10

The Production of Prosody[1]

A. Cutler and S. D. Isard   *University of Sussex*

I. Introduction

Prosody is the sauce of the sentence—it adds to, enhances or subtly changes the flavour of the original. And like a good sauce, the realization of a sentence's prosodic structure is a blend of different ingredients none of which can be separately identified in the final product. Thus it is rarely possible to say: syllable A is longer than syllable B simply and solely because syllable A bears lexical stress and syllable B does not; or: this particular fall in pitch is due exclusively to the presence of a clause boundary.

Accordingly, we will not attempt in this paper to describe separately the determinants of each component of the suprasegmental pattern i.e. the specific factors which lead to durational variation, pitch changes and variations in amplitude. Rather, we will concentrate on what we hold to be the major sources of prosodic effects, which can be grouped into four main categories: lexical stress patterns of individual words; the placement of sentence accent; syntactic structure; and a variety of pragmatic factors such as choice of speech act and attitudinal indicators, which influence the overall shape of the intonation contour.

We shall not attempt to relate the influence of these sources directly to numbers of milliseconds or precise changes in pitch. Instead we will try to describe their effect at an abstract prosodic level, whose units can then be realized as specific pitches and durations, rather in the way that the abstract phonological level is realized in surface phonetic form.

To describe the units of the abstract prosodic level, we shall draw on terminology that is more or less common among British writers on the subject. In particular we shall speak of "tone groups", intonational units realized as a major pitch movement, possibly preceded by a preparatory run-up and followed by a subsequent tailing off. Halliday (1967), Crystal (1975), and O'Connor and Arnold (1961) give classifications of the major pitch movements and of the run-ups and aftermaths that can go with each.

There are three main decisions that must be taken with respect to the tone-group structure of a sentence:

(i) Where the tone group boundaries will fall,

(ii) Which syllable within a tone group will have the major pitch movement associated with it,

(iii) What the major pitch movement will be.

(i) is determined largely by the syntactic structure of the sentence. One tone group per clause seems to be a sort of neutral, default case (see Halliday (1967) and Crystal (1975)).

(ii) is primarily influenced by placement of accent, while

(iii) is largely the product of pragmatic factors. It is possible, however, for the three sorts of influence to poach on one another's territory, as when the decision to emphasize two words in a single clause forces the clause to take two tone groups, or when the syntactic structure of a compound noun phrase dictates which syllable to accent in order to emphasize the noun phrase as a whole.

Consider, by way of illustration, the "that one sank" clause of (1)-(3) below.

(1) (My first boat blew up, and so I bought another and) that one *sank*.
(2) (I had a boat, but it sank, and so I bought another and) *that* one sank.
(3) (A: You mean your old boat sank? B: No, I bought another boat and) *that* one *sank*.

In (1) and (2) the clauses are covered by a single tone group. Placement of emphasis dictates that the major pitch movements occur on "sank" and "that" respectively. The type of pitch movement in both cases is a fall, appropriate to a simple contrast between the accented item and the corresponding item in the previous clause.

In (3) the choice to emphasize both "that" and "sank" leads to a two tone group clause. Again the contrasted item "that" is given a fall in pitch, but "sank", which is carried over from the question, and in a sense provides the setting within which "that" is contrasted, is given a pitch contour which first falls and then rises.

A fourth decision which we want to place at the abstract prosodic level is one which ultimately determines the lengths of the phonetic components which make up the utterance. Some writers, e.g. Klatt (1975) and Cooper (this volume), take the position that segmental durations such as vowel lengths are determined directly on the basis of syntax and the intrinsic nature of the phoneme concerned. Others, notably Abercrombie (1965), Halliday (1967) and Lehiste (1977) prefer to divide the utterance into "feet", stretching from one stressed syllable to the next. According to this scheme, the feet are assigned target lengths, which are shared out among the syllables composing the feet and then among the phonemes composing the syllables. Witten (1977) suggests one algorithm for achieving the sorts of feet discussed by Abercrombie (1965).

We shall take this second approach, and in particular we shall discuss the influence of syntactic structure on the placing of foot boundaries, rather than directly upon phoneme lengths.

It would, of course, be desirable to display a detailed model of the way in

which all of these decisions are taken. Unfortunately, if not surprisingly, we are not in a position to attempt such a feat. We do, however, have certain remarks to make about the role of each of the factors mentioned above and we shall devote one of the four following sections to each.

II. Lexical Stress

At some point during the sentence production process the words which will be uttered are looked up in the mental lexicon—that is to say, the appropriate phonological realization of each semantically specified unit is located. In each polysyllabic word one syllable is marked for heavier stress than the other(s), and the eventual phonological instantiation of the word will include this stress marking. There are several acoustic consequences of lexical stress: the stressed syllable can be longer in relative duration than the unstressed syllables, and may also be spoken with greater physical intensity. If sentence accent falls on a polysyllabic word, it is on the stressed syllable that the pitch movement associated with accent will be realized. The vowels in unstressed syllables may reduce to / /. Lehiste (1970) describes these effects in detail.

There is evidence that words are listed in the mental lexicon in a phonological form more abstract than the surface phonetic form. This evidence comes largely from the study of speech errors, particularly the classic work of Fromkin (1971, 1973). She observes, for instance, that consonant clusters involving the nasal [ ] and a stop may split into [n] plus stop, as in (4):

    (4) The ban will pake 5-6% interest
          (Target: The bank will pay . . .)

This suggests that underlying the surface form of bank [bæ k] is a more abstract form with the final cluster /nk/. Fromkin also observes that velar nasals may themselves split into two segments, [n] and [g]:

    (5) swin and swaig
          (Target: swing and sway)

which again suggests the psychological reality in speech production of an abstract /ng/ underlying [ ]. Many phonologists, notably Chomsky and Halle (1968), have argued that [ ] is in fact derived from such an underlying form.

Further support comes from a study of semantically unrelated word substitution errors, or malapropisms, by Fay and Cutler (1977). In this study the target (word which was intended) and error (word which was uttered) pairs were found to be very similar in sound, particularly in their initial segments. It was hypothesized that malapropisms arise when the speech production device picks from the mental lexicon, instead of the word it is seeking, one of that word's near neighbours; close neighbours in the lexicon are very similar in sound because the lexicon is primarily arranged by phonological similarity. At the point (counting from left to right) at which a particular target and error departed from identity, the two words were in general very similar in distinctive feature marking, which Fay and Cutler took as a suggestion that

"phonological similarity" in the mental lexicon is defined in terms of distinctive features.

There were however some exceptions to this generalization, and among these was a group of errors in which at the point of departure from identity one of the two words contained a vowel whereas the other contained the glide [y], e.g. *movie* (Target: *music*), *genuine* (Target: *general*), *musicians* (Target: *magicians*). Chomsky and Halle (1968) argue that such words as *music* do not contain the segment /y/ in their underlying representation, but that it is inserted by rule into the surface form. Fay and Cutler pointed out that a comparison between the underlying representations of such pairs would involve the distinctive feature difference between two vowels rather than between a vowel and a glide and hence would result in more similar feature marking which in turn would reduce the discrepancy between this small group of errors and the major body of the malapropism corpus.

If we assume a mental lexicon in which the phonological representations of the listed words are in a form more abstract than the surface form, then a question arises as to the representation of lexical stress patterns. Chomsky and Halle (1968) have formulated rules which derive the surface stress patterns of English words. The analogous rules which derive the surface phonetic form of a word from an underlying abstract form appear to be actively involved in the speech production process; is lexical stress also determined mechanically by rule in production?

The alternative to application of the stress rules in production is a listing, in the mental lexicon, of the appropriate stress marking as a part of the phonological representation of each word.

Evidence from speech errors involving erroneous placement of lexical stress appears to support the latter proposal. Typical lexical stress errors include (6)-(9):

    (6)   Now the paradigm involves presenting—presenting . . .

    (7)   I've got my book so we don't have any conflicts.
        (Target: conflicts)

    (8)   Everyone knows that economists—that economists . . .

    (9)   I need the number of the Psychol—Psychological Corporation.

In (6), the verb *present* has mistakenly been stressed on the first syllable, i.e. has the stress pattern of the noun *present;* in (7) the reverse is the case: *conflict* (N) has been stressed as if it were *conflict* (V); in (8) *economists* bears stress on the third syllable, suggesting the adjective *economic (al);* in (9) the stress has moved from the third to the second syllable, which is where it falls in *psychologist* and *psychology.*

These correspondences are not atypical; all lexical stress errors exhibit them. The erroneously produced stress pattern is always that of another word (thus no such errors as * *administrative* are observed); and this word is always morphologically related to the intended word. As a consequence of this, lexical stress errors only ever occur in morphologically complex, or derived, words (thus an error such as **window* does not occur). A more extensive analysis of this type of error may be found in Cutler (1979); see also Fromkin (1977).

It will be argued that these errors arise as a result of confusion between two differently stressed forms in the mental lexicon and that they provide evidence for the inclusion of stress marking in the lexical entry. Alternative explanations are unable to account for the regularities which these errors exhibit. Thus the suggestion that they are blends between two alternative candidates for utterance founders on the observation that all known types of blends involve two words of the same grammatical category (or two equivalent constituents) whereas lexical stress errors always involve two words of different grammatical categories. The possibility that they might result from simple exchange or shift of stress features cannot in any way account for the failure of non-derived words to show such errors or for the constraint that the errors always give the stress pattern of a morphological relative of the intended word. The same is true of the proposal that misplaced lexical stress is a consequence of confusion between words at the articulatory program level. And, importantly, it is also the case that misapplication of stress rules could not account for these regularities; the stress rules apply to derived and nonderived words alike, and it is unclear why errors should occur in one instance but not in the other, and unclear why misapplied rules should always assign stress to a syllable which does bear it in some related word.

An explanation which accounts for all the features shown by lexical stress errors is the following: words derived from a common base are stored together in a single lexical entry, with, inter alia, each word's stress pattern being specified as part of its representation. Lexical stress errors arise as a result of confusion within the lexical entry—the stress pattern selected is not that of the intended word but of some other member of the entry. Thus the pattern is always that of a related word since the error occurs within the common lexical entry; and stress errors only ever appear in derived words because only derived words share a lexical entry with other words—non-derived words have private lexical entries.

Thus the characteristics of lexical stress errors argue against the mechanical application of stress assignment rules as part of the speech production process. A question which is not however resolved by the available data is whether the specification of stress internal to the lexical entry is in terms of stress features marked for each syllable, or whether the appropriate rule for each word is appended to its phonological representation. These two proposals may in fact be indistinguishable with respect to their realization in the type of lexical stress error which occurs. However, it is not clear that application of the stress rule appropriate for, say, an adjective to a noun derived from the same base might actually lead to the stress falling on that syllable which bears it in the adjective. The Chomsky-Halle stress rules invoke as a major determinant of lexical stress the number of syllables in the word as well as such factors as the phonemes which terminate the word, i.e. whether the word ends with a vowel, a consonant or a consonant cluster. Related nouns and verbs and adjectives very often differ on exactly these characteristics, the exception being such noun-verb pairs as *object, conflict* etc. If only pairs of this latter type were involved in lexical stress errors, we might

have no basis for feeling that the specification of the appropriate stress *rule* as part of the lexical entry was a less satisfactory description than the specification of the stress *features*. However, most of the lexical stress errors involve word pairs with different endings and different number of syllables (e.g. examples (8) and (9) above). In these cases it is at least a reasonable supposition that the rules would not, when applied to, for example, the shorter form, result in stress falling on that syllable which bears it in the longer word. These considerations are by no means conclusive, but they suffice to render preferable to us an account of stress specification in the lexical entry which includes the actual listing of the stress marking, or stress features, for each syllable.

It may be felt that our conclusion that stress is listed in the lexicon and *not* determined by mechanical application of stress assignment rules is in conflict with recent linguistic evidence concerning the psychological reality of these rules. Nessly (1974), for example, has shown that English speakers can, with great reliability, decide upon the appropriate pronunciation (including stress pattern) of invented "words" which they have never seen before, and that the stress patterns they choose are in accord with English stress assignment rules. This, not in itself surprising, finding indicates that native speakers at some level know the principles upon which lexical stress assignment in their language functions. But "psychological reality" of stress assignment rules does not imply at all that the rules are applied every time a polysyllabic word is uttered. Nor is it necessarily "inefficient" or "uneconomical" to postulate a system which includes both knowledge of the rules and stress marking of individual words. Computational efficiency, to which end the system is presumably designed, involves a trade-off between the speed with which operations can,be performed on the one hand, and the size of the system, i.e. the storage space it takes up, on the other. Thus it is not unreasonable to suppose that the gain in speed of operation resulting from stress specification within the lexical entry is sufficient to justify any increase in storage demands which result from it. The internalized stress rules are as a consequence made redundant in normal production, but cannot be jettisoned since they are needed to cope with new words, names, or nonsense, i.e. with any item which does not already have an entry in the mental lexicon.

III. Accent

Just as within a polysyllabic word one syllable has greater prominence than the others, so within an utterance of more than one word greater prominence is given to one word than to others. The syllable on which sentence accent falls is the syllable which bears the lexical stress of the accented word, so there is a sense in which we can think of lexical stress as embodying the potential for sentence accent. In this section we will consider the way in which the placement of sentence accent is determined during the production of a sentence.

Perhaps the simplest proposal (for implementation in a production model) is that accent placement is a function of syntactic structure. This, in skeleton form, is the claim made by Chomsky and Halle (1968), and by others, for example Bresnan (1971). It is not important for the present discussion to describe in detail the rules which they propose for accent placement. Two facts about such systems, however, are very important. The first is that accent placement by rule is determined with reference to syntactic structure alone, i.e. without reference to semantic or pragmatic considerations. The second point follows from the first; it is clear that syntactically driven rules will apply only one accent pattern to a particular syntactic structure, but it is immediately obvious that for any sentence there are many options for accent placement. Any of the seven words in (10) for instance, *could* bear the primary sentence accent:

(10)   They don't grow bananas in Northern England.

Thus it is a necessary characteristic of systems such as that of Chomsky and Halle that they claim that each sentence has a "normal" or "neutral" accent placement, i.e. the placement which is determined by the syntactically driven rules. Alternative placements are all special cases—in the Chomsky/Halle system non-neutral accent placement is described as expressing *contrast* on the accented word with another word, or constituent, in the sentence, in another sentence, or implied by the context.

Appealingly simple as the syntactic proposal is, it does not appear appropriate to incorporate it in a speech production model. Rules such as Chomsky and Halle's constitute a procedure taking only syntactic information as input to produce an accent pattern. It is not clear how, or whether, the rules can be integrated into a larger system that takes semantic and pragmatic factors into account.

Furthermore, considerable linguistic effort has recently been devoted to demonstrating that a syntactic model does not correctly describe the placement of sentence accent in neutral cases. In the vanguard of the attack have been Bolinger (e.g. 1972) and Schmerling (1974, 1976). Criticism has been directed at both of the above-mentioned aspects of the Chomsky-Halle position, namely that accent placement can be determined by syntax alone, and that each sentence has one "neutral" accent pattern.

Schmerling (1976) provides a compelling illustration that for even a very simple two-word sentence the accent placement is determined by contextual factors. In the year in which two ex-presidents of the United States died, Schmerling reports, she was informed of their respective deaths in an interestingly different manner. Harry S. Truman died after a long illness which was extensively reported in the media, and Schmerling was informed of his death in the following words:

(11)   Truman died.

Lyndon Johnson, on the other hand, died of a sudden heart attack; Schmerling heard of this as follows:

(12)   Johnson died.

Schmerling is undoubtedly correct in her claim that the differing accent placement in these two sentence resulted from the differing contexts. Truman

was known to be sick; the new information in (11) concerns his death, so the verb is accented. In (12) the new information is that something happened to Johnson, to whom nothing in particular was expected to happen.

Another telling counter-example to the syntactic determination of accent was provided by Ladd (1978). It was first pointed out by Newman (1946) that the class of sentences of the type of (13)

(13)   I have plans to write

customarily takes sentence accent on the noun when the noun is the direct object of the verb ("I must write plans"), but on the final verb when that verb is a complement to the noun ("I plan to write"). Ladd (1978)†, however, devised contexts for a version of this construction in which the syntactic factors determining accent placement were completely over-ridden by contextual factors. These examples are worth quoting in full. In (14) the verb-as-complement reading is appropriate despite accent on the noun:

(14)   a. George had no idea he was supposed to follow Helen.

       b. Whaddya mean—Helen left directions for George to follow!

and in (15) the noun-as-direct-object reading is appropriate even though accent falls on the verb:

(15)   a. George feels pretty bad about ruining dinner, but the package had no directions.

       b. Gee—didn't you look over on the counter by the toaster? Helen left directions for George to follow!

       (Ladd, 1978: pp. 138-139)

Likewise, many counter-examples have been offered to the normal/contrastive distinction. It has been observed that certain sentences appear to have only a "contrastive" accent pattern and no "normal" pattern:

(16)   Even a child can build it!

Speakers unanimously place the accent in (16) on "child", although the Chomsky-Halle sentence accent rules would deem that a contrastive rather than the normal placement for this sentence.

Similarly, many emphatic accent placements do not appear to contrast with anything; it seems far-fetched to claim that (17) expresses a contrast with, for example, (18):

(17)   There's no way I'm going to go along with that!

(18)   *There's some way I'm going to go along with that!

Furthermore, it is often the case that accent is assigned to a word for no other reason than to avoid placing it on some other word. Consider the (b) sentences in (19) and (20) below:

(19)   a. John's gone to North Dakota to study the mating habits of the native linguist.

       b. But there aren't any linguists in North Dakota!

(20)   a. I'd hate to be a dentist.

       b. Me too—I'm sure glad there are people who want to be dentists.

       (from Ladd, 1978)

†Counter-examples were also provided by Bolinger (1958, 1972) and by Berman and Szamosi (1972).

It is clear that the accenting of *in* and *be* respectively does not arise from a contrast with another preposition or another verb, but from the fact that the words which might otherwise have been accented have been used in the previous utterance. Accenting them might therefore give an impression of redundancy. Thus the accent has been moved away from the repeated words, and has ended up on words which do not themselves in this instance have any reason for claiming prominence. In other words, the accent pattern results not from *accenting* of the emphasized word but from *de-accenting* of others. An excellent discussion of the phenomenon of deaccenting is given by Ladd (1978). Some recent examples from the authors' own experience are (21) and (22):

> (21)   I didn't read any newspapers all the time I was writing my Ph.D.—but that was because I was in Austin, Texas, where there aren't any newspapers to read.

> (22)   If you'd like to gather to the little bath we'll look at that next.

(22) was spoken by a tour-guide giving a commentary on a series of Roman baths. The "little bath" itself had not been previously mentioned; many other baths had been. In fact it is not necessary for a de-accented word to have actually occurred in the preceding context; it can be implied by it:

> (23)   a. Where'll we have dinner—the Dim Symptom?
>   b. Shirley won't eat Chinese food.

Conversely, de-accenting can be precipitated by preceding use of the de-accented word in a different sense:

> (24)   He's so sharp he's even called Sharp.†

Sometimes de-accenting can lead to an accent pattern which could in a different context have resulted from accenting, or emphasis; ‡ compare, for example, the indignant utterance (26) as spoken by an adult in reply to (27) or by a child in reply to (28):

> (26)   I was reading the book!

> (27)   I put away that book and the other stuff on the table.

> (28)   Did oo have fun playing with the bookie-wookie?

In the former instance the accent on *reading* results from de-accenting *of book*, in the latter it results from focus on new information, since the act of reading has not, in (28), been taken for granted as what one does with books.

We will not attempt to develop here an original and detailed case in favour of the determination of sentence accent by semantic and pragmatic (contextual) factors rather than by syntax; for the complete arguments we refer the reader to the authors we have cited. We would however like to utter a cautionary word or two. The arguments against the syntactic position have often amounted to outright rejection of its concepts. Thus Schmerling (1974) claims that the notion of "neutral" accent is quite useless; any accent placement embodies presuppositions about the discourse context. Ladd (1978) argues that there is no such thing as contrastive accent—only

---

†Note that there are limits to the indirectness:
  (25) * She's so sharp she's even called Cutler.
‡ A nice example of this is given by Ladd (1978; p. 117).

differences in focus as a result of reference to various aspects of the context. We believe that both of these rejections are a little too sweeping. Let us take first the case of contrastive accent. It should be obvious from the above discussion that we agree with Ladd and other critics of the syntactic position that "contrastive" is not an appropriate catchall term for any semantically placed accent. Nevertheless, that contrast exists apart from focus can be demonstrated by manoeuvering it into a sentence along with focus. Thus in (29b), several words are focussed; some of them are contrastively accented, some not.

> (29) a. London's the capital of Scotland, isn't it?
>   b. No, Edinburgh's the capital of Scotland, London's the capital of England. †

The *kind* of accent which falls on those words which occurred in (29a) (London, Scotland) differs from the kind of accent falling on the words expressing new information (Edinburgh, England). The former bear a fall-rise accent, the latter a simple falling accent (see our discussion of examples(1)-(3) in the introduction to this paper). All four words are accented, or, in Ladd's terms, focussed; but those which also express a contrast have a falling accent, those which don't express contrast bear a fall-rise. (The types of accent are of course reversed if (29b) serves as an answer to:

> (30)   Edinburgh's the capital of England, isn't it?)

That is to say, when several items in a sentence are focussed, those which are contrasted can be distinguished intonationally from those which are not; the accentual system is richer than would appear from a description, for example Ladd's, which seeks to subsume contrast under the general rubric of focus.

Now let us consider the possibility of a role for "neutral" accent placement. In some sentences semantic and pragmatic reasons do not pick out a particular word for sentence accent. In (31), for example, a contrast is drawn between two constituents as wholes:

> (31)   I was not surprised to hear Susan was mad about old movies, but that she didn't like Chinese food surprised me.

Movies and *food* are accented not because they are contrasted with other particular items (e.g. *clothes, men*), but simply because they are in each case the rightmost items in the constituent; they bear accent on behalf of their constituents, so to speak.

This kind of default accent is, we believe, the one sense in which the notion of neutral accent can be justified. The way in which accent is placed in the default case was formulated in detail by Newman (1946); as re-stated by Schmerling (1976) the principle is:

> (32)   Given a sequence of stresses which are equal and greater than other stresses within the intonational unit, the last such stress will be more prominent than the others.

(By "intonational unit" Newman referred to what we have called "tone group".)

†We are grateful to Christopher Longuet-Higgins for this example. More detailed remarks on contrast can be found in Isard (1978).

The principle can be seen in operation in a sentence such as (33), from Chomsky (1971) and Jackendoff (1972):

(33)   Was he warned to look out for an ex-convict in a red shirt?

The accent on *shirt* might represent contrast with some other word, in which case (34) might be a good reply, or simply accent on the constituent as a whole, in which case any of (35)—(37) would be acceptable:

(34)   No, an ex-convict in a red hat.
(35)   No, an automobile salesman.
(36)   No, an ex-convict in a dinner suit.
(37)   No, the FBI.

Shifting the accent in (33) to, for example, "red" or to "convict", renders it no longer neutral, and (35)—(37) no longer appropriate replies.

Ladd (1978) gives a comprehensive account of default accent, which he describes as focus on the entire constituent. We would prefer to shift the emphasis slightly and call it focus on *what* the constituent *denotes,* rather than, for instance, on *how* it denotes it. In (31), for example, although *movies* and *food* are not contrasted with other comparable words, there *is* a contrast which is being drawn, namely between that attribute of Susan's which did not surprise the speaker and the other attribute which did. The hearer is intended to appreciate this contrast without paying particular attention to any of the individual words in each embedded sentence. So the speaker accents the embedded sentences, using however the neutral or default accent.

A complete account of sentence accent placement, therefore, includes neutral accent (in a minority of cases rather than the majority which syntactically motivated accent rules would claim to account for). It also includes contrast, and focus, as well as accentuation of a particular word achieved by de-accenting some other word. We will shortly discuss the order in which these various factors exercise their effects in sentence production. First, however, we will round out the picture we have given of above-word-level stress with a few remarks on compound stress.

"Blackbird" expresses a meaning different from "black bird" and "English professor" from "English professor". The former of each pair, more heavily stressed on the initial element, is a compound noun. Others, for example Chomsky and Halle (1968), have given thorough accounts of compound stress patterns. What is important for our present discussion is that compounds act as a unit in the competition for accent. However complex a compound, whether the initial element is for instance itself a compound, as in (38), or is an Adj-N sequence, as in (39), the compound as a whole has one and only one most prominent syllable:

(38)   Chemistry research laboratory.
(39)   Indo-European syntax text.

When the neutral or default accent falls on a compound as the rightmost element of a constituent, then, it accents that syllable which is the most prominent in the compound (just as accent falling on a word is realized in that syllable which bears lexical stress). Thus the (a) sentences of (40) and (41) bear a possible neutral accent; the (b) sentences are acceptable replies.

(40)   a. Was he working at a chemistry research laboratory?
       b. No, at a garbage dump.
(41)   a. Did you buy an Indo-European syntax text?
       b. No, a flowerpot.

Compound stress is determined at an early level in the production process, prior to the placement of sentence accent. Some compounds (e.g. *blackbird)* are presumably lexicalized, and their stress patterns would accordingly be retrieved along with their phonemic specification from the mental lexicon. Others would be constructed when the search for a lexical unit to express a complex meaning (e.g. "Indo-European syntax text") failed to come up with an entry (e.g. "Lehmann"), and various components of the complex meaning had to be separately retrieved and combined to form it. By the time accent placement operates, in any event, compound nouns enter into the calculation essentially as single words.

Thus the sentence production operation which determines which of all the syllables in a sentence will be the most prominent is carried out on a string of words which (a) if polysyllabic have their stressed syllable marked, and (b) are grouped if appropriate into compounds with the most prominent syllable in the compound marked. Semantic and pragmatic factors then determine accent placement to express focus, contrast or deaccentuation (we know of no evidence that these are separate, ordered operations).

It is interesting to note that the semantic/pragmatic factors can over-ride the earlier assignments of syllable prominence. For instance it is quite possible to pick out a single component of a compound for contrast:

(42)   No, of course I didn't say Indo-Germanic syntax text, I said Indo-
       European syntax text.

This is hardly surprising, since contrast can even over-ride lexical stress, as in the well-known example:

(43)   This whisky wasn't exported, it was deported.

Similarly, de-accenting can also result in accent falling on some other syllable in a compound than the usually stressed one; the stress shifts (possibly by the operation of some form *of* the default principle) to the rightmost nearest word, i.e. to the next word to the right, e.g. (44), or to the next to the left if nothing not de-accented remains on the right, e.g. (45):

(44)   I thought you said you knew nothing at all about syntax—so how
       come you've got an Indo-European syntax text?

(45)   I thought you said you owned no syntax texts at all—that there's an
       Indo-European syntax text.

The only remaining component of the accent assignment operations is the neutral or default accent which applies last of all, and indeed only applies if the semantically motivated accent placement has left something for it to apply to—a string of words marked for focus or contrast as a whole but containing within it several equally prominent lexical (or compound) stresses. The default principle makes the rightmost of these most prominent and the operation of accent placement is therewith complete.

Correct description of the determination of sentence accent, we have seen, is

not a simple matter; it incorporates reference not only to the semantics of the message, but also to the structure of the sentence and the discourse context. The language production model, therefore, must also allow for these differing inputs when accounting for accent placement.

IV. Isochrony and syntactic boundaries

Pike (1945) makes a distinction between what he calls syllable-timed languages, in which each syllable is of roughly equal length, and stress-timed languages, in which stressed syllables occur at roughly equal intervals. French is supposed to be an example of the former sort of language, and English of the latter. It is important to note that the stressed syllables in question here are not necessarily accented, but are generally just the syllables marked for stress in the lexicon. Syllables not so marked are also stressed for these purposes if they do receive accent (e.g. "no linguists in North Dakota"). The intervals from one stressed syllable to the next are termed feet.

Attempts to find isochronous (equally timed) feet by measuring wave forms of English speech have generally failed. A "tendency toward isochrony" is sometimes detected in studies where an extra unstressed syllable is inserted between two stressed syllables, and although the duration between the stressed syllables does go up, it goes up by less than the length of the extra syllable, the other syllables having been compressed to make up for its presence (Huggins, 1975; Fowler, 1977). As Fowler notes, this phenomenon may indicate that the speaker is trying to make the next beat occur as close as possible to "the right time", but it is not in itself evidence that the "right times" are evenly spaced.

Abercrombie (1965) explains some of the wider deviations from isochrony by positing "silent stresses", which are essentially skipped beats. He points out, among other things, that such silent stresses also occur in verse, where the existence of the regular beats themselves is less controversial. He cites

My / sire / ` is of a / noble / line /

from Coleridge's "Christabel", as well as

To / be or / not to be / ` / that is the / question.

Some may find even more compelling the extra, silent, beats that he postulates at the ends of the first two lines of the limerick form, as in

There / was a young / man from Cape / Horn /

Who / wished that he'd / never been / born /

Lehiste (1977) reviews the isochrony literature and comes to the conclusion that isochrony is at least in part a perceptual phenomenon. That is, we hear the times between stressed syllables as being more nearly equal than they really are. This conclusion is supported by recent work of Donovan and Darwin (1979).

Donovan and Darwin present subjects with sentences all of whose stressed syllables begin with the same phoneme, say /t/, as in

(46)   He turned up by ten talking of terrorism.

The subjects are then asked to adjust a series of clicks so as to make them occur with the same timing as the /t/s in the sentence. The subjects can hear the sentence and the clicks as often as they like, but they cannot hear them simultaneously. The subjects tend to space the clicks more evenly than the /t/s are spaced in the sentence, suggesting that they hear the /t/s to be spaced more evenly than they really are.

At this point we are faced with the possibility that hearers might simply impose an isochronous interpretation on anything that they take to be spoken English, and that even if there is a tendency toward isochrony in the physical signal, it is not necessary to the perception of isochrony. Lehiste (1977) rejects this idea on the grounds that differences in the length of interstress intervals can not only be perceived, but they can be used to convey linguistic distinctions. In particular, she claims that speakers use lengthened interstress intervals to mark syntactic boundaries, and that in order for this to be possible, speaker and hearer must both have some notion of an unlengthened interval to use as a standard.

Lehiste asked subjects to read aloud sentences in which ambiguous phrases like "old men and women" were embedded. The subjects were asked to read the sentence in two different ways, one in which the phrase was supposed to be grouped as "(old men) and women" and the other in which it was supposed to be grouped as "old (men and women)". Further groups of subjects listened to the sentences to make certain that they were perceived as intended.

When the length of the segment "men and women" was measured, it was found to be dramatically longer in the case where "(old men) and women" was the intended reading. Lehiste makes the proposal that this lengthened segment constitutes a deliberate disruption of isochrony, and that such disruptions are used to signal the presence of syntactic boundaries. "It is in this sense," she writes, "that isochrony is integrated into the grammar of English at the syntactic level" (Lehiste, 1977. p. 262).

Lehiste's theory still leaves us with a number of questions to consider. Perhaps the most immediate is whether it is possible to say anything further about the amount by which speakers will lengthen interstress intervals in order to achieve their purpose of marking syntactic boundaries. Will the amount of extra length be related to the length of a notional unlengthened interval in some systematic way, or will it perhaps increase in an otherwise unpredictable fashion with the "amount of emphasis" that the speaker wants to achieve? And if the relation is systematic, is it possible that the increased duration results not from a complete disruption of isochrony, but rather from "skipping a beat", so that the duration of longer intervals is twice that of shorter ones, but an underlying rhythm is maintained?

Such a possibility is at least consistent with data presented in Lehiste (1973), where subjects were asked to disambiguate sentences of the form "Steve or Sam and Bob will come". The distance between "Steve" and "Sam" was roughly twice as great in the case where the intended reading was "Steve or (Sam and Bob)", and similarly the distance between "Sam" and "Bob" was approximately twice as great when "(Steve or Sam) and Bob" was intended.

In her report of the study involving the "old men and women" examples, Lehiste gives only the comparative lengths for the entire segment "men and women", and not the lengths of individual feet. We have run a small study on a set of similar sentences and measured the lengths of the feet. The sentences used were

(47) I'm allergic to ripe marrows, melons and cucumbers.

(48) We bought expensive brandy, port and cigars.

(49) He sells used cars, bikes and trailers.

In Lehiste's sentences, the ambiguity to be resolved was whether the adjective, "old", applied to just one item, "men", or two, "men and women". We chose sentences where the adjective could apply to three items instead of just two, because we thought that in these cases we might detect a tendency to restore isochrony by lengthening all three items, and not just the first.

We asked each of five subjects to read each sentence in both possible ways, e.g. in the case of the first sentence to give one reading corresponding to "ripe (marrows, melons and cucumbers)" and another corresponding to "(ripe marrows), melons and cucumbers". We then measured two interstress intervals for each sentence: one from the vowel onset of the stressed syllable of the first listed item to the corresponding point in the second item, and then from there to the third listed item. In the case of the first sentence this means taking the time from the beginning of the first vowel of "marrows" to the beginning of the first vowel *of* "melons", and similarly from "melons" to "cucumbers". There are other proposals for the way in which interstress intervals should be measured (see, for example, Morton *et al.,* 1976), but they would not give very different results for our purposes, and these measurements are relatively straightforward to make.

Measurements were made on digitized wave forms sampled at a rate of 8000 Hz on a PDP-12 computer. We used a wave form editing program written by C. J. Darwin which displays a wave form on a screen and allows one to insert pointers into the wave form. The distance between pointers can be measured to within 0T ms. Each measurement was performed twice, and discrepancies were well within 5 ms, so we feel reasonably confident in claiming accuracy to within 20 ms.

Our measurements showed that in the sentences where the adjective was meant to apply to all three nouns, e.g. "used (cars, bikes and trailers)", the subjects made the two feet we measured nearly equal. The average ratios of the first foot to the second foot in these sentences are shown in Table I.

When the subjects read the sentences with the other meaning intended, the first foot was considerably lengthened, as Lehiste's theory would predict. The average ratios of the lengthened first feet to the original first feet (e.g. "cars" in "(used cars), etc." to "cars" in "used (cars, etc.)") are given in Table II. The increase very nearly amounts to a doubling of the original length.

If we consider the ratios of first feet to second feet in the sentences where first feet are lengthened, the averages are again in the neighbourhood of 2 (see Table III). However these averages conceal what appear to be systematic differences among the individual subjects. For instance, one *of* the subjects

TABLE I

Mean ratio across subjects of first measured foot to second measured foot when adjective was meant to apply to all nouns

Sentence 1	("ripe marrows, etc.")	0 885
Sentence 2	("expensive brandy, etc.")	1 004
Sentence 3	("used cars, etc.")	1 071

TABLE II

Mean ratio across subjects of first measured foot when adjective was meant to apply to first noun only to same foot when adjective was meant to apply to all nouns.

Sentence 1	("ripe marrows, etc.")	2 021
Sentence 2	("expensive brandy, etc.")	2 010
Sentence 3	("used cars, etc.")	1 974

TABLE III

Mean ratio across subjects of first measured foot to second measured foot when adjective was meant to apply to first noun only.

Sentence 1	("ripe marrows, etc.")	1 857
Sentence 2	("expensive brandy, etc.")	1 812
Sentence 3	("used cars, etc.")	1 724

Mel, had ratios of 2 126, 2 246, and 2 629, consistently above 2, while another, Derek, had ratios of 1 087, 1 355 and 0 856, much nearer to 1.

Our intuitive impression from listening to the tapes *of* the subjects' utterances is that there is a tradeoff between the use of comparatively longer first feet, and the use of intonation. That is, Derek's pitch changes were much more marked than Mel's.

If speakers of English do in fact make such a trade-off, it provides further justification for an abstract level of prosodic groupings, where different speakers would have in common the intention of marking off a syntactic unit by assigning it a grouping of its own, and would then diverge as to the way in which the presence of this grouping would be signalled, in one case by pitch movement and in another by adjusting the timing.

We can note in this connection that while Lehiste wants to use disruption of isochrony as a way of setting off syntactic units, Crystal (1975: pp. 16-21) suggests rules for doing the same job with intonation, adjusting the domains of pitch movements in order to show which words should be grouped together. Neither set of rules takes account of the phenomena on which the other is based. However, Halliday (1967) proposes that the tone group should be viewed both as an intonational unit, and as a rhythmic unit consisting of a

number of feet. If we adopt this proposal, we can postulate that the syntactic units in question are given different tone groups, and then that the tone groups may be distinguished by pronounced intonation contours, or by a change of timing.

In Section III we rejected the notion that the syntactic component of sentence production generally determines accent placement. The syntactic component nevertheless exercises a considerable effect on the prosodic structure of the sentence by specifying the tone group divisions. This information may then be realized either as durational or as pitch variations.

V. Holistic Contours

At the Chicago Linguistic Society meeting in 1974 Mark Liberman and Ivan Sag produced an amusing demonstration that an intonation (fundamental frequency) contour could by itself convey a certain amount of meaning. The contour in question was the one borne, for example, by the incredulous reply (50b):

(50) a. I've got elephantiasis, I'm gonna die.

b. Elephantiasis isn't incurable!

In reply to the question:

(51) Ivan, would you mind dropping my pet whale off at the aquarium on your way to school today?

the contour alone was performed on the kazoo; it was clear to the audience that the reply was an indignant objection to the request. Liberman and Sag called this intonational pattern the "contradiction contour". In a subsequent paper (Sag and Liberman, 1975) they isolated other holistic contours which they also associated with particular meanings.

The notion that intonation contours have intrinsic meaning has been espoused by a number of linguists, and, indeed, has been postulated in a much stronger version than that claimed by Liberman and Sag. Pike (1945; p. 20), for instance, stated: "Many intonation contours are explicit in meaning. Whenever a certain sequence of relative pitches is heard, one concludes that the speaker means certain things over and above the specific meanings of the words themselves. A change of pitch contour will change the meaning of the sentence."

In this strong version, the contours-have-meaning claim poses some interesting possibilities for a language production model. It would be relatively simple to incorporate into such a model an intonational lexicon in which contours were paired with their fixed meanings, and to divide the meaning of an utterance into that part to be conveyed by the words and

†The drawn contour represents the variation of fundamental frequency against time (allowing for some inaccuracy due to the mismatch between acoustic duration and orthography) on a Kay SonaGraph spectrogram of the utterance.

another part to be conveyed by the intonation contour. This latter part could then be looked up in the intonational lexicon and the appropriate contour retrieved in the same way that looking up meanings in the word lexicon results in the phonetic forms of words being retrieved.

Once again, however, we find that an apparently simple and appealing model which seems to be suitable for incorporation into a production model turns out not to account for the prosodic facts. The problem is that for such a proposal to work the meanings assigned to the contours must be, to a certain extent at least, supra-contextual; for each contour there must be an element of common meaning (or, if the contour is ambiguous, a finite set of meanings) which can be observed in every instance of the contour in use. It can be demonstrated that this is not the case.

Liberman and Sag's "contradiction contour", for example, seems to express above all impatience in the following context:

> (52)   Father (to son who has been ignoring a friend's attempt to attract attention from outside the window):

> Go and see what the fellow wants!

To say that what is common to the contexts of (50), (51) and (52) is, for instance, that the speaker disapproves of his audience's attitude, is to fail to do justice to the richness of the effects of the contour in each context.

Even in Liberman and Sag's own contexts the effects are richer than can be captured by the general term "contradiction". Liberman and Sag point out that Ivan's answer to Mark's request in (51) could be a wide range of quite different utterances, e.g.:

> (53)   a. I'm not having that smelly beast in my car.
>   b. You don't have a pet whale.
>   c. You know it wouldn't fit in my VW.
>   and so on.

They also point out that certain other utterances don't work so well—"require some fairly unnatural assumptions in order to be construed as contradictions", in their words (p. 422)—for example:

> (54)   a. I'm not very fond of that animal.
>   b. I'm more than happy to take him along.

But there are also whole classes of contradictions which couldn't take their contour; for instance, it only "fits" on declaratives and some imperatives, not on questions so that (55 a and b) would sound very odd if intoned in that manner:

> (55)   a. Why on earth should I do that?
>   b. How could I ever fit it into the VW?

Moreover, it is possible to make minor changes to those sentences which will fit, e.g. (53 a-c), which render them unsuitable for the contradiction contour without in the least altering their contradictory import:

> (56)   a. I'm just not having that smelly beast in my car.
>   b. You don't even have a pet whale.

c. You know it just wouldn't fit in my VW.

Similarly, replacing (53c) with (57) does not change the speaker's message but does make the "contradiction contour" inappropriate:

(57)   It wouldn't lit in my VW as you very well know.

A more complicated objection seems to be that the semantics of the utterance are more closely constrained than Liberman and Sag realized; not only does the speaker have to object strongly to the request, but he has to state the reason for his objection (as in 53 a-c). (58 a and b), objections without a reason given, do not work:

(58)   a. That's the most outrageous thing I've ever been asked.

b. I wouldn't do that for anything.

Oddly enough, there is an alternative contour which seems to express contradiction in this context and which is appropriate for all of (53), (55), (56) and (58) as well as (54) and (57):

(59)

You know I never give lifts to whales.

The acoutic characteristics of this contour are a high initial section followed by a slight rise and rapid fall to a low, flat terminal section. The positioning of the fall depends on the positioning of sentence accent in the utterance (on "have" in 53, on "fond" in 54a, on "any" in 58b, for example). We do not by any means wish to claim, however, that this contour is a synonym of Liberman and Sag's contradiction contour; it is too easy to think up contexts in which only the contour of (59) is appropriate (e.g. 55-58) or in which the two have markedly different effects.

(60)   a. Now I see why we couldn't do it.

b. Now I see why we couldn't do it.

(60a), for example, expresses sudden enlightenment—Aha!—and, importantly, suggests that the speaker is about to amplify, to reveal the reason he has just discovered; (60) on the other hand suggests irritation, perhaps because the speaker has had to repeat the utterance.

The point of this extended set of examples is simply that intonational meaning is contextually constrained. The effect of a particular contour differs with the context in which it occurs, and different contours can, depending on context, have similar or radically different effects. Not even the simplest of intonational effects is free from contextual variation. Take, for example, a pair of contours described by Sag and Liberman (1975). They point out that (61) can be either a suggestion or a genuine question, whereas (62) can only be a genuine question:

(61)   Why don't you move to California':

(62)   Why don't you move to California?

One of the functions of the contour of (62), they claim, is to "freeze" the utterance into a literal interpretation and rule out the indirect speech act of suggestion which can be carried by many questions (e.g. just about any question beginning "Why don't you . . ."). This effect again turns out to be dependent on the utterance itself and its context. For instance, (63) is a suggestion with either contour:

(63)   Why don't you go away?

With the contour of (62) it is a more direct suggestion, in fact, and certainly a more offensive one, than with the contour of (61).

A more detailed treatment of the context-dependence of the intonational effects described by Liberman and Sag is given by Cutler (1977).

It is, unfortunately, a negative kind of argument that we have made in this section; we have been concerned to show only that the claim that intonation contours have fixed meanings, or even constant pragmatic effects (e.g. on the literalness of a question), does not hold up. The effect of a particular contour is strongly constrained by the utterance which carries it and by the context in which this utterance is spoken. Accordingly, the choice of contour must be made with reference to contextual factors. We now turn our attention to the incorporation of contour selection into a language production model.

Gazdar (this volume) has demonstrated that the number of language production decisions which are affected by pragmatic, or contextual, factors is very large indeed. Selection of an appropriate intonation contour is, indeed, one of the phenomena he has cited. We believe that the choice of contour can be compared in complexity to the choice of a particular syntactic structure; both are pragmatically determined decisions between a restricted number of alternative ways of expressing the intended message. In other words, the speaker chooses between, say, the contours of (50) and (59) with reference to a particular message-in-context in much the same way that he chooses between the active and the passive voice. It is superfluous to point out that very little is known so far of the way in which such decisions are carried out. The implication for a production model, however, is that the output of the contour selection component is not uniquely determined by the output of other components (syntactic, semantic) but, like the syntactic and the semantic components, makes reference to the discourse context of the utterance.

Among the tasks of the contour selector we have discussed so far only one, choice of a holistic pitch contour; but it is likely that certain other effects have their origin at essentially the same point in sentence production. One of these is ironic tone of voice (Cutler, 1974), the sneering way of saying a sentence so as to give it a conveyed meaning which is the reverse of the literal meaning. A

speaker's decision to say, e.g. (64) in an ironic manner, rather than saying (65), is determined by pragmatic factors.

(64) Looks like it's going to be a really groovy party.

(65) I'm afraid this party is going to be rather boring.

Another such phenomenon is sentence tempo; to give just one instance, an effect of imminent threat can be achieved by slowing down the utterance in an exaggerated manner:

(66) Were—you—thinking—of—hitting—that—child?

Finally, we should point out the obvious fact that the output of the contour selector is presumably in an abstract form; its eventual realization in speech may be modified by other aspects of the prosodic structure. Contrastive accent may, for example, result in a pitch peak falling on the accented word without altering the effect of the contour, for example:

(67) a. Everyone's moving to California, I'll be all alone.

        b. Why don't you move to California?

In the "pure" form of this contour the fall would occur on the final syllable, as in (61), but (67b), as (61), is a suggestion rather than a genuine question. Liberman and Sag (1974) also point out that their "contradiction" contour coexists with shifts in sentence accent. Similarly, tempo is independent of accent placement (accent could plausibly fall in (66) on *were, you, hitting, child,* for example), as is ironic tone of voice:

(68) Looks like that's going to be a really groovy party.

Syntactic boundary placement, of course, determines the length of the constituent over which the contour is extended, and the lexical stress patterns of the chosen lexical items determine the syllables upon which the pitch movements demanded by a contour will be carried out. Selection of a holistic contour is, however, independent of these factors; interaction occurs at the point at which the output of the many components of the language production process are phonetically realized.

VI. Conclusion

We cannot claim to have given a complete description of the production of prosody. We have, for instance, not considered the phonetic realization of prosodic effects, but have confined our discussion to a more abstract level. Moreover, we have treated our four sources of prosodic variance as independent, although it is clear that at the level at which they are realized in the utterance they must interact.

Nevertheless, we feel that the four main divisions of this paper reflect the major decisions affecting the prosodic shape of a sentence. It will be apparent that we do not believe that all prosodic effects simply "fall out" of decisions taken, during the course of sentence production, about, for example, lexical items and syntactic structure. We feel that at certain stages during the

production of an utterance specifically prosodic decisions are taken. This conclusion has obvious consequences for the formulation of a model of language production. While on the one hand lexical stress patterns are determined by the output of the lexical component, and tone group boundaries are at least in part set by the choice of syntactic structure, accent placement and selection of intonation contour are decisions which have exclusively prosodic results. The production model should incorporate both an accent determination and a contour selection component.

All parts of the production process can make errors, including those parts which have major prosodic effects; many different types of prosodic error are described by Cutler (in press). Consideration of prosodic errors can materially influence the way we model language production; an example of this is the argument in Section II of this Chapter. Furthermore, the manner in which prosody interacts with errors of all types can prove highly instructive, as many speech error researchers have noted. For example, it has often been remarked (e.g. Meringer and Mayer, 1895; Boomer and Laver, 1968; Garrett, 1975) that exchanges of elements below the level of the word preserve lexical stress: stressed vowels and syllables exchange with each other, unstressed vowels and syllables likewise, but stressed do not exchange with unstressed. This regularity forces the assumption that such exchanges take place at a level at which the utterance is divided into feet. Prosodic characteristics, in other words, assist in identifying the level at which a particular error arises. They can also assist in classifying particular errors. Cutler (in press) cites the error:

(69)   Do you talk on the telephone with which ear?

which, as the drawn contour indicates, was spoken with the intonation appropriate for a yes-no question rather than for a wh-question, suggesting that it may have arisen as a result of a blend with an alternative yes-no question rather than simply as a wh-question which got its word order mixed up. Similarly, inferences about detection of one's own errors can be drawn from prosodic characteristics; the speaker of (69), for example, can be assumed not to have detected the anticipation error *words* as it occurred:

(70)   Notice that these are the only two words that apply above the word level.
       (Target: the only two rules . . .)

This conclusion arises from the fact that the second, intended, occurrence of the word *words* bears sentence accent, whereas, as the discussion in Section III made clear, the second occurrence of a given lexical item in a sentence is normally deaccented. We can assume that the erroneous earlier occurrence of *words* was not available to the accent placement system.

Sentence accent in fact interacts in a very interesting way with word shift and exchange errors. When two lexical items exchange places, the accent structure of the sentence customarily remains unchanged, e.g.:

(71)   We have a laboratory in our computer.
       (Target: we have a computer in our laboratory)
       from Fromkin, 1973.
(72)   Something funny smells!
       (Target: something smells funny)

Fromkin (1971) drew on this regularity in constructing an early model of language production based on speech error data—a model which, unusually among such models, attempted to account specifically for prosody as well as the rest of the sentence. She suggested on the basis of the exchange error findings that accent placement might be determined by the syntactic and semantic structure of the utterance prior to the retrieval of words from the lexicon. The rationale for this ordering was that it would allow words to be inserted from the lexicon into the wrong slot in a syntactic frame, i.e. into a slot marked for accent rather than into a slot which was not so marked, or vice versa.

There are several reasons for preferring, instead of this early model of Fromkin's, an account in which lexical look-up precedes accent placement. For one thing, the de-accenting process can make reference to prior occurrence of particular words rather than meanings. For another, the word on which accent falls is sometimes determined by the particular lexical item chosen, as when a choice is made between a simple verb and a synonymous verb-particle combination:†

(75)   a. John promised to house the visitors, but not to feed them.
       b. John promised to put the visitors up, but not to feed them.

We would therefore not agree that errors such as (71) and (72) necessarily imply that sentence accent must be placed before lexical lookup. The regularity which Fromkin noted about such errors is, however, particularly interesting, since it allows us to draw a contrast with errors such as (76-78), in which the words which have exchanged places are not lexical words but members of the vocabulary's closed class (e.g. prepositions, pronouns, etc.):‡

(76)   Can I turn off this?
       (Target: Can I turn this off?)
(77)   Well I much would have preferred the owl.
       (Target: I would have much . . .)
(78)   It's useful so that they don't know how far they in are.
       (Target: . . . how far in they are)

As may be seen, the accent in such errors falls on the particular word, not on the slot, which should have borne it in the intended utterance. (Further examples are given by Cutler (1979).) The prosodic characteristics of exchange

---

† Evidence that choices of this nature are made at the lexical level is provided by lexical blends (assumed to occur when two synonyms are available in the lexicon) between simple verbs and verb-particle combinations, e.g.:

(73)   I just snabbed it! (From Garrett. 1975; explained as a blend *of snap up* and *nab*—the reference is to finding a bargain while shopping.)
(74)   Aren't you going to telephone her up? *(telephone* and *call up:* from Fromkin, 1973).

‡ See Bradley (1978) for evidence that closed class words do not form part of the main lexicon.

errors therefore provide strong support for the contention of Garrett (1975) that lexical words and closed class words are involved in fundamentally different types of error at different levels of the production process.

The moral we wish to draw is this: not only is the production of sentence prosody intrinsically interesting, but attention to prosodic phenomena can prove of great value with many differing aspects of the design of a language production model. Although we have not attempted in this paper so rash a project as the construction of such a model, we hope that our remarks have shed some light on how it should be done.

References

Abercrombie, D. (1965). Syllable quantity and enclitics in English. *In* "Studies in Phonetics and Linguistics," Oxford University Press, London.

Berman, A. and Szamosi, M. (1972). Observations on sentential stress. *Language,* 48, 304-325.

Bolinger, D. (1958). Stress and information. *American Speech,* 33, 5-20.

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language,* 48, 633-644.

Boomer, D. S. and Laver, J. D. M. (1968). Slips of the tongue. *British Journal of Disorders of Communication,* 3, 1-12.

Bradley, D. C. (1978). *Computational Analysis of Vocabulary Type.* Ph.D. dissertation, M.I.T.

Bresnan, J. (1971). Sentence stress and syntactic transformations. *Language,* 47, 257-280.

Chomsky, N. (1971). Deep structure, surface structure and semantic interpretation. *In* "Semantics" (D. D. Steinberg and L. A. Jakobovits, eds). Cambridge University Press, London.

Chomsky, N. and Halle, M. (1968). "The Sound Pattern of English". Harper and Row, New York.

Crystal, D. (1975). "The English Tone of Voice." Edward Arnold, London.

Cutler, A. (1974). On saying what you mean without meaning what you say. *Papers from the Tenth Regional Meeting, Chicago Linguistic Society,* pp. 117-127.

Cutler, A. (1977). The context-dependence of "intonational meanings". *Papers from the Thirteenth Regional Meeting, Chicago Linguistic Society,* pp. 104-115.

Cutler, A. (1979). Errors of stress and intonation. *In* "Errors of Linguistic Performance: Slips of the Tongue, Ear, Pen and Hands" (V. A. Fromkin, ed.). Academic Press, New York.

Donovan, A. and Darwin, C. J. (1979). The perceived rhythm of speech. Ninth International Congress of Phonetic Sciences, Copenhagen, 1979.

Fay, D. A. and Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry,* 8, 505-520.

Fowler, C. A. (1977). *Timing Control in Speech Production.* Ph.D. dissertation, University of Connecticut.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language 41,* 27-52.

Fromkin, V. A. (ed.) (1973). "Speech Errors as Linguistic Evidence." Mouton, The Hague.

Fromkin, V. A. (1977). Putting the emPHAsis on the wrong syLABle. *In* "Studies in Stress and Accent" (L. M. Hyman, ed.). University of Southern California, Los Angeles.

Garrett, M. F. (1975). The analysis of sentence production. *In* "Psychology of Learning and Motivation", vol. 9 (G. Bovver, ed.). Academic Press. New York.

Halliday, M. S. (1967). "Intonation and Grammar in British English". Mouton, The Hague.

Huggins, A. W. F. (1975). On isochrony and syntax. *In* "Auditory Analysis and Perception of Speech" (G. Fant and M. A. A. Tatham, eds). Academic Press, London.

Isard, S. D. (1978). A note on contrastive stress. *Pragmatics Microfiche,* 3, 2-G1.

Jackendoff, R. S. (1972). "Semantic Interpretation in Generative Grammar". MIT Press, Cambridge, Mass.

Klatt, D. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics* 3, 129-140.

Ladd, D. R. (1978). *The Structure of Intonational Meaning.* Ph.D. dissertation, Cornell University.

Lehiste, I. (1970). "Suprasegmentals". MIT Press, Cambridge, Mass.

Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America* 54, 1228-1234.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics,* 5, 253-263.

Lehmann, W. P. (1974). "Proto-Indo-European Syntax". University of Texas Press, Austin, Texas.

Liberman, M. and Sag, I. (1974). Prosodic form and discourse function. *Papers from the Tenth Regional Meeting, Chicago Linguistic Society,* pp. 416—427.

Meringer, R. and Mayer, K. (1895). "Versprechen und Verlesen: eine psychologisch—linguistische Studie". Goschen, Stuttgart. Re-issue, (A. Cutler and D. A. Fay, eds). John Benjamins, Amsterdam, 1978.

Morton, J., Marcus, S. and Frankish, C. (1976). Perceptual centers (P-centers). *Psychological Review,* 83, 405-408.

Nessly, L. (1974). *English Stress and Synchronic Descriptions.* Ph.D. dissertation, University of Michigan.

Newman, S. S. (1946). On the stress system of English. *Word,* 2, 171-187.

O'Connor, J. D. and Arnold, G. F. (1961). "Intonation of Colloquial English", Longmans, London,

Pike, K. L. (1945). "The Intonation of American English". University of Michigan Press, Ann Arbor.

Sag, I. and Liberman, M. (1975). The intonational disambiguation of indirect speech acts. *Papers from the Eleventh Regional Meeting, Chicago Linguistic Society,* pp. 487-497.

Schmerling, S. F. (1976). "Aspects of English Sentence Stress." University of Texas Press, Austin, Texas.

Schmerling, S. F. (1974). A re-examination of "normal stress". *Language,* 50, 66-73.

Witten, I. H. (1977). A flexible scheme for assigning timing and pitch to synthetic speech. *Language and Speech,* 20, 240-260.