined...
```

This example has a particularly simple alignment map, which only marks the positions of its time points. If there were data about the exact timing, the alignment map might look like this:

```
<align type=time units=sec>
<loc id=P1 absolute="12:00 EST">
<loc id=P2 distance=1 since=P1>
<loc id=P3 distance=10 since=P2>
<loc id=P4> distance=2 since=P3>
```

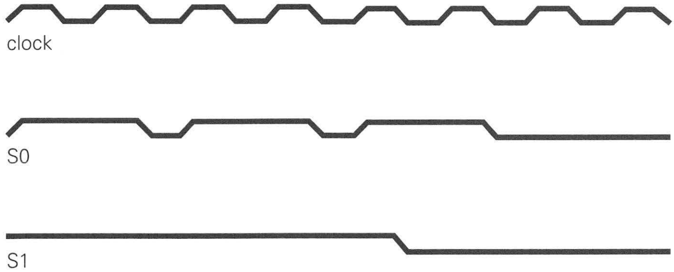
However, standard desktop computers have sufficient power to run software which can turn off all the markup at the user's request, thereby showing pure, readable text. An adequately designed program could do this in selective ways, permitting the scholar to easily emphasize all the speech satisfying particular linguistic properties described in the markup. For a discussion of a program that strips markup, ICECUP, see Quinn (1993).

While these systems are all descriptively adequate, their comprehensiveness creates a text that is full of markup and that is thus visually difficult for the user to read. In fact, all of Edwards' (1992) criteria for readability discussed above are violated, particularly in the ICE markup: related events are not placed spatially together, none of the markup (asterisks, curly braces, brackets) is iconic and the symbols are not concise but instead quite lengthy.

Mixed systems

The requirement to have both precise description and easy visual interpretation of concurrent events is not unique to linguistic scholarship. Traditional western musical notation strikes such a balance successfully, but so do specialized representation schemes in some other disciplines. For example, timing diagrams for digital electronics have a very music–like notation, but have the timing signature indicated on its own "staff line," the line labeled "clock" in the following example. All other timing is indicated relative to the pulses provided by the clock.

For speech, a number of markup systems attempt to strike a balance between readability and descriptive adequacy, and while these systems succeed quite well in this regard, they have individual inadequacies.



Typical integrated circuit timing diagrams indicating the synchrony of two signals compared to the clock signal. S0 and S1 must start rising at the same time, but they begin falling at different times. S1 has a duration of five clock ticks, whereas S0 has a repetition three times, each with duration two clock ticks.

In the system of annotation to be used in the Corpus of Spoken American English (CSAE) (Du Bois et al., 1993), an elaborate and elegant system has been devised to annotate both paired and multiple overlaps. In the example below, the first pair of overlaps, *to leave* and *They don't,* are vertically aligned and enclosed in single brackets. The next pair of overlaps are not vertically aligned but are enclosed in numerically indexed brackets: the lack of vertical alignment insures that the subsequent overlaps can be vertically aligned. The final overlaps are vertically aligned and numbered. In this system, the numbering is kept to a minimum both to insure readability and to prevent the numeral *1* from being confused with lower case *l* or the first person pronoun *I.*

In the published version of this example, the overlaps shown here as [2out2] and [2Berkeley2] were indicated with double brackets: [[out]] and [[Berkeley]]. The numeric notation reflects current CSAE practice. (J. Du Bois, personal communication)

```
B: Nobody wants [to leave].
A:              [They don't] move [2out2]
S: [2Berkeley2] just keeps [3getting3] bigger [4and bigger4]
B:                         [3Yeah3],
                                              [4Yeah4]
   ... Well it's amazing to me
(Du Bois et al. 1993:51)
```

The CSAE system is the best of the systems we have surveyed so far because it both adequately describes single and multiple overlaps and, additionally, can present some overlaps in a visually clear manner. However, raw CSAE coding is subject (from a readability perspective) to the same problems we noted above with respect to vertical alignment approaches. Its advantage over those approaches is that because it is descriptively adequate, vertical alignment problems can be corrected by software or by sufficiently energetic human intervention.

A different approach is taken in the HIAT transcription system (Ehlich, 1993). HIAT is perhaps the most fully iconic system currently in use, as it adopts a visual approach not only to overlapping speech but also to aspects such as tone, volume, tempo, emphasis, interpretation and even non–verbal communication. HIAT's approach to overlapping speech is inspired by musical scores; it is a rigorous formalization and extension of the vertical alignment approach we've discussed earlier.

In most vertical alignment schemes, new lines indicate turn boundaries. Consequently, overlap in turns too long to fit on a line cannot be iconically represented. Ehlich's HIAT scheme (illustrated below) maintains iconicity by placing all overlapping dialogue within simultaneity areas, denoted by the rulings or simultaneity braces. When one simultaneity area is insufficient, another may easily be added. This approach is not limited to displaying simultaneity; for instance, in the example below it also displays transcription conjectures (in parentheses), both simple and double. Note that iconicity is maintained at the expense of delimiting turn boundaries by line ends, though it seems to us that a boundary marker could easily be added to the scheme. In the example below, T begins by saying *Yeah,* then pauses while S1 says *Exactly!,* and then goes on to say *If you had simplified...,* with *If* overlapping the *act* of S2's *Exact.* It is impossible to tell whether this is one turn or two on T's part.

This notation is limited in the precision with which it can indicate overlaps. It is hard to tell from the display below whether T's *No* really overlaps the garbled conclusion of H's speech. Ehlich suggests that extra rules can be superimposed on the display to resolve questions like this (Ehlich, 1993:131).

```
   T:                                                   Yeah            If you had
   H:   One could / one could uh divide by six at once.
  S1:                                                       Exactly!
  S2:                                                              Exact
1 Sy:                                                                     (I got the
```

```
   T:   simplified by six at once, then the same res/      No, leave it! Would've
   H:                              Shall I (wipe it out)?
  S1:
  S2:
2 Sy:   same)
```

```
   T:   been (immediately) the same result.
             (instantly)
   H:
  S1:
  S2:
3 Sy:
```

(Ehlich, 1993:134)

For some purposes this lack of precision may be a blessing; to the extent that the transcriber is struggling to make sense of a somewhat indeterminate dialogue, this imprecision may be an accurate reflection of the audible record. But from the point of view of a linguist who's interested in overall analysis of a corpus rather than detailed analysis of a sample, this imprecision just makes this particular sample somewhat less useful.

Our work

We describe next the advances we have made in two areas. First, we have invented a music-like strategy that retains all of the advantages of such strategies but which is as comprehensive as the token-based strategies. Second, we have developed software that allows reading and annotation using our music-like strategy to be done even when there is a preferred underlying token-based strategy; thus a user can have the best of both worlds.

Like Ehlich, we find the case for a music–like strategy compelling. Our approach makes use of the most common approach to explicit formatting alignment: the ruled table. We use rulings to indicate overlap boundaries and thickened rulings to indicate turn boundaries. Our major departure from other approaches is to present the flow of the dialog vertically, with each speaker getting a column of the table. Below we show some of the previous examples in this form.

A	B
it's too oh it seems like there's fat deposits in there you know what I	
mean	no
	it's not like
when you go into	that at all
Portugal you know they don't even it's not even like it's refined	

This is a version of an example shown previously. Speaker A has the floor to begin with, yields it to speaker B during the first overlap, then uses the second overlap to retake it.

In the following three–speaker example, most of the dialogue belongs to speakers A and C... but B's *uhm–uhm* can't be neglected.

A	B	C
but this is a good illustration of how studying something forces you to		
re	uhm-uhm	
conceptualize as you just said what how language works because language does not work necessarily by genre but by parameter		
perhaps		this is
		more of a scientific methods question what happens when you get half-way through a research project and you realize that your original conceptualization was all wrong you have to go back and start all over

A: yes that's why you have to know how long we've been planning this project
B: I think a very long time indeed
A: Yes

Note that in this case the annotator chose to locate speaker A's second turn outside the tabular representation of overlap. That turn could just as well have gone in A's cell in an extra row, denoted by a thickened rule and added to the bottom, so as not to indicate an overlap with speaker C's turn. Indeed, a dialogue could be recorded entirely within a table. While perfectly readable, this would result in excessive white space wherever there were no overlaps. More likely, the annotator will choose a mixture comprising tables for the overlapped and closely related portions of the dialogue and conventional layout for the rest.

Like HIAT, there's a sense in which this notation is music-like, but it is most definitely not music. Perhaps the biggest difference is that music treats duration (at least relative duration) directly. Notes not only have position relative to each other and to the bars in which they are found, they have duration as well.

In contrast, rows of our tables have irregular durations — they "last" as long as the overlap they denote and deny the possibility of meaningfully indicating duration at a lower level. Also, our scheme can lose iconic representation of word boundaries. For reasons such as these, our notation is unsuitable for some linguistic purposes.

The software

The main software developed so far is an annotator's assistant for the ICE markup described earlier. Its purpose is to help prevent errors during annotation. Spelling errors in the tagging are the most obvious kind of error. Most word-processing programs in use by corpus linguists for annotation have mechanisms that eliminate spelling errors in token-based tagging. The most common scheme uses the so-called "macro capability" available on many personal computers. Macros allow the user to define sequences of characters which will be automatically inserted in the text when specific keys are pressed. Since most computer keyboards have keys corresponding to more than the usual assortment of characters in a single (western) language, these extra keys are often used for such special, user-defined purposes.

Thus, misspelling tag names is a very simple error to prevent. Preventing constraint violations is more difficult. These errors arise because the marking scheme includes certain stated or deducible constraints. Unless the software enforces them, an annotator can enter illegal or nonsensical annotation leading to confounding of subsequent processing programs. For example, something denoted as a paragraph should never appear in the middle of something denoted as a sentence, although something denoted as a quotation sometimes may and sometimes may not.

There are no textually implied constraints as to where overlaps begin and end, so any strategy can, at best, offer the annotator easy ways to examine the nature of the overlaps and compare it to the original source. However, there are obviously restrictions on the relations between the time boundaries of speech fragments and the linguistic structure. For example, a word does not normally begin in the turn of one speaker and end in the turn of another (perhaps only in a play by Eugene Ionesco!). In our software, the linguist can express such constraints in a relatively simple fashion: the entire set of valid tags is given a hierarchical structure and this hierarchy is examined at each point at which a menu of tag choices is to be offered. The tags currently in the document are parsed and no tags are offered other than those which can fit at that level of the hierarchy. Note, however, that "tags" marking the boundary of speech phenomena are not explicitly entered by the annotator, nor shown as such in the text. They are in fact represented by the table cell boundaries.

See English (1990) for a technical description of the facilities which enable the building of specialized editors such as ours.

The software is built on a powerful commercial document preparation system which can run on engineering workstations and medium-powered PC's. It has a built-in table editor which had most of the capabilities we needed for our speech representation. That editor largely supplied adequate capability to our design, though a few of its table ruling restrictions hinder us at some points. For example, it does not quite support our representation in case a table breaks at a page boundary. Using the page break control of the underlying software, we can insure that this never happens, but this potentially distracts the reader and may make the transcriber's task harder. Details of our constraint-based editor are available in Morris et al. (1994).

Limits to readability

While we believe that we have found a mixed system approach that offers good readability without sacrificing descriptive accuracy, we are also aware that there are underlying limits to readability that no existing system, and perhaps no system at all, can escape.

One such limit is typographical in nature. Any attempt to portray speech in writing must attempt to deal with the differing granularities of the two — the phoneme in speech, the printed character in writing. The size of written phoneme representations is not guaranteed to (and often does not) match the size of the phoneme; as a result, written representations of low-level overlaps will look "jerky" as the indicators of overlap will begin to take up as much space as the phoneme representations themselves. Worse, in a big table, these might even break across pages.

In fact, this "jerkiness" can occur even when the overlaps are somewhat less fine-grained, as in the following example:

Ken	Joanne
No. I don't thi	
nk so.	Isn't Ni
	ca
No.	ra
	gua one of the things you
I don't think so.	places you can't
	go?
No, we have – No	I think is is
I don't think so. We have f- f- Techni-cally speaking, full diplomatic re	
lations	Oh yeah?
with Nicaragua	
Yeah, they have an an ambassador and, you know an American ambas	
sador there who –	I – then why does
	everybod
like a nest of CIA spies	y always have to go through Mexi
	co
That's just –there– there isn't a direct transportation	
but there- it's not-	I thought - no but I
	th
it's	ought it
not like going to Cuba	
Wh	
ere they have to	Oh –
connections through Mexi	
co because	I knew someone
	who went to Cuba and had to go

From the Corpus of Spoken American English, with permission.

With overlaps like these, we believe the tradeoff between readability and descriptive accuracy can't be avoided.

```
Ken:       ... (TSK) No.
            .. I don't thi[nk so].
Joanne:                   [Isn't Ni]ca[2ra2]gua one of the things you –
Ken:                                  [2No2].
Joanne:    .. [3places you can't3] go?
Ken:          [3I don't think so3].
           [4No,
Joanne:    [4I think is i=s4].
Ken:       we have –
           No4]=,
           I don't think so.
           We have f- f- –
           .. (H) T=echnically speaking=,
           full= .. d=iplomatic re[lations] with Nicaragua,
Joanne:                            [Oh yeah]?
Ken:       Yeah,
           they have a n- an ambassador=,
           and,
           .. you know an American ambas[sador there=,
Joanne:                                 [I –
           then why does] everybod[2y always have to go through Me=xi2]co=?
Ken:       who] –
                                  [2like a nest of CIA spies2].
           .. (H) That' just –
           there –
           there isn't a direct transportation,
           [but there,
Joanne:    [I thought,
Ken:       it's not],
Joanne:    no,
           but I] th[2ought it <X w- X>2] –
Ken:                [2(H) it's2] not like going to Cuba.
           .. Wh[3ere they have to3] make connections thr[4ough Mexico because4],
Joanne:         [3Oh=3].
           [4(TSK) (H) I knew someone4] who went to Cuba,
           and had to go,
```

This is the original transcription of the example on page 128. Overlap numbering indicates which segments overlap. We extracted only the turn boundaries and overlap from this transcription. Although our software can support it, the markup given here is richer than what we are illustrating in this paper. In particular, the parenthetical expressions in this sample indicate inhalations and other noises which occupy time; since they are not indicated in our tabular markup, we have attempted to show in our table where the actual overlap begins. This leads to an illusory variance from the alignments here, for example in the last overlap, in which Joanne's noises begin in the middle of Ken's "through," but her speech in the second syllable of his "Mexico." Finally, note that turn boundaries are based on linguistic judgements made by the transcriber. Because overlaps make precise determination of turn boundaries difficult, our placement of them may not agree with CSAE's.

We took this monospaced sample directly from a diskette graciously provided us by the CSAE.

Another fundamental limit on readability is the physical page. In traditional vertical alignment approaches (and even HIAT), the limiting factor is the page width and resulting line breaks. In our notation, it is page depth and page breaks. In either case, some dialogue cannot be contained within the limits imposed by paper, and either iconicity breaks down entirely, or some convention is adopted that preserves iconicity while increasing the interpretive burden on the reader. In the HIAT representation, the numbering of areas alerts the reader to the continuation of a dialogue across line breaks; in our representation, when a table is continued across a page break, the bottom rule on one page and the top rule on the next page are omitted to indicate the continuation.

Page width becomes a limitation of our scheme when there are a large number of speakers. If there are not too many speakers, we can set the table on a turned page.

Ultimately, however, these limits reflect the embryonic nature of current technology. Any scheme with pretensions to iconicity is really an attempt to visualize the nonvisual phenomenon of overlapping speech. Visualization of nonvisual processes is currently an active area of computer science research. While efforts to date have focused on data-intensive processes (e.g., weather prediction), the spread of high-performance technology will inevitably bring attention to bear on more mundane visualization problems as well. In the end, the best way to see overlapping speech might not look anything like a printed page, nor be subject to the restrictions of print. And it will almost certainly allow the linguist to hear the speech at the same time that he or she sees it.

The use of text as the visual medium has one great disadvantage and one great advantage. The disadvantage is that there is no convenient relationship between the length of a written word and the duration of its utterance. The advantage is that the linguist or other consumer of the visualization has a rich and powerful set of mental skills which are based on written language. These skills outweigh anyone's vision of the near term capability of computers and can not be sacrificed to technological convenience. For now, we believe that approaches like ours provide an effective way to put current technology at the service of linguists.

Endnote 1

Most electronic corpora have detailed indexing schemes which are useful for citation as well as in support of linguistic analysis. For citations in our paper, we sketch these schemes here.

ICE

ICE citations have the form

```
ICE-<country>-<sample-type>-<sample-number>.
```

Thus, all of our samples are taken from the preliminary transcriptions of the American corpus (ICE-USA), are of type S1A, denoting face-to-face speech, and have three-digit sample numbers.

London-Lund

London-Lund citations have the form

```
LLC S.<sample-type>.<sample-number> <first-tone-unit>-<
last-tone-unit>
```

'S' indicates that this is speech. (All LLC samples are speech, but the citation form is derived from that of another corpus, the Survey of English Usage, which also has written samples). Our citations all have sample-type 1, denoting face-to-face conversations. Tone units are intonation groups demarcated in the text by integers.

Endnote 2

T	H	S1	S2	Sy
	One could / one could uh divide by six at once.			
Yeah				
		Exactly!		
			Ex	
If			act	
you had simpli same)				(I got the
fied by six at once,				
then the same res/	Shall I (wipe it			
	out)?			
No, leave it! Would've been (immediately) (instantly) the same result.				

Above is our rendering of the example rendered in the HIAT notation on page 123. We have broken overlap cells at phoneme boundaries, which makes our rendering harder to read than HIAT's. However, it is more precise because it is not subject to inaccuracies arising from words whose typographic rendering is long but whose spoken duration is short. Some linguistic inquiries might not require phonemic granularity. Although we have not implemented it, in such cases our software could automatically place the entire word in the cell containing the last phoneme. Finally, note that in re-rendering this example, we have made judgements about the precision of overlap and the turn boundaries which may not be unambiguous in the HIAT rendering.

References

Bloomfield, L. 1933. *Language.* New York: Holt, Rinehart and Winston.

Chomsky, N. 1957. *Syntactic Structures.* The Hague: Mouton.

Chomsky, N. 1975. *The Logical Structure of Linguistic Theory.* New York: Plenum.

Church, K.W. and M. Liberman. 1991. A Status Report on the ACL/DCL. *Using Corpora: Proceedings from the New OED Conference.* Waterloo, Ontario: The University of Waterloo Centre for the New OED and Text Research. 84-91.

Curme, G. 1947. *English Grammar.* New York: Harper and Row.

Du Bois, J. 1991. Transcription Design Principles for Spoken Discourse Research. *Pragmatics* 1:1, 71-106.

Du Bois, J. et al. 1993. Outline of Discourse Transcription. In Edwards and Lampert (1993). 45-89.

Edwards, J. 1992. Design Principles in the Transcription of Spoken Discourse. In Svartvik (1992). 129-144.

Edwards, J. and M. Lampert, eds. 1993. *Talking Data.* Hillsdale, New Jersey: Lawrence Erlbaum.

Ehlich, K. 1993. HIAT: A Transcription System for Discourse Data. In Edwards and Lampert (1993). 123-48.

Garside, R., G. Leech and G. Sampson. 1987. *The Computational Analysis of English.* London: Longman.

Greenbaum, S. 1992. A New Corpus of English: ICE. In Svartvik (1992). 171-9.

Hafez, O.H. 1991. Turn-Taking in Egyptian Arabic. *Journal of Pragmatics* 15, 59-81.

Jespersen, O. 1909-49. *A Modern English Grammar on Historical Principles.* Copenhagen: Munksgaard.

Kucera, H. and W.N. Francis. 1967. *Computational Analysis of Present-Day English.* Providence, Rhode Island: Brown University Press.

Morris, R., E. Blachman, and C. Meyer. 1994. A Constraint Based Editor for Linguistic Scholars. *Electronic Publishing* 6:4, 349-360.

Poutsma, H. 1926-9. *A Grammar of Late Modern English.* Groningen: Noordhoff.

Quinn, Akiva. 1993. An Object-oriented Design for a Corpus Utility Program. In Aarts, J., P. de Haan and N. Oostdijk, eds. *English language corpora: Design, analysis and exploitation.* Amsterdam: Rodopi, 215-225.

Quirk, R. 1992. On Corpus Principles and Designs. In Svartvik (1992). 457-69.

Sacks, H., E.A. Schegloff and G. Jefferson. 1974. A Simplest Systematics for the Organization of Turn-taking in Conversation. *Language* 50, 696-735.

Schenkein, J., ed. 1978. *Studies in the Organization of Conversational Interaction.* New York: Academic Press.

Shastri, S.V. 1988. The Kolhapur Corpus of Indian English and Work Done on Its Basis So Far. *ICAME Journal* 12,15-26.

Sperberg-McQueen, C.M. and L. Burnhard. 1994. *Guidelines for Electronic Text Encoding.* Available by Internet file transfer from sgm19.ex.ac.uk

Svartvik, J. and R. Quirk. 1980. *A Corpus of English Conversation.* Lund, Sweden: Lund University Press.

Svartvik, J., ed. 1992. *Directions in Corpus Linguistics.* Berlin: Mouton de Gruyter.

Sweet, H. 1891-98. *A New English Grammar.* Oxford: Clarendon Press.

Walker, D.E. 1992. Developing Computational Lexical Resources. In Kittay, E.F. and A. Lehrer, eds. *Frames, Fields, and Contrasts: New Essays in Semantic and Lexical Organization.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Acknowledgments

We are grateful to the Santa Barbara Corpus of Spoken American English for permission to use an unpublished sample, especially to its director, Jack Du Bois, for thoughtful comments on an early draft and to Rob Norris for providing us with the sample in machine readable form and helpful discussion to rationalize our notation with CSAE's. Any remaining inaccuracies are our own.

Gerald Grow is a professor in the School of Journalism, Media and Graphic Arts at Florida State University in Tallahassee, where he teaches courses on magazines. He is the author of many articles in magazines and academic journals, a popular book on camping in Florida and is a widely unrecognized cartoonist.

Division of Journalism
Florida Agricultural and
Mechanical University
Tallahassee, Florida 32307

Visible Language 28.2
Gerald Grow, 134-161

Rhode Island School of Design
Providence, Rhode Island 02903