Chapter 10
Generating Phonetic Plans for Connected Speech

The generation of connected speech involves more than the mere concatenation of word forms retrieved from memory. Words participate in the larger gesture of the utterance as a whole, and the speaker's phonetic plan expresses this participation in myriad ways.

There are, first, morphological and segmental accommodations of various sorts. A speaker will choose allomorphs that are tuned to the context. In chapter 8 auxiliary reduction was given as an example. Speakers normally prefer *I've bought it* over *I have bought it*, and *he'll go* over *he will go*. They may also cliticize other elements to neighboring words. Small words such as *to* and *of* are reduced and cliticized under certain conditions, as in *I wanna go* or *a bottle'o milk*. Segments may get lost, changed, or added at word boundaries, as in *jus fine* for *just fine* and *got [tʃ]ou* for *got you*. This often goes with resyllabification at word boundaries. In short, the syllable plans retrieved in connected speech often do not conform to the syllabification of the individual words' citation forms. This is because it is a main function of phonological encoding to prepare for fluent connected articulation. Long strings of spelled-out "citation" forms must be translated into fluently pronounceable strings of syllables.

Second, there is the speaker's prosodic planning. Words participate in the overall metrical structure of the utterance; they are grouped in smaller or larger rhythmic phrases. This phrasal togetherness is realized by the manipulation of the loudness, the duration, and the pitch of successive syllables in the utterance, and by the insertion of pauses. The speaker will, in particular, chunk his running speech in intonational phrases, which are the domain for the assignment of pitch contours. In the speaker's phonetic plan, words participate in this melodic line, creating peaks or troughs when they carry pitch accent. The melodic line is, in addition, expressive of attitude and emotion over and above the propositional meaning expressed in the utterance.

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

The present chapter will review how prosodic plans for connected speech are generated by the speaker, and how these affect the generation of word form. These two aspects of phonetic planning are closely interwoven. It is, for instance, impossible to generate a metrical structure for an utterance as a whole without having access to the syllabicity of the constituent words. In turn, however, the computed metrical parameters for the overall utterance must eventually be realized in the phonetic spellout of the individual words. There is a back-and-forth between stages of word-form spellout and stages of prosodic planning. But very little is known about the processes involved in the phonological encoding of connected speech.

The chapter will begin with a rough sketch of a possible architecture underlying the generation of connected speech (section 10.1). A *Prosody Generator* figures rather centrally in this architecture. It produces incrementally, and in close interaction with word-form spellout, the metrical and intonational parameters of an utterance. These are, we will suppose, eventually fed to the phonetic spellout procedures. After this global sketch of the architecture we will turn to a more detailed treatment of the Prosody Generator and of its metrical and its intonational planning (sections 10.2 and 10.3, respectively). Section 10.4 will discuss how the Prosody Generator affects the processes of segmental and phonetic spellout—in particular, how it mediates in the syllabification and the segmental accommodation of words in connected speech.

10.1 A Sketch of the Planning Architecture

10.1.1 Processing Components

In the "blueprint for the speaker" (figure 1.1), the box labeled "phonological encoding" represents a processor that is supposed to generate phonetic plans for connected speech. Let us begin by filling that box with some further details, as in figure 10.1.

The main input to phonological encoding is the unfolding surface structure. First, its terminal nodes with their diacritical parameters are pointers to word-form addresses. The previous chapter outlined how these word forms are retrieved from memory and transformed into phonetic plans. The main steps in this process—morphological/metrical spellout, segmental spellout, and phonetic spellout, are depicted on the left side of figure 10.1. Second, the surface phrase structure plays an important role in the generation of phonetic plans for connected speech. It is the main input to the *Prosody Generator*—a processing component that computes, among other things, the metrical and intonational properties of the utterance.

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

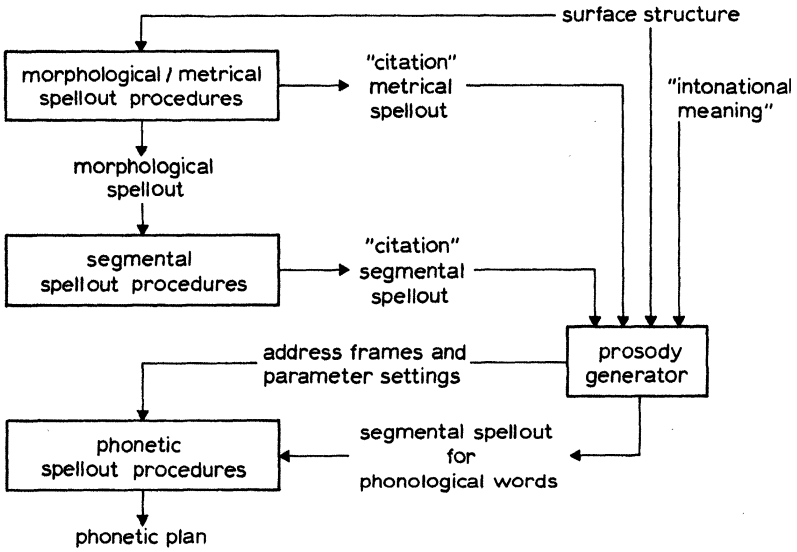


Figure 10.1
An outline of the architecture for the phonological encoding of connected speech.

Additional input to the Prosody Generator is what we called "intonational meaning" in chapter 8: rhetorical intentions, emotions, and attitudes. They cause the Prosody Generator to select certain tunes and tones, and to set key and register. Further essential input is the metrical spellout—i.e., each word's number of syllabic peaks, the location of the one that carries word accent, and the word's diacritical pitch-accent feature (if any). This, in combination with the relevant phrase-structural information, suffices to compute a metrical grid (see subsection 8.2.2) for the utterance, as well as a pitch contour.

Some aspects of phonological encoding are under executive control. Intonational meaning may perhaps be considered a form of executive control. The speaker can also, within bounds, freely insert pauses and vary the rate of speech. Speech rate is an important factor in phonological encoding. Not only does it affect the size of phonological and intonational phrases; it also has consequences at the segmental and phonetic levels, as we will see. When speech is fast, phonetic spellout is affected across the board. In the previous chapter we saw that Dell (1986) made the fruitful assumption that speaking rate does not affect the speed of spreading activation; it only affects the number of syllable frames to be filled per second. More generally, we will assume that the rate parameter sets the

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

speed of frame production at all levels of processing in phonological encoding. In the terms activation-spreading: It determines how long a "current" node stays current.

Let us now turn to the output of the Prosody Generator. It was argued in the previous chapter that the address frames of phonetic spellout are triggered by metrical spellout, each peak initiating the construction of a syllable frame. This can now be further specified. Figure 10.1 expresses the assumption that the string of peaks is channeled through the Prosody Generator. The phonetic spellout procedure subsequently receives a highly enriched signal. It specifies for each successive syllable frame its duration, its loudness, and its contribution to the pitch contour. It also inserts pauses. Phonetic spellout, then, is not the mere retrieval of a stored phonetic syllable plan; it is also a parametrization of that plan in terms of duration, loudness, and pitch movement.

There is reason to assume that prosody generation also affects segmental spellout. Phrasal boundaries can become determinants of how a word's syllables will be spelled out. A good example is the French phenomenon of *liaison*.* There are many French words in which the final consonant is, as a rule, not pronounced (e.g.,*trè(s), cour(t), peti(t)*). But this final consonant may fail to delete when the next word in connected speech begins with a vowel (as in *très intelligent, court ajournement, petit enfant*). This bears some resemblance to what happens with the word *an* in English. As a rule, its final /n/ is not pronounced, except if the next word begins with a vowel (*a ball, an animal, an intelligent animal*). So far this just shows that the segmental spellout of a word's final syllable may be dependent on the onset of the next word. This in itself is an important property of connected speech; it shows that segmental spellout is not merely a word-internal process but can be dependent on context. The case of liaison demonstrates, in addition, that there are phrasal conditions on this context-dependency. The following examples from Kaisse 1985 will illustrate this:

très intelligent
(very intelligent)

trè(s) intelligent et modeste
(very intelligent and modest)

In the first of these examples there is normal liaison; the /s/ is pronounced. In the second, however, there is no liaison if *très* modifies the whole conjunction *intelligent et modeste*, i.e., if the phrasal composition is (*très*

* For an experimental study of liaison see Zwanenburg, Ouweneel, and Levelt 1977.

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

(*intelligent et modeste*)). The phrase boundary between *très* and *intelligent* apparently blocks liaison in this case.* The precise formulation of the phrasal relations that allow or block liaison is not at issue here (see Kaisse 1985 for a detailed analysis), but only the fact that phrasal relations between subsequent words can, on occasion, affect their segmental spellout. It should, furthermore, be noticed that liaison (or, for that matter, /n/-deletion in *an*) is, within broad limits, independent of speaking rate. They are general phenomena of connected speech.

In summary: The Prosody Generator computes phrasal conditions relevant to segmental spellout, as well as a range of prosodic parameters for phonetic spellout. Before turning to these activities of the Prosody Generator in sections 10.2 through 10.4, we will consider where some further phenomena of connected speech are generated in the framework of figure 10.1.

10.1.2 Casual Speech

It is essential to distinguish between phenomena of connected speech, of casual speech, and of fast speech (Kaisse 1985). Connected speech need neither be casual nor fast. There are general properties of connected speech that arise independent of its speed or its formality. French liaison is such a case, and so are many of the metrical and intonational phenomena to be discussed in this chapter.

Casual speech differs from formal speech, but it need by no means be fast. There is slow casual speech, just as there is fast formal speech. Casual speech is a *register*,† a variety of the language, which may have characteristic syntactic, lexical, and phonological properties. A speaker may or may not have several such registers at his or her disposal ("motherese" and "telegraphic speech" are two examples). In casual talk the speaker is biased not only toward using a particular subset of his lexicon (e.g., *cop* rather than *policeman*) but also toward using particular allomorphs (e.g., *I've* rather than *I have*). We will assume that not only the former lexical choices but also the latter allomorphic ones are made during grammatical encoding. In other words, they are indicated at the level of surface structure, and hence are contained in the input to phonological encoding. The surface structure's terminal nodes point to the intended "casual" forms.

---

* But one wonders what Kaisse's data base for this claim was. Native speakers of French seem to have difficulties with the particular example.

† This use of the term *register* should not be confused with the notion of pitch level discussed in chapter 8.

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

Still, some aspects of casual speech have their origins in phonological encoding itself. When a speaker says *lea'me alone*, he is not using a casual allomorph of *leave*, but rather deleting a syllable-final consonant. This is, no doubt, a casual-speech phenomenon that arises at the level of word-form planning. We will, however, assume that phonological phenomena of casual speech are only *indirectly* caused by the casual register. The casual register allows for much faster speech than the formal register, and it is only at high rates that forms like *lea'me alone* arise. The proximal cause of such phonological phenomena is therefore rate, not casuality of register. Hence, the present chapter should concentrate on general phenomena of connected speech and of fast speech only.

10.1.3 Fast Speech

Among the most prominent properties of fast speech are *reduction* and *assimilation*. Reductions can arise at different levels of processing. A speaker can increase his rate of communication by generating short messages, by using a telegraphic register, and/or by accessing reduced (casual) allomorphs. All this is planned above the level of phonological encoding.

A speaker can also gain speed by reducing small unaccented words, such as pronouns and prepositions: *Give'm attention*, *Think o'money*. Such reductions may originate at the level of segmental spellout, following some very general rules of reduced spellout. These rules are sensitive to the character of immediately adjacent elements (e.g., *a bottle o'milk* but not *a bottle o'applejuice*) and to prosodic phrase structure (e.g., *Think o'money* but not *What are you thinking o'? Money?*).

Speed can also be gained by reducing segments across the board. A speaker can, for instance, reduce all word-initial unstressed vowels, as in *p'tato* or *t'mato* (Zwicky 1972). This depends neither on adjacent words nor on phrase structure. There are, however, restrictions on the kinds of new clusters that arise. For instance, *m'ternal* and *r'member* are unlikely reductions (Zwicky 1972; Kaisse 1985). These restrictions are, however, not phonotactic in the sense that only well-formed onset clusters are allowed; /pt/ and /tm/ are ill formed as syllable onsets in English. It is likely, therefore, that this kind of reduction takes place *after* the syllable plans are addressed. The correct syllables ([pə], [tə]) are addressed, but these unstressed syllables are given such minimal settings for their duration and loudness parameters that their vowels just about disappear in articulation. Why such extreme minimal settings are not possible for [mə] or [ri] is unclear, but it probably has to do with the high-sonorous onsets of these syllables.

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

To sum up: There can be reduction at all three spellout levels. The surface-structure input can induce morphological spellout to address reduced allomorphs, segmental spellout can generate reduced forms following general structure-dependent rules, and phonetic spellout can be subject to extreme parameter settings for the duration and loudness of a syllable.

Assimilation is a quite general phenomenon in connected speech, but it spreads wider in fast speech than in slow speech. It involves the change of some segment under the influence of another one, and the change makes the two speech sounds more similar. The phrase *ten books*, for example, is pronounced as [tɛm bʊks], where /n/ assimilates to the adjacent /b/ by adopting its bilabiality feature. Most forms of assimilation are to be located at the segmental spellout level. There can be substantial structure dependency in assimilation, as is testified by the much studied case of *wanna* (the assimilation of *want* and *to*). Assimilation is possible in *Who do you want to (wanna) succeed*, but not in *Who do you want to succeed you?* (Dogil 1984). (See also example 9 below.) The dependency on phrasal relations, however, is quite different for different kinds of assimilation. This will be taken up again in section 10.4.

Assimilation should be distinguished from *coarticulation*. Adjacent speech sounds interact because of the physiology and the mechanics of articulation. These interactions become more intense at higher speech rates; they depend on the time allotted to the articulation of syllables. These parameters are set at the phonetic spellout level. Coarticulation, therefore, occurs at the same level as the vowel reduction in *p'tato*, discussed above.

Reduction and assimilation often combine with *cliticization*, which is also widespread in connected speech but which is especially prominent when talk is fast. Cliticization consists of adjoining reduced materials to immediately adjacent words. The assimilation *wanna* above is such a case. Cliticization is necessary where a reduced morpheme has no syllabicity of its own. The allomorph *-ve* in *I've* and *you've* is nonsyllabic; it generates no peak in metrical spellout. It cannot be a free-standing word. It must, therefore, attach to the preceding word during segmental spellout. The result is a *phonological word*. The domain of segmental spellout is the phonological word (Nespor and Vogel 1986). The Phrase *I've* is not spelled out for *I* and *have* and subsequently reduced. It is, rather, the phonological word *I've* that is spelled out (as /aɪv/). By and large, each word form pointed to by the surface structure is a phonological word, with nonsyllabic allomorphs as a main exception. In addition, new phonological words can arise during phonological encoding, with cliticization as a main case. The Prosody Generator computes these new phonological words on the basis of

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

phrasal configurations in surface structure and syllabicity information from metrical spellout. It can also decide to further reduce and adjoin elements, even if they are both syllabic. The above assimilation *wanna* is such a case; it has become a single phonological word, and it is spelled out as such during segmental spellout.

These few remarks on fast speech have served to provisionally localize some major phenomena of fast speech, such as reduction, assimilation, and cliticization, in the framework of figure 10.1. More definite conclusions about the origins of these phenomena during fluent speech await thorough process analyses. Only when we understand the structure that controls the generation of connected speech will we be able to propose a more definite partitioning of the system.

10.1.4 Shifts

Another phenomenon to be localized in the scheme of figure 10.1 was already touched upon in chapter 7 in the discussion of Garrett's (1982) example:

(1) Did you stay up late vEry last night?

Garrett called this kind of speech error, where a word jumps over one or two adjacent ones, a *shift*. Garrett's view of shifts, with which I concur, is that they are caused not at the level of grammatical encoding but during phonological encoding. Shifts ignore the syntactic-category constraints that are so characteristic of word exchanges. The interchanged words in example 1, *very* and *late*, are of different syntactic categories, and so are *it* and *making* in the following example (from Stemberger 1985a):

(2) We tried it mAking ... mAking it with gravy.

Here it is a closed-class word (*it*) that is anticipated; an open-class one is jumped over. Garrett observed that closed-class elements (such as pronouns, prepositions, and articles) predominate in shifts, and this also holds for Stemberger's data.

Where can shifts be located in the framework of figure 10.1? My suggestion is in the transition from surface structure to morphological/metrical spellout. As surface structure is incrementally produced, its terminal pointers become available "from left to right." They will be fillers for successive address frames, and each completed address will occasion the retrieval of an item's morphological/metrical form information. Even if the surface structure's terminal elements are generated in impeccable order, a "later" element may happen to be spelled out more rapidly than an earlier element.

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

One would expect such anticipations especially in the case of highly frequent words, whose forms are more easily accessed than the forms of rare words. Stemberger (1985a) confirmed this prediction. It should, in addition, be noted that closed-class words are very high in frequency, which accounts for their preponderance in shifts.

One consequence of order reversals in morphological/metrical spellout is that there is a concomitant reversal of order in the metrical patterns received by the Prosody Generator. This predicts that a shifted word's pitch accent will stick to it, and that is what one observes in examples 1 and 2. Garrett (1982), Cutler (1980a), and Stemberger (1985a) observed and confirmed this property of shifts. It contrasts with errors in grammatical encoding, which, as we saw in subsection 7.1.2, usually show stranding of pitch accent.

Misordering at the level of morphological/metrical spellout can probably also account for affix shifts, as in the following (from Garrett 1982):

(3) I had forgot about*en* that

Here the affix of *forgot-ten* jumped over *about*, to which it attached. Presumably, the two morphological spellout procedures for *forgotten* and *about* ran more or less in parallel. The morpheme *about* became available just after the stem *forgot*, and just before its suffix *ten*. If this account is anywhere near correct, however, it is still surprising that such sublexical shift errors are not much more frequent in fluent speech.

This completes the initial sketch of the phonological encoding architecture underlying the generation of connected speech. The following three sections will deal more specifically with the workings of the Prosody Generator.

10.2 The Generation of Rhythm

The rhythm of connected speech appears in the alternation of more or less stressed syllables and the insertion of pauses. There are several ways in which a speaker can stress a syllable. One is to make it louder than neighboring syllables, another is to stretch it in time, and still another is to give it an accenting pitch movement. Though independently variable, these three tend to go together. It makes sense, therefore, to begin by considering the generation of rhythm at a fairly abstract level, namely as the generation of a pattern of stresses and pauses. This abstract pattern was called *metrical structure* in chapter 8.

Let us recapitulate what kinds of metrical structure are built by the speaker. There are, first, the words, with their internal stress patterns. The

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

basic stress patterns are retrieved from memory during metrical spellout. One of a word's syllables is marked for word accent; it will attract pitch accent, if there is to be any. In addition to retrieved words, the Prosody Generator will have to deal with other phonological words created by cliticization. Second, there are phonological phrases to be built. They can be seen either as "absolute" prosodic units—successive stretches of speech leading up to lexical (nonpronominal) heads-of-phrase (Nespor and Vogel 1986)—or as "relative" units leading up to stronger or weaker break options (Selkirk 1984a). Third, there are intonational phrases. One could say that they run from one *actual* prosodic break (i.e., a *taken* break option) to the next. These phrases are the domain for the assignment of meaningful pitch contours. There is, finally, the utterance as a whole, which may have utterance-initial or utterance-final metrical properties relevant to turn-taking, such as anacruses or utterance-final lengthenings.

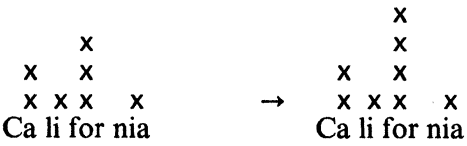
A main principle for a processing theory of rhythm is that, at all these levels, production should take place *incrementally*. This means that the metrical pattern should be created as surface phrase structure and morphological/metrical spellout become available. The Prosody Generator should not buffer large amounts of input in order to make current decisions dependent on later materials. It should, rather, be able to work with very little lookahead. In the following we will successively consider the metrical planning of (phonological) words, of phonological phrases, and of intonational phrases from this incremental point of view. We will then turn to aspects of timing, i.e., the duration of segments and syllables in the contexts of words, phrases, sentences, and larger units, and finally to the issue of isochrony, i.e., the presumed regular temporal spacing of stressed syllables in the connected speech of "stress-timed" languages.

10.2.1 Phonological Words

In speakers of languages (such as English) that have limited lexical productivity, the basic or "citation" metrical pattern of most words is stored in the mental lexicon. This metrical pattern is one of the first features to be spelled out in word-form access. The Prosody Generator accepts it as the basis for further metrical processing. If the lexical pointer to the form address has the diacritical feature "pitch accent", this information is also transmitted to the Prosody Generator. A first task, then, is to translate this information as an extra beat on the peak that carries word accent. This procedure is exemplified in the following, which depicts the pitch accenting of *California*:

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

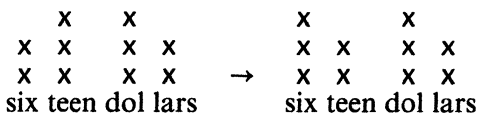
(4) **Basic metrical grid** **Pitch-accented grid**



In this example the word *California* is written out for convenience only; the adjustment of the metrical pattern can be made before the individual syllables have been spelled out.

In subsection 10.2.2 I will argue that this accenting operation needs no lookahead whatsoever—i.e., that it can be done incrementally, as new words are metrically spelled out. This is slightly different for operations that prevent stress clashes. In chapter 8 the example phrase *abstract art* was given, which is pronounced with alternating stress—*Abstract Art*—in spite of the fact that the "citation" accentuation of the adjective is *abstrAct*. A "beat movement" (Selkirk 1984a) prevents a stress clash between two adjoining syllables. The same type of beat movement can be observed in the phrase *sixteen dollars*. The stored accent pattern for the constituent words are *sixtEEn* and *dOllars*, but in the phrase there is alternating stress: *sIxteen dOllars*. This movement operation is depicted in the following:

(5) **Beat movement** (i) (ii)



Beat movement does require some minimal amount of lookahead. It is of two sorts. First, since the condition for the shift is a threatening succession of two stressed syllables in subsequent words, the metrical pattern of the second word must be at hand in order to effectuate a beat movement in the first one. This requires that the Prosody Generator minimally buffer the metrical patterns of two subsequent words (but see subsection 10.2.2). Of course, it does not always do so. But if it doesn't (because of a high speech rate, or for some other reason), there will be no beat movement. Second, there are phrasal restrictions on beat movement. If the utterance to be developed is

(6) Dimes I have sixteen, dollars just one

there will be no beat movement. It is prevented by the phrase boundary following *sixteen*. This shows that the input for the beat-movement operation is not only the metrical structure of two consecutive words but also phrasal information. The latter information is quite local in nature, how-

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

ever. The only relevant feature is whether or not there is a phrasal boundary following the first word of the pair (and if so, of what kind it is).

The Prosody Generator can also create new phonological words by cliticization. Unstressed closed-class words are easily cliticized to adjoining open-class words, and the tendency to cliticize increases with the rate of speech. The sentence *They have it* will normally be uttered with *it* cliticized to the lexical head of phrase *have*—i.e., with *have-it* as a single phonological word. The phonological word is the domain of syllabification. The phonetic spellout for *have-it* will not consist of the syllables [hæv] and [ɪt], but of [hæ] and [vɪt].

Can cliticizations be incrementally generated? That is, can the Prosody Generator produce these phonological words without lookahead? Let us begin with the dominant case of *enclitics* (cases where the "little" word follows the "big" word to which it adjoins). The above *have-it* is an example. In English, most enclitics derive from unstressed monosyllabic closed-class words, particularly pronouns, auxiliaries, and particles. But there are two cases to be distinguished:

(i) Cliticization is, of course, obligatory or necessary when the "little" element is nonsyllabic. This is generally the case for cliticized auxiliaries. It was suggested above that auxiliary forms such as *'ve* and *'ll* are allomorphs of the full forms. These allomorphs are already referred to by appropriate lexical pointers in surface structure, and are directly addressed at the level of morphological/metrical spellout. The metrical pattern of such an element is empty, since these morphemes have no syllabic peak. The only thing to be done by the Prosody Generator is to add the empty element to the previous nonempty one. Having access to the developing surface-structural information, the Prosody Generator recognizes the empty element as the metrical realization of a particular lexical pointer. It is in this way that phonological words such as *I've* and *you'll* arise initially. At the next stage, the Prosody Generator must occasion the correct segmental spellouts of these prosodic words. The recognition of such nonsyllabic elements, and the subsequent decision to adjoin them to their predecessors, is obviously a completely local affair; there is no relevant "later" information. This conclusion leaves unimpeded the possibility that, during grammatical encoding, the choice of such an allomorph might depend on the following syntactic context. Compare, for instance, the sentences 7 and 8, which are derived from an example in Pullum and Zwicky 1988:

(7) I know where it's located

(8) I know where it is

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

In sentence 7, the nonsyllabic allomorph is generated. However, sentence 8 requires the full form; *I know where it's* is ill formed. This difference suggests that later context can be relevant for the choice of allomorph.*

(ii) When the "little" word is syllabic and can stand alone as a phonological word, the Prosody Generator can still cliticize it to the foregoing "big" word. A much-studied case is the infinitival particle *to*. It is easily cliticized in a sentence like the following:

(9) Who do you want to see?
/wɒntə/
/wɒnə/

Here the new phonological word *want-to* is formed, which becomes spelled-out as /wɒn-tə/ and, by further reduction, as /wɒnə/. This kind of *to*-cliticization is quite general (consider *ought-to*, *used-to*, and *supposed-to*, all of which show resyllabification, testifying to their status as phonological words). Still, the Prosody Generator cannot leave it at completely local decisions, i.e., decisions involving only the pair X + *to*. This is apparent from example 10, where the same pair, *want to*, cannot be adjoined:

(10) Who do you want to see this memo?

Examples 9 and 10 are from the work of Pullum and Zwicky (1988), who present a concise review of the extensive literature on *to*-contraction. The upshot of this literature is that there are phrase-structural conditions on the cliticization of infinitival *to*. The Prosody Generator must refer to surface structure in order to decide whether these conditions are fulfilled. However, the main question for our present purposes is whether these are local conditions or whether they can involve much later parts of surface structure. Pullum and Zwicky conclude their review of the evidence with the statement that "a very small portion of the surface syntactic context, local in terms of both adjacency and bracketing . . . is relevant for the determination of whether a given word sequence can have the contracted pronunciation." In other words, little lookahead is required for the Prosody Generator to cliticize these structure-dependent cases. It should be added that for other varieties of encliticization, as well, there is no convincing counterevidence against this locality assumption.

The situation is only slightly different for *procliticization*, where the "little" word is adjoined to the following "big" word. This is far less widespread in English. The pronoun *it* in subject position can, in certain

*Because *is* is an auxiliary in sentence 7 only, there is also a *local* syntactic difference.

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

dialects of English, adjoin to the following auxiliary or main verb, as in *'t is winter* or *'t went away*. And there are dialects of English in which one can adjoin an indefinite article to the head noun (as in *anapple*, which then becomes syllabified as [ə-næpl]). Utterance-initial conjunctions, as in *And go now*, can also procliticize: *Ngonow*. These kinds of cliticization, of course, require a lookahead of one word, but probably no more. The phrase-structural condition on procliticization is probably just the absence of a major phrase boundary right after the potential clitic. Cliticization blocks in a sentence such as *John, who hated it, went away*, where [twɛnt] cannot be formed.* The one-word metrical and structural lookahead required here is the same as the minimal lookahead required for beat movement. And if this lookahead fails (for instance, because the "big" word is not retrieved in time), there will be no procliticization. The speaker will say *it – went away*, not *'t – went away*.

In conclusion: The speaker can generate phonological words incrementally. The phrase-structural and metrical conditions for cliticization are, it seems, locally available; a lookahead of no more than one word is required.

Two further closing remarks should be made on the generation of phonological words. The first one concerns the distinction made above between reduced auxiliaries and clitics of other kinds. Auxiliary clitics, such as *'ve* and *'ll*, it was argued, are indicated at the surface-structure level. Their lexical pointers have a diacritic feature that selects for the reduced allomorph. The reduction of most other small elements, such as *to* and it, was not treated as allomorphic; it was considered purely a matter for the Prosody Generator. Why not include the auxiliaries in this more general "late" account of reduction and cliticization? Kaisse (1985) gave various reasons for giving a lexical account of auxiliary reduction. A first one is that the reduced forms of auxiliaries are irregular, and therefore are probably stored as such. Take, for instance, the reduced forms of *will* and *would*: *'ll* and *'d*. No other English *w*-words reduce in this manner. (It would yield something like *'ch* for *which*.) A second reason is that there are slight distributional differences between the full and reduced forms of auxiliaries, testify-

* In addition to the phrase boundary after *it*, there are two other factors that might preclude cliticization here. First, *it* is not the subject of *went away*. Second, there is no c-command relation between *it* and *went*. A surface-structure constituent *A* c-commands a constituent *B* if, of every constituent of which *A* is a proper part, *B* is also a proper part, but without *B*'s being a proper part of *A*. Kaisse (1985) argues for the role of c-command conditions on cliticization. Both factors, however, are strictly local in surface structure.

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

ing to their lexical status. One can say *Where's the lions?*, but there is no correct slow-speech equivalent *Where is the lions?*

The other remark concerns the difference between inflections and clitics. If a reduced auxiliary is treated as a spelled-out morpheme that becomes cliticized, why not treat inflections in just the same way? To produce the form *walked*, there should be two lexical pointers in surface structure: one for the stem *walk* and one for the past inflection *-d*. The Prosody Generator would then encliticize the latter to the former, and induce the regular segmental spellout /wɔkt/. This would indeed be very similar to the production of *I've*. When inflections are treated as just a kind of closed-class elements, one also has an easier account of inflectional-shift errors, such as Garrett's example *I had forgot about*en *that*. The inflection *en* shifts, just as any closed class element can shift. Still, there are strong reasons for distinguishing inflections from clitics. These reasons are reviewed by Zwicky and Pullum (1983). Among them are the following: (i) Clitics are not very "choosy" about their hosts, whereas inflections are. The clitic auxiliary *'s* can attach to any kind of host, not only to a subject noun or pronoun. Here it adjoins, for instance, to a preposition: *The person I was talking to's going to be angry with me*. Inflectional suffixes, in contrast, attach only to a specific host category. Plural *s*, for instance, attaches only to noun stems. (ii) There is much irregular inflection (*give – gave*) but no irregular cliticization. (iii) Clitics can attach to other clitics, as in *I'd've done it*, but inflections cannot attach to inflections (except in speech errors such as *people read the backses of boxes*). These and other reasons make it necessary to distinguish carefully between the etiologies of inflections and those of clitics.

10.2.2 Phonological Phrases, the Grid, and Incremental Production

As the surface structure unfolds "from left to right," the speaker incrementally constructs phonological words. Can he also incrementally group these words into larger prosodic phrases—in particular, into phonological and intonational phrases? In this subsection I will argue that this is almost always possible, in spite of a theoretical counterargument. The speaker can normally construct these phrases without much "preview" of later surface structure. Let us begin with phonological phrases.

Chapter 8 presented a strict view and a more lenient view of phonological phrases. On the strict view (Nespor and Vogel 1986), an utterance is a concatenation of phonological phrases. They are, roughly, defined as stretches of speech leading up to and including a lexical head of phrase. The

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

more relativistic conception (Selkirk 1984a) is that the phonological phrase is a stretch of speech leading up to a weaker or stronger "break option." Let us consider the incrementality issue from both points of view. We will begin with the strict view, and consider example 3 of chapter 8, repeated here as example 11:

(11) //The detective /1 remembered //2 that the station /3 could be entered /4 from the other side as well //5.

All single and double slashes indicate phonological phrase boundaries.

As a first approximation, the incremental construction of a phonological phrase by the Prosody Generator can be straightforward:

Main Procedure Concatenate phonological words until one appears that is or contains a lexical head of phrase (i.e., head of NP, VP, or AP). Terminate the phrase right after that phonological word, except if the conditions for the Coda Procedure apply.

Ignoring for the moment the Coda Procedure, we can observe that this Main Procedure gives the correct result for positions /1 through /4: *detective*, *remembered*, *station*, and *entered* are lexical phrase heads. The head-of-phrase function is locally indicated in the developing surface structure (see subsection 5.1.3). The procedure requires no preview.

There is a problem, however, for the last phonological phrase. Its lexical head is *side*, but the phrase continues till after *as well*. How does the Prosody Generator know that the phrase should not be ended after *side*? Should it, for instance, know that there is no further lexical head of phrase in the offing? That would be a "preview" requirement.

That, however, is not necessary. The local surface-structural information (i.e., just between *side* and *as*) tells the Prosody Generator that (i) the current PP is finished and (ii) the new phrase is not a VP, a PP, an AP, or an NP. This suffices to add any newly created phonological words to the current phonological phrase. And it involves, again, strictly local information. The more general formulation can be the following:

Coda Procedure If a phonological word containing a lexical head of phrase completes that major constituent but is followed by a minor constituent boundary (i.e., not a VP, a PP, an AP, or an NP boundary), then add the following phonological words to the current phonological phrase until no more words follow or until a major constituent begins.

In sentence 11, this procedure will add the minor constituent *as well* to the current phrase *from the other side*. It will, in fact, complete it, since no more words follow. But the speaker might have continued with, say, *through a*

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

gate. In that case he would have had to round off the current phrase after *well* and begin a new one, because *through* opens a major (PP) constituent.

The Main and Coda Procedures guarantee incrementality of phonological phrase construction for most cases. One remaining problem concerns "nonlexical" heads of major constituents—in particular, pronoun heads of NPs. What, for instance, if example 11 had ended as follows?

/ 4 from the other side of it //5

The coda procedure does not apply here, because a new major constituent begins after *side*: the PP *of it*. Still, *of it* cannot be an independent phonological phrase, because its NP has (and is) a "nonlexical" head: the closed-class word *it*. So, it is to be added to the current phrase. But to decide this, the Prosody Generator should be able to "preview" the upcoming nonlexical head *it*. There are at least two possible reactions to this problem. The first one is to consider *side-of-it* as an encliticization, i.e., as a single phonological word. In that case the Main Procedure will build the correct phonological phrase. The phonological word *side-of-it* contains a lexical head of phrase (*side*); hence, the boundary follows that phonological word. Whether this solution suffices remains to be seen. It may, in particular, not be the case that encliticization of such phrases materializes in slow speech. A second reaction could be this: Major constituents with nonlexical heads tend to have this head in first or second position (*I saw it*, or *I heard of it*), because pronouns do not take complements to the left. The preview required, therefore, spans no more than two closed-class words, i.e., two syllable peaks.

Our provisional conclusion, therefore, is that phonological phrases can be incrementally produced. No substantial buffering of surface structure is required. The phrases can be produced as the surface structure unfolds "from left to right."

Let us now turn to the "lenient" view of phonological phrases. It should be remembered (see chapter 8) that in Selkirk's (1984a) theoretical framework the phonological phrase is only a derived notion. What really matters is the distribution of "silent beats" over the between-word positions of the metrical grid. When two words are separated by many beats, one can speak of a phonological phrase boundary. The number of silent beats between words is determined by a variety of factors, which we called "break options." According to the theory, the main break options are the end of an intonational phrase; the end of a sentence constituent; the end of a multiword NP, VP, PP, or AP; after a lexical head of NP, VP, or AP; and after a content or open-class word (see subsection 8.2.2). If each of these

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

factors contributes one silent beat when it applies, the sentence *Mary finished her Russian novel* displays the following distribution of silent beats:

(12) Mary xxxx finished xx her Russian x novel xxxxx

The size of the beat strings determines whether particular metrical phenomena can take place. For instance, stress clashes between adjoining words will be prevented (by beat movement) only if they are separated by no or few beats. The size of the strings also determines the use of various boundary markers, such as pitch movements, pauses, glottal stops, and syllable lengthening. (For a systematic study of these boundary markers in reading that strongly confirms the relevance of the just-mentioned factors, see de Rooy 1979.)

The issue of incrementality now involves two questions: (i) How much lookahead is needed for the Prosody Generator to insert the correct number of silent beats between one word and the next? (ii) How much lookahead is needed to compute each new word's stress level? Let us take up these questions in this order.

The number of silent beats inserted after a word depends on the number of prevailing break options. For each of the options we should, therefore, ask: Can it be locally recognized, or does it need structural lookahead? The least obvious factor in this respect is the first one, end of intonational phrase. We will return to it in subsection 10.2.3, where we will conclude that its status as a factor is circular. The end of a sentence constituent—i.e., a constituent immediately dominated by S (subject phrase, predicate phrase)—is locally given in surface structure. This also holds for the end of a multiword NP, VP, AP, or PP; the Prosody Generator must remember only that the phrase was multiword. Lexical heads of NP, VP, and AP are also immediately recognizable as such as the surface structure unfolds, and so are content-word lemmas. This means that, at the end of each word, the Prosody Generator can, without preview of later surface structure, determine which of these conditions are fulfilled. The answer to the first question is, therefore, that the distribution of silent beats can be incrementally computed.

The second question requires discussion of the metrical processes involved in the generation of a metrical grid. Take beat movement, demonstrated in example 5 with the generation of *sIxteen dOllars*. According to Nespor and Vogel (1986), the domain of beat movement is the phonological phrase; i.e., there will be no beat movement if the two words are separated by a phonological phrase boundary, as in sentence 6. In terms of grids, a separation by two or more silent beats will probably suffice to block

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

beat movement. The Prosody Generator can compute the beat movement for the first word if it knows this word's stress distribution, the number of silent beats following the word, and the second word's stress distribution. Since the silent beats are computed without lookahead, the only preview required is the second word's metrical form. Can it be known with just a one-word lookahead? It can in most cases, though it is theoretically not obvious. The case would be trivial if beat movement were to depend only on the "citation" metrical patterns of the two words involved. In that case, a preview of the next word's metrical spellout would suffice for the Prosody Generator to take a decision on beat movement. But according to Selkirk's theory, beat movement applies to materials that are already metrically processed to some extent, not to the spelled-out base forms. If this preprocessing is a condition for beat movement, one should first find out how much lookahead the preprocessing requires.

This preprocessing, called *text-to-grid-alignment* by Selkirk, involves various kinds of stress-assignment rules. For our present purposes we can refrain from reviewing most of them. With only two exceptions, they concern the composition of the citation forms of words, including word compounds. We are, however, assuming that the native English speaker has these basic patterns stored for all words he uses, except the extremely infrequent ones. A speaker can probably apply the rules when he forms a brand-new word, but normally he won't have to refer to them in his incremental phonological encoding. The two exceptions are the "Pitch-Accent Prominence Rule" and the "Nuclear-Stress rule." Can these rules be applied incrementally?

The *Pitch-Accent Prominence Rule* says that a pitch-accented syllable should be more prominent than any syllable that is not associated with pitch accent. Moreover, this rule overrides any other metrical rule.

This latter addition makes the rule very simple to apply in incremental fashion. When the Prosody Generator receives the basic metrical pattern of a pitch-accented word, it will process the diacritical pitch-accent feature by adding one or more extra beats to the syllable carrying word accent. This was already discussed (see example 4). In order to apply the rule correctly, the Prosody Generator must take care of two things. First, so many extra beats have to be given to the pitch-accented syllable that it is more prominent than any earlier non-pitch-accented word in the current intonational phrase. This requires a record of previous stress assignments, but no lookahead. Second, any following word in the phrase that has no pitch accent should be given less prominence. That is what "overriding" means. But at the moment of assigning pitch accent, the Prosody Generator need

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

not know these following words. It will implement this requirement as further words are received for processing. In short, pitch accent can be assigned on a purely incremental basis.

The *Nuclear-Stress Rule* says: Assign primary stress to the last nonanaphoric word in a major category or sentence (at least, this is one of its many formulations).

Selkirk's (1984a) treatment of this rule in terms of metrical grids was discussed in chapter 8 above. In each major phrase, so many beats are added to the accented syllable of the last word that it becomes the most stressed word in the phrase. Take the sentence *The sixteen dollars were lost*. After beat movement on *sixteen*, the grid looks like this:

```
(13)      x         x             x
          x   x     x   x     x   x
      x   x   x     x   x     x   x
     the six teen  dol lars  suf ficed
```

The subject NP, *the sixteen dollars*, has *dollars* as its last word. To have primary stress in the phrase (i.e., more stress than *sixteen*), it needs one additional best:

```
(14)                x
          x         x             x
          x   x     x x     x     x
      x   x   x     x x     x     x
     the six teen  dollars  suf ficed
```

But the Nuclear-Stress Rule will apply again on the level of the sentence as a whole. Here *sufficed* is the last word of the constituent. It needs two extra beats to surpass the stress of *dollars*. The end result is this:

```
(15)                                x
                    x               x
          x         x               x
          x   x     x   x     x     x
      x   x   x     x   x     x     x
     the six teen  dol lars  suf ficed
```

What does the Prosody Generator have to know in order to apply the Nuclear-Stress Rule? Can it generate the correct metrical pattern just going from left to right without knowledge of later surface structure? It can in the following way: In going from left to right, the Generator checks whether each word completes a major constituent. If a word does so, its stress is increased just enough to surpass the stress of all other words in the phrase. This, of course, requires a record of previous stress assignments in the phrase. But it requires no preview. The nuclear stress on *dollars* is independent of what follows. It is, in particular, independent of the later nuclear stress on *sufficed*. But the Prosody Generator must keep a record of

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

previous surface structure. When it gets to *sufficed* and notices that it completes a sentence, it must know where the sentence began, since *sufficed* is to be given more stress than any other element in the sentence. In short: The incremental assignment of nuclear stress requires memory of phrase structure and of previous stress assignments, but no lookahead.

One complication (which doesn't affect this conclusion) is the assignment of pitch accent. If the phrase doesn't contain a pitch accent, the procedure goes as outlined. If there is a pitch accent, the rule doesn't apply; it is "overridden." If there are two or more pitch accents and the phrase-final word has a pitch accent, then the Nuclear-Stress Rule applies again, and it gives the phrase-final pitch-accented element the highest prominence.

It is, therefore, safe to say that the metrical "preprocessing" does not require lookahead, but only memory. Thus, beat movement can always apply with one-word lookahead, i.e., after the next word's "preprocessed" metrical pattern has been computed. And this preprocessing requires no lookahead. Beat Movement is one of three "grid euphony rules," which create the alternating rhythm in speech. We will leave the other two rules, Beat Addition and Beat Deletion, untouched here. They do not change the picture, as they require no more preview than the stress level of the next word's first syllable.

However, potential "domino" effects must be discussed. Consider Beat Movement again. The current word's stress pattern shifts because of the next word's. But then, couldn't the current word's adapted stress pattern affect the previous word's? Theoretically, it could. For example:

(16) sixtEEn abstrAct pAIntings →
sixtEEn Abstract pAIntings →
sIxteen Abstract pAIntings

Here, the clash with *pAIntings* requires *abstrAct* to become *Abstract*. This, however, causes a stress clash with *sixtEEn*, which requires it to change to *sIxteen*. And indeed, this would be the "ideal delivery" of this phrase. This ideal delivery requires a two-word lookahead. But there is, of course, no theoretical upper limit on the domino effect; *any* amount of lookahead may be required. What is the psycholinguistic consequence?

What we called "preprocessing" (i.e., the assignment of pitch accent and nuclear stress to basic word patterns) is not subject to the domino effect. A speaker can always do this incrementally, without any previewing. Making speech really rhythmic, however, means applying the euphony rules, and this does theoretically require infinite lookahead. In practice, however,

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

cases like example 16 are rare—it is, actually, pretty hard to construct four- or five-word cases. With a one-word preview, the speaker can almost always come up with the correct rhythmic result. Whether he actually does is another issue, to which we will return in subsection 10.2.5.

When there is good reason for the speaker to approach "ideal delivery" of the utterance, he will presumably buffer more than a single word, so that cases such as example 16 will be recognized in due time. The conjecture can be made that one consequence of using a formal register is to increase the "window" or buffer of phonological encoding. We should, in addition, predict that, in fast speech, rhythm rules are the first to be disturbed, leaving the assignment of pitch accent and nuclear stress intact.

10.2.3 Intonational Phrases

"Break options" were mentioned in the discussion of the assignment of silent beats in the previous subsection. The speaker may or may not express silent beats, depending on register and rate of speech. The expression can take the form of lengthening a phrase-final syllable, inserting a small pause, making a pitch movement, and so on. But in all these cases the speaker basically continues; he does not really take a break. When the speaker does take a break, however, he factually completes an intonational phrase. This means that he selects an appropriate nuclear tone, and that after the break (which is usually followed by a pause of more than 200 milliseconds) he resets the baseline pitch to begin a new intonational phrase (if any). The former point will be taken up in section 10.3, the latter in the next chapter. The question here is: What determines whether a speaker will take a break option?

The break decision is, to some extent, under the speaker's executive control. The speaker may want to be highly intelligible to his listener(s), so he may speak slowly in short, high-keyed intonational phrases. He will then take every major break option to complete an intonational phrase. (The *reductio ad absurdum* of this speaking style can be observed in stewardesses' announcements during airline trips.) This freedom in taking break options immediately defeats all efforts to give a principled linguistic definition of intonational phrases (see also Ladd 1986). Intonational phrases are, to some extent, pragmatic devices under the speaker's intentional control.

There is, second, a general relation with speaking rate. Fewer options are taken at high rates than at low rates. A speaker concatenates much more in fast speech than in slow speech. In fact, the speeding up of speech is due largely to the leaving out of pauses. In addition, there is apparently a tendency to avoid making very short or very long intonational phrases, i.e.,

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

to distribute breaks evenly (Grosjean, Grosjean, and Lane 1979; Gee and Grosjean 1983). The reason for this tendency is unclear, but together with the just-mentioned relation to speaking rate it implies that speakers prefer to make intonational phrases within a particular range of duration. An intonational phrase should ideally span some 2 seconds and range between 1 and 3 seconds. Deese's (1984) data on spontaneous speech are in agreement with these estimates. The following example, taken from his table 6–5, gives the pause durations between intonational phrases and the duration of each phrase in milliseconds:

(17)

1,249 1,831
.../ 1,580 /before they discover that/ 499 /the bankers and the landowners/
1,593
680 /are going to take all the profits/
3,060
230 /and then insist that the Holbrook family owes them/
1,910 1,535
510 /so the idyll of farm life/ 420 /proves an illusion as well/

Such a tendency to equal durations (if it can be substantiated for spontaneous speech at all) may have its *raison d'être* in articulatory motor programming, either in the size of the Articulatory Buffer, or in the convenient amount of air to be inhaled, or in both.

A third set of factors are syntactic. Speakers usually break when they reach a sentence boundary. Taking pauses and change of pitch as evidence, Deese (1984) found that 76 percent of the sentence boundaries in his large sample of spontaneous speech were prosodically marked. Still, it should be noticed that an intonational phrase *can* cross a sentence boundary. Chapter 8 above mentioned the boundaries of parentheticals, tag questions, nonrestrictive relative clauses as attractors of intonational breaks. They are all sentential constituents. Another syntactic factor discussed in chapter 8 is the succession of three or more constituents of the same type (*the barbecue, the charcoal, and the icebox*), which invites "listing intonation"—i.e., one intonational phrase per constituent.

A fourth factor is semantic in origin. A speaker may be inclined to break shortly after a very prominent pitch accent. The latter will then become the nucleus of the intonational phrase.

A final factor is, one could say, an "operational" one: the availability of new surface-structure fragments. If, at a potential break point, no further surface materials have become available for phonological encoding, the

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

speaker must take the option in order to gain processing time. The lack of new ammunition may have different causes, as we have seen; there may be planning trouble at the message level, or at the level of grammatical encoding.

It should now be clear why I have called the intonational-phrase boundary a "circular" factor as a phonological break option: The decision to break *creates* the intonational-phrase boundary; it is in no way determined by it.

10.2.4 Metrical Structure and Phonic Durations

Metrical structure reflects itself, in part, in varying durations of segments, syllables, words, phrases, and pauses. The relation is not a simple one. A stressed syllable tends to be longer than an unstressed one, but a phrase-final unstressed syllable may also be stretched. Strings of silent beats in metrical structure will probably correspond to pauses, but Selkirk (1984a) suggests a more complicated relation: that the number of silent positions following a word will reflect itself in the sum of the last syllable's lengthening and the following pause. Thus, both metrical stress and silence may be mapped onto syllable length. The Prosody Generator should compute durational parameters for successive syllables and for the pauses between them. These parameters should be fed to the phonetic spellout mechanism, which generates the phonetic plans for successive syllables. Little is known about the computation of these temporal phonetic parameters, but a host of empirical studies, especially on syllable duration, are relevant to this issue. We will successively consider studies of syllable length in words, in phrases, and in larger utterances, and will complete this section with a few remarks on pauses.

It is a well-established fact that syllables in longer words tend to be shorter than the same syllables in shorter words. Nooteboom (1972) retrieved a publication by Roudet (1910), who gave measurements for the syllable [pâ] in French words of increasing length:

	centiseconds
pâte	27
pâté	20
pâtisserie	14
pâtisserie St. Germain	12

And this finding has been repeatedly reconfirmed by phoneticians (see, e.g., Lindblom 1968; Lehiste 1970). Lehiste (1970) suggested that this phenomenon might be due to a tendency of speakers to make words equally long. Syllables in longer words would then necessarily be shorter

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

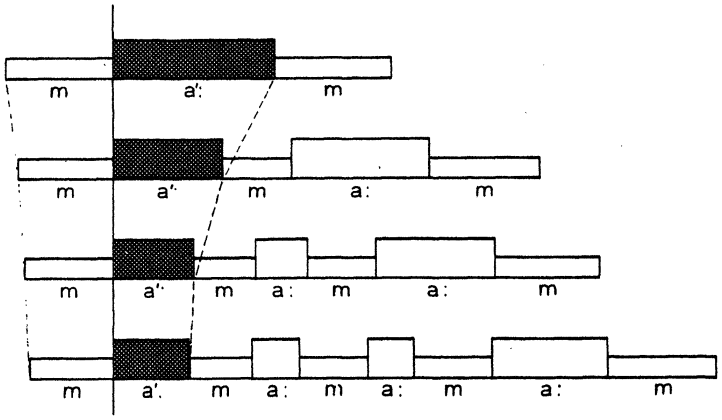


Figure 10.2
Duration of a stressed vowel in a word's initial syllable as a function of the number of syllables following the word. (After Nooteboom 1972.)

than syllables in shorter ones. But Nooteboom (1972) showed that this is not so. It is not the *total* number of syllables in a word that determines a particular syllable's duration, but the number of syllables *following*. To show this, Nooteboom had subjects pronounce nonsense words containing one stressed syllable *mAm* plus zero or more preceding and following syllables, all of the type *mam*. Figure 10.2 shows the average findings for zero, one, and two syllables following the stressed one. Only the number of syllables *following* had a substantial effect. The stressed vowel's duration is about halved going from zero to two following syllables. The consonant duration is also reduced, but only by about 20 percent. (See Gay 1981 for a similar result.) Fujimura (1981) calls the less reducible parts of an articulatory gesture "icebergs." A syllable is not stored with fixed relations between the durations of its parts. Lengthening and shortening a syllable affects its icebergs (in particular, its C parts) far less than it affects the waves in between (in particular, its peak or V part). But different consonants are "icebergy" to different degrees, and they tend to be softer in syllable-final position (Cooper and Danly 1981).

Nooteboom's results were confirmed by Nakatani, O'Connor, and Aston (1981) for stressed syllables only. Otherwise there was no tendency to compress syllables in longer words or to stretch short words. Word duration was strictly a linear function of the word's number of syllables.

Taken together, these results make it likely that there is no preferred duration parameter stored with the word as a whole. Rather, for each syllable the duration parameter is set as a function of the number of syl-

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

lables to follow in the word. The number of syllables (or peaks) in a word becomes available during metrical spellout. This can therefore be a basis for the Prosody Generator's computation of these parameters. Clearly, the Generator needs a full word's lookahead to decide on its first syllable's temporal parameter. Syllable length is, moreover, a function of metrical stress. Nakatani et al. (1981) found that the syllable carrying word accent is longer than a word's other syllables, and that this difference is more marked if the word is pitch accented.

But syllable length also depends on conditions outside the word—in particular, on the number of syllables following in the phrase. Fujimura (1981) gives the following examples:

It was yEllow ice cream

It was yEllow, I screamed

In the second utterance, where *yEllow* is phrase-final, that word is uttered at a far slower rate than in the context of the first utterance. Phrase-final lengthening of syllables has been shown repeatedly (Klatt 1975; Kloker 1975; Umeda 1975; Cooper 1976; Cooper and Cooper 1980; Cooper and Danly 1981; Nakatani et al. 1981; see also the short review by Vaissière 1983). Nakatani et al. (1981) found that stressed syllables become progressively longer toward the end of a (phonological) phrase. But this was not so for unstressed syllables, which are not sensitive to phrase position. On the other hand, both stressed and unstressed syllables did show the word-final lengthening effect.

Cooper and Danly (1981) reported that the phrase-position effect is stronger in the final phrase of an utterance (as in example 18 below) than in a nonfinal phrase (example 19):

(18) Bob made a counter-offer to the largest *bid*

(19) The couple made the largest *bid* on the cottage

This may be due to Nuclear Stress, which can only occur sentence-finally. If so, it confirms the finding of Nakatani et al. (1981) that a syllable's length varies with its degree of stress.

Also closely related to the phrase-position effect is what one could call the "word-in-isolation effect." Words in isolation are substantially longer than words in phrases. This may be a special case of the phrase-position effect; an isolated word is, of course, the last one in its phrase. Whether isolated words are *as long as* (other) phrase-final words remains to be tested.

Phrase-final syllable lengthening can be computed by the Prosody Generator without much foresight. It concerns the last phonological word in

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

the phrase; from there, the phrase boundary is "visible" to the Prosody Generator. A progressive lengthening of stressed syllables over the course of the entire intonational phrase, as reported by Nakatani et al., asks for a different kind of explanation. If it can be substantiated at all in more natural speech situations (Nakatani et al. used a modified form of Nooteboom's task), one might conjecture that the speaker "blindly" increases the duration of successive stressed syllables till he reaches the end of the intonational phrase, and that he then resets the durational parameter to the initial value for the next phrase. Another possible explanation involves nuclear stress. Phrase-final stresses naturally "grow" toward the end of the sentence, owing to the mechanics of the Nuclear-Stress Rule. And more heavily stressed syllables tend to be longer. The mechanics of nuclear-stress assignment do not require much preview, as we have seen.

Let us now turn to effects above the phrase level. Lehiste (1975) reported that readers stretch the last sentence of a (read) paragraph. It is not known whether this generalizes to turn-final sentences in spontaneous speech. Turn-taking intentions play an important role in spontaneous speech. A sentence-*shortening* effect, reported by Deese (1980, 1984), was mentioned in chapter 8 above. In contrast with the other studies mentioned in this section, Deese analyzed natural conversations. He found that they contained stretches of accelerated speech with a speaking rate of about ten syllables per second (the normal rate was five to six syllables per second). These could be expressive of modesty, or they could serve a floor-keeping function: "I am approaching the end of my sentence, but not of my utterance; there is more to come." Indeed, speakers finished them with the appropriate "continuation rise" as a boundary tone.

These utterance-level findings are, essentially, findings on speaking rate. And speaking rate is, to some extent, under direct intentional control. The speaker increases his rate mainly by cutting back on pausing. Compression of syllable duration hardly ever surpasses 25 percent.

This brings us to pausing. Let us first recall the study by Gee and Grosjean (1983), discussed in section 7.2. They analyzed the distribution of pauses in materials that were obtained in a reading task. In this task subjects read simple sentences, one by one, at different rates of speaking. Pause lengths could be perfectly predicted from a complex index, involving syntactic and prosodic features. The more such features coincided at a break, the longer the pause. The main prosodic features were whether the break involved a boundary between phonological phrases or between intonational phrases, but allowance was also made for a break before the

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

(phonological) word carrying sentence-final nuclear stress. Van Wijk (1987) published a reanalysis of these findings, which we already touched upon in section 7.2. He argued that prosodic factors alone sufficed to make the same predictions: The least pausing should occur between a content word and a function word adjoined to it. One could say that this protects the integrity of phonological words. The next larger pause option is between content words within a phonological phrase. Still larger pauses may be expected at boundaries between phonological phrases. Here van Wijk distinguished between "neutral" and "marked" phrases boundaries, which are both predictors of pausing. The latter type can only be interpreted as intonational phrase boundaries, because they may involve a nuclear pitch movement, such as a continuation rise. Still, van Wijk denies that Gee and Grosjean's intonational-phrase boundary feature is relevant. At any rate, however, it seems that the prosodic phrasing structure alone suffices to predict the pausing pattern in a reading task.

So far, we may conclude that speakers who can prepare for "ideal delivery"—and this is the case when the task is to read a single sentence—make their pausing durations dependent on the phonological phrase structure. One might want to test whether the number of silent beats of Selkirk's algorithm would be an equally good predictor of pause durations. One would also want to see further evidence for her conjecture that the number of silent beats following a word predicts the *sum* of syllable-final lengthening and following pause—i.e., evidence for a negative correlation between syllable lengthening and pausing. Some indirect evidence for the latter conjecture can be found in Scott 1982.

But one should never forget that pausing is multiply determined (see O'Connell and Kowal 1983 for a review). Pauses and nonpauses may, in particular, serve subtle communicative functions, as Kowal, Bassett, and O'Connell (1985) showed in an analysis of the reading and interviewing styles of two media professionals. In reading, sentence-final pauses were almost always made; in fact, the whole pausing pattern followed the norms of "ideal delivery". However, omission of pauses between sentences, and probably extension of the intonational phrase over the sentence boundary, occurred in almost 40 percent of the cases in the interviews. This happened especially in high-speed, ego-involved utterances. On the other hand, pauses can be inserted at odd places to create rhetorical effects, to suggest spontaneity, and so forth. There is much executive control in the management of pausing; one should not expect strict phonological rules to govern the distribution of silence in speech.

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

Also, one should exercise much restraint in generalizing findings on prosody in reading aloud to spontaneous speech.

10.2.5 Isochrony

It has been argued that English is a *stress-timed* language (Pike 1945) in which speakers tend to produce stressed syllables at regular and roughly isochronous intervals. Other languages, such as French and Spanish, are supposed to be *syllable-timed*, i.e., to give about equal duration to each syllable. Traditionally, the interval that begins at a stressed or salient syllable and ends just before the next stressed syllable is called a *foot* (Abercrombie 1967). The following example (from Halliday 1970) shows a partitioning in feet:

each / fOOt in / tUrn con / sIsts of a / nUmber of / sYllables /

Stress-timing would mean that feet tend to be equally long. This notion of the foot as a prosodically relevant entity has been largely abandoned in linguistics, and I see no role for it in a theory of language production either. The modern notion of foot is entirely internal to the phonological word (Nespor and Vogel 1986). Still, it is an empirical issue whether the spacing of stressed syllables in English tends toward isochrony.

How could isochrony be attained in the production of speech? In two ways. Speakers could, first, stretch or compress syllable durations, depending on the number of syllables in a foot. When there are many syllabels in a foot, they should be pronounced at a higher rate than when there are only a few. Second, speakers could add, delete, or shift accents so as to make feet that are about equally long. As things stand, there is only conflicting evidence for the hypothesis that speakers vary syllable durations in order to establish isochrony. But there is good reason to suppose that speakers shift accents to create a more even distribution of stressed syllables. Let us discuss these two ways of establishing isochrony in turn.

Lehiste (1977) reviewed the isochrony research and concluded that some findings spoke for and other findings spoke against isochrony. She mentioned in particular that speakers often violate isochrony in order to mark syntactic boundaries (see Cutler and Isard 1980 for some experimental evidence).

Since the publication of Lehiste's review, some further studies have appeared, again with ambivalent results. Nakatani et al. (1981) found absolutely no evidence for isochrony in their experimental data. Their measurements concerned so-called reiterated speech, in which a subject

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

would read an adjective-noun pair like *remote stream* and pronounce it as *mamAmAm*. The pair was embedded in a sentence, like *the remote stream was perfect for fishing*, and the rest of the sentence was spoken normally; the full stretch of speech would therefore be *the mamAmAm was perfect for fishing*. The adjective-noun pairs were constructed in such a way that they contained shorter or longer feet. (The above sentence, for instance, contains the one-syllable foot/mʊt/.) Nakatani et al. analyzed the durations of one-, two-, three-, and four-syllable feet as spoken by their most fluent subjects. If there is a tendency toward isochrony, the foot duration should not increase linearly with the number of syllables. But it did. Each syllable added a fixed amount of time (about 150 msec) to a foot's duration, irrespective of the foot size. Each syllable consisted of precisely two segments, [m] and [æ], so that the average segment duration was about 75 msec. The relevance of this will soon be apparent.

Jassem, Hill, and Witten (1984), on the other hand, found some evidence for isochrony in a detailed segment-by-segment analysis of the recorded materials that go with Halliday's (1970) course in spoken English. They found that segments (and thus syllables) were significantly shorter in long feet than in short feet, and that the average segment rate was 13.3 per second (which corresponds to an average segment duration of 75 msec). The speech rate in the two studies considered here was therefore precisely the same. Still, the findings on isochrony differ. Nakatani et al. caution against overgeneralizing their findings. In particular, their experimental materials were reiterated adjective-noun phrases; they did not contain function words. It could be that some degree of isochrony is obtained by manipulating the duration of function words only.

There are other interesting findings in the study by Jassem et al. One of them is that there is no isochrony in so-called *anacruses* (short stretches of high-rate speech). As was mentioned above, Deese (1984) showed that such stretches are quite normal in spontaneous speech. It is relevant here that whatever there is in isochrony breaks down at these high speaking rates. There may be isochrony in trot, but there is none in gallop.

Dauer (1983) compared the isochronous tendencies of five different languages, including "stress-timed" English and "syllable-timed" Spanish. The materials were literary texts read by native speakers. That choice is regrettable, because literary texts may very well have been *designed* to be rhythmic. Still, the results were somewhat surprising. The variability of inter-stress intervals (i.e., feet) was the same for stress-timed as for syllable-timed languages. English feet were as variable in duration as Spanish feet. And for all languages the average foot length was statisti-

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

cally the same, namely between 400 and 500 msec. But is this isochrony? On the present interpretation, isochrony can only mean that foot length is not a linear function of the number of syllables it contains; the more syllables there are in a foot, the shorter they should be. Dauer showed, however, that for her data the function is a linear one. Each additional syllable added 110 msec to the interstress interval, and in all languages. In other words, there was no isochrony, even for readings of literary texts.

The contradicting results of Dauer and Jassem et al. leave the issue as undecided as it was to start with. There is, at any rate, no reason so far to conjecture an internal clock or pacemaker that induces the speaker to deliver speech in isochronous feet. In spite of her own data, Dauer aligns herself with such a notion.

The second possible way for a speaker to establish isochrony is by manipulating the placement of stress. Beat Movement does just that, as do Selkirk's other euphony rules. They promote a rhythm of alternating stresses, and thus they promote isochrony. This view of isochrony is linguistically far better motivated than the previous one. It should, in particular, be noticed that it has no consequences for syllable length. Isochrony is established by evading strings of adjoining stressed syllables or of adjoining unstressed syllables, not by squeezing more or fewer syllables into a fixed temporal frame.

The data one would need in order to prove this version of isochrony would be that euphony adjustments are indeed made in spontaneous speech. For instance, is it statistically more often the case that speakers shift clashing stresses apart (as in *sIxteen dOllars*) than that they shift them together (as in the unlikely *he becOmes sIxteen*)? Two recent studies were devoted to this issue, one by Cooper and Eady (1986) and one by Kelly and Bock (1988).

Cooper and Eady report five experiments in which they created conditions for beat movement, such as clashing stresses. The following two sentences, for instance, appeared on a list to be read by subjects:

(20) Thirteen corporations submitted bids to build the new shopping mall

(21) Thirteen companies submitted bids to build the new shopping mall

In example 20, *thirteen* can be normally pronounced with main stress on the second syllable, since it is followed by two unstressed syllables. In example 21, however, there is the risk of clashing stresses, since *company* begins with a stressed syllable. An isochrony or beat-movement tendency on the part of the speaker would induce the pronunciation *thIrteen*. And

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

Cooper and Eady constructed several other cases where metrical analysis would predict a shift toward a more isochronous rhythm. All the sentences were read aloud, and the critical syllables were analyzed with respect to their duration and pitch (F_0); a stressed syllable should be of longer duration and/or of higher pitch than an unstressed one. The results of this extensive and careful study were completely negative. There was no measurable difference between examples 20 and 21 insofar as the syllables of *thirteen* were concerned. And, similarly, there was no evidence for stress shifts in any of the other conditions tested. Are we being deceived by metrical phonology?

Kelly and Bock (1988) restored confidence. Their experimental approach was very different. They provided their subjects with sentences containing a two-syllable nonsense word, such as *colvane*. In the context of a given sentence, the nonsense word would function either as a noun or as verb. In examples 22 and 23, for instance, *colvane* plays the role of a noun, whereas it figures as verb in examples 24 and 25.

(22) Use the colvane proudly. [noun, trochaic biasing context]
(23) The proud colvane proposed. [noun, iambic biasing context]
(24) Planes will colvane pilots. [verb, trochaic biasing context]
(25) The pins colvane balloons. [verb, iambic biasing context]

Kelly and Bock predicted that noun function would induce a trochaic word accent, i.e., *cOlvane*, following the majority rule for English nouns (*tIger*, *sOldier*, etc.). Conversely, they expected the nonsense verb to receive iambic accent, i.e., *colvAne*, in accordance with the majority rule for English verbs (*convEne*, *expEct*). The critical variable in the experiment, however, was the metrical context in which the nonsense word appeared. The noun in example 22 is preceded by a normally unstressed syllable (*the*) and followed by a stressed one (*proud*). This environment would support the expected trochaic rhythm of the "noun" *colvane*. But in example 23 the trochaic rhythm *cOlvane* would clash with the metrical context. It would create a stress clash between *proud* and *col*. The context biases for a iambic pattern: *the prOud colvAne propOsed*. Similarly for the nonsense verbs in examples 24 and 25: In the former, the context biases toward a trochaic word accent, contrary to the normal iambic pattern for verbs; in the latter, however, the context supports the iambic verb pattern.

In the Kelly-Bock experiment, subjects read such sentences and their speech was tape-recorded. The critical nonsense words were then excised from the tapes, and two judges categorized the word accents as either

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

iambic or trochaic. The results, which were consistent between the two judges, amounted to this: Although trochaic patterns dominated, nouns (as in examples 22 and 23) received significantly more trochaic pronunciations than verbs (as in examples 24 and 25). Trochaic biasing contexts (as in examples 22 and 24) released more trochaic pronunciations than iambic biasing contexts (as in examples 23 and 25). The latter result shows that speakers do tend to impose an alternating stress rhythm, in spite of the fact that they have preferential accentuations of verbs and of nouns. There is a tendency toward isochrony, in that stress clashes are, to some extent, evaded by adjusting a word's metrical pattern.

The question remains why these results are so different from those obtained by Cooper and Eady. A first point to be noticed is that Kelly and Bock's contextual biasing effect, though highly significant, is not exceedingly strong. The trochaic biasing context released 84 percent trochaic patterns, the iambic context 77 percent. When effects are this small, one needs many observations to detect them. Kelly and Bock had more than 50 times as many observations per condition than Cooper and Eady. A second potentially important difference was that Kelly and Bock used perceptual judgments of stress, whereas Cooper and Eady had physical measurements. Perceived stress is a complex function of a syllable's composition, loudness, duration, pitch movement, and (maybe) precision of articulation. All these features were available to the judges, and they were able to weigh them. Cooper and Eady had only duration and pitch measurements, and may thus have missed other subtle features contributing to stress. Third, Cooper and Eady used real words. It may be the case that the effect observed by Kelly and Bock is even smaller when real words are used.

Together, these two studies strongly invite further experimental exploration of metrical euphony rules, including experiments that involve spontaneous speech rather than reading aloud. Almost all experimental research in prosody involves reading tasks, but in reading the normal conditions for incremental speech planning are not met. Results on a speaker's prosodic planning and lookahead in reading cannot be generalized to normal spontaneous speech.

A study by Cutler (1980b), finally, provides interesting evidence that a speaker's tendency to impose an even distribution of stresses can induce characteristic speech errors. In particular, Cutler studied spontaneous errors involving either syllable omission (as in example 26) or stress shifts (as in example 27).

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

(26) Next we have this bicEntial rug [bicentEnnial]

(27) We do think in spEcific terms [specIfic]

Cutler checked whether errors like these established more isochrony than there would have been in the target utterance. For instance, the target utterance for example 26 would have had the following foot structure; the error foot structure is given as a comparison:

Target: / Next we / have this bicen/tennial / rug /

Error: / Next we / have this bi/cential / rug /

The foot containing the error (the second foot) is closer in number of syllables to the surrounding feet than is the second foot in the target sentence. Hence, the error establishes more isochrony than the target would have displayed. Cutler found that this tendency toward more isochrony was highly significant for her corpus of syllable omission errors.

In a similar test for stress errors (such as in example 27), Cutler established that the resulting patterns of feet were more isochronous than the target patterns.

Where could these errors arise in the model framework of figure 10.1? Example 26 involves the deletion of a syllable. One might guess that the Prosody Generator skips the delivery of an unstressed syllable frame to the phonetic spellout level. The error in example 27 is harder to account for. As Cutler (1980a) showed, such stress-placement errors originate in the simultaneous activation of a morphologically related word. For *specific*, the related word would be *specify*, which has word accent on the first syllable. One might conjecture that the metrical patterns of both words were spelled out in metrical spellout, and that both stress patterns were simultaneously fed to the Prosody Generator, which then made its choice in such a way that euphony or rhythm could be established with least effort. This explanation, however, falls short of accounting for word-stress errors that involve a mixture of the metrical patterns of the two concurring words (as in *articulAtory*, which has the number of syllables of *artIculatory* but the accent placement of *articulAtion*). The ways of the Prosody Generator are still quite enigmatic.

This completes our review of isochrony, the main conclusion of which is that the Prosody Generator's metrical planning promotes an alternating distribution of more-stressed and less-stressed syllables. This establishes some degree of isochrony. The original notion of isochrony, however, finds rather little support. Syllable lengths are seldom or never adapted for the purpose of spacing stressed syllables evenly over time.

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

This also completes our review of metrical planning, the first main job of the Prosody Generator in the generation of connected speech. We sketched how a metrical pattern could be computed for phonological words, and for phrases up to the level of intonational phrases. We analyzed whether this generation could be done incrementally, without more preview of surface structure than a single word. The answer was encouraging, in spite of the fact that the euphony operations, which establish rhythm, theoretically require infinite lookahead. We also considered one of the Prosody Generator's main types of output: a pattern of stress and durational parameters for successive syllables and silences. These parameters are to be implemented during the phonetic spellout operations. They are sensitive to a syllable's position in a word, phrase, paragraph, or turn. Finally, empirical research on isochrony was reviewed. There is only very limited support for the original idea that a speaker adjusts the lengths of unstressed syllables so as to make intervals between stressed syllables more isochronous. There is a better empirical basis for supposing that, at least to some extent, speakers like to impose an even distribution of stressed syllables.

10.3 The Generation of Intonation

The second main job for the Prosody Generator is to compute pitch contours for successive intonational phrases. The melody of an intonational phrase, as we saw in subsection 8.2.3, is the result of a variety of forces. It expresses the speaker's affective involvement, especially in key and register. Its tune is raised when a new topic is introduced, or in response to the interlocutor's introduction of a new topic. It also expresses, by way of continuation rise or final fall, whether the speaker intends to continue or not. It signals, through pitch accents, where there is prominent, new, or contrastive information. And it is, through its nuclear tone, an important instrument for expressing the utterance's illocutionary force.

There is no process model that provides an on-line computation of the sentence melody resulting from all these forces. In the following we will, therefore, set ourselves a fairly limited task. We will consider some of the factors affecting each of the global and local properties of sentence melody, and the effects they have. The global properties are declination, key, and register. The local ones are prenuclear tune and nuclear tone. The main message of this section will be that intonational planning can probably be done incrementally, without much lookahead. But "euphony" or melody can be improved when there is a small amount of

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

preview. This is in full harmony with what we found for metrical planning in the previous section.

10.3.1 Declination

There is evidence that, at least in many languages, pitch gradually drifts down in the course of an intonational phrase. Cohen and 't Hart (1965) called this phenomenon *declination.* An example is given in figure 10.3. It is as if the pitch movements that go with pitch accent and nuclear tone are superimposed on a generally downward-drifting "declination line." Willems (1983) measured an average declination of 0.3 semitones per second for spontaneous British English speech (there are twelve semitones in an octave). The variability in such measurements is, however, so large that doubts about the universality of the declination phenomenon are warranted (Lieberman, Katz, and Jongman 1985; see also Ladd 1984 on the statistical claims with respect to declination).

Where declination is systematic, it may be due not to the speaker's phonetic plan but rather to physiological factors such as diminishing subglottal air pressure (see chapter 11). At the end of an intonational phrase, the speaker inhales and the pitch level is reset to the higher starting position. As a result, a kind of sawtooth pattern of declination arises over successive intonational phrases.

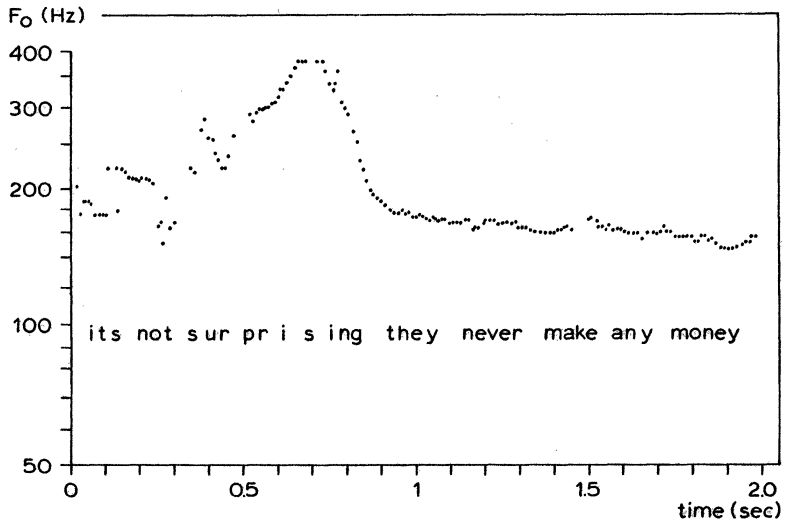


Figure 10.3
The gradual declination of pitch over an intonational phrase. The pitch-accent peak is superimposed on the "declination line." (From Cohen et al. 1982.)

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

But it has been suggested that some planning of declination is possible. There is, in particular, some evidence that declination is steeper for short utterances than for long ones (Ohala 1978; Cooper and Sorensen 1981; De Pijper 1983; Collier and Gelfer 1983). The latter finding is, on first view, not easy to reconcile with an incremental theory of intonation. How can the speaker know in advance how long his intonational phrase is going to be, if he has not even generated the full surface structure for that phrase? It is too early, however, to draw dramatic conclusions. First, the effect is most apparent in reading. But in reading a speaker does have a preview in the most literal sense. Second, the causal relation (if any) between phrase length and slope of declination may be inverse. If a speaker, for whatever reason, makes his pitch decline rapidly, he will sooner feel the urge to reset. This may induce him to take an early break option. Consequently, the running intonational phrase will be a short one.

10.3.2 Setting Key and Register

The range of pitch movement in an intonational phrase—the key—depends in particular on the "news value" of that phrase. When a phrase expresses "main-track" or foregrounded information (subsection 4.3.5), the key will, as a rule, be higher than when it expresses "side-track" or background information. Brown, Currie, and Kenworthy (1980) found that the key is also higher when a speaker introduces a new topic, and that the pitch excursions diminish when a topic becomes exhausted. They also showed that the baseline of the pitch range (i.e., the register) is lifted as a whole when a speaker introduces a new topic. Key and register are, moreover, lifted together when the speaker has a strong ego involvement in what he is saying. That ego involvement can be due to general communicative tension (Heeschen, Ryalls, Hagoort, and Bloem, forthcoming), to surprise, or to enthusiasm. A higher register is also chosen to express friendliness, helplessness, and so on.

These settings of key and register do not require any lookahead. As far as register is concerned, the Prosody Generator will set a global pitch parameter for phonetic spellout. This is the default pitch level for successive syllables. Pitch excursions for pitch accenting or nuclear tones are then programmed as deviations from the baseline. The key (i.e., the size of the pitch excursions) may be set as a global parameter to be instantiated every time a pitch accent is to be made. In this way, foregrounded and backgrounded information may be globally opposed by a speaker. Alternatively, the size of the excursion may be set anew for each syllable

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

requiring a pitch excursion, depending on the accessibility or the contrastiveness of the particular lexical item.

10.3.3 Planning the Nuclear Tone

A tone, as we saw in chapter 8, is not an indivisible whole. There is, first, the *nuclear pitch movement*: a step up or down to the nucleus, plus a fall or rise (or steady level) from the nucleus. There is, second, a *boundary tone*: the pitch movement that takes place at the final syllable of the intonational phrase. These two pitch movements can be separated by several syllables, or they can follow one another within a single syllable. This is exemplified in figure 10.4, which gives three cases of tone V (the fall-rise). In figure 10.4a the nuclear movement is on *po*, the boundary tone on *leave*. In figure 10.4b the nucleus is still *po*, but now *bear* is the end-of-phrase syllable, and it carries the boundary tone. In figure 10.4c *bear* is both the nucleus and the boundary syllable, and both pitch movements are then projected on that same single syllable.

These two components of a tone can probably be set independently, and they play different expressive roles. The nuclear pitch movement is mostly a focusing device. It indicates the most prominent lexical item in the intonational phrase. But in addition it has an illocutionary function:

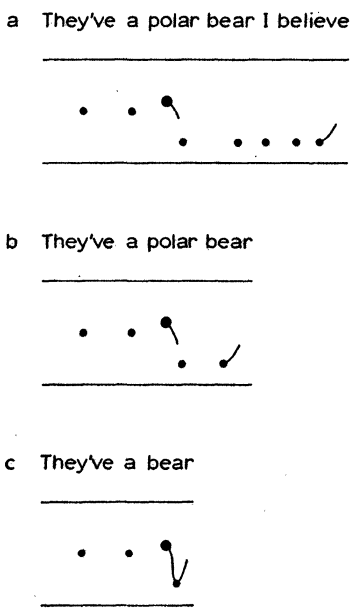


Figure 10.4
Decreasing separation of nuclear pitch movement and boundary tone.

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

expressing matter-of-factness, being reassuring, or the like. The boundary tone can perform several different functions. It can indicate a mood of finality or nonfinality. It can express the utterance's illocutionary force, the kind of commitment the speaker is making. It can express the speaker's intention to finish a turn or to continue. And it can be expressive of attitude, as in the rising tone of friendliness.

In order to generate the nuclear pitch movement, the Prosody Generator must select the nuclear syllable. How much lookahead is needed to do this? In subsection 10.2.2 we saw that the assignment of pitch accent or nuclear stress requires no lookahead. But in order to assign a nuclear pitch movement, the Prosody Generator must know which pitch-accented or nuclear-stressed syllable is the last one of the intonational phrase. Here we should return to the conclusion of subsection 10.2.3: "the decision to break *creates* the intonational phrase boundary; it is in no way *determined* by it." In other words, every time the speaker reaches a pitch-accented or nuclear-stressed syllable, a decision can be made to give it the nuclear tone and to break at the next convenient break option. This requires no lookahead whatsoever. In particular, the convenient break option need not be in view. There is only the decision that it should be taken as soon as it appears. It was suggested in subsection 10.2.3 that very prominent pitch accents, in particular, may induce a speaker to decide on a break.

The fact that lookahead is not a necessary condition for positioning the nuclear pitch movement by no means excludes the possibility that speakers do look ahead in deciding where to begin the nuclear tone. As we have seen for metrical planning, lookahead can increase the "euphony" of speech. Especially in slow speech, where the generation of surface structure can be far ahead of phonological encoding, a more ideal delivery can be planned when the speaker has preview of a sentence boundary. He can then decide to complete the intonational phrase at that sentence boundary, and hence to make the nuclear pitch movement at the last pitch-accented or nuclear-stressed syllable of the sentence.

The generation of the boundary tone obviously requires no more lookahead than a single syllable. It is made on the last syllable before the break; only that break must be in view.

The Prosody Generator must, in some way or another, know what nuclear tone to impose. It depends, we saw, on a multitude of factors. Some of these factors are probably indicated at the message level. Illocutionary force, mood, and modality are planned by the speaker when he generates a message to be expressed (see subsection 3.5.1). At the level of surface

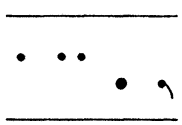
Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

structure these are translated in part into syntactic sentence mood (declarative, interrogative, imperative), in part into the use of modal verbs or adverbials, and in part into indicators for nuclear pitch movements and boundary tones (see section 5.2). These indicators or parameters in surface structure are recognized by the Prosody Generator, or so we assume.

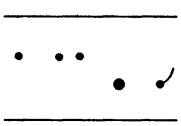
Still, it is unlikely that these parameters are the sole determinants of nuclear tone. Take, for instance, the choice of boundary tone. Many questions are asked with a falling boundary tone (Brown et al. 1980; de Pijper 1983). Consider example 28a, with its low-fall nuclear tone.

(28)

a Did Peter call you



b Did Peter call you (He said he would)



Let us assume that the falling boundary tone was indicated at the level of surface structure. This indication will easily be overruled if the speaker decides, shortly before finishing the sentence, not to finish his turn but to add another clause (*He said he would*). He will now make a continuation rise as boundary tone on *you*, as in example 28b. This looks like direct executive control rather than replanning of a surface structure. Also, the emotional and attitudinal aspects of tone, such as friendly rising boundary tones, are probably directly induced by the emotional system without mediation through message and surface structure. The Prosody Generator will, in some as yet unknown manner, integrate these various sources of activation in making the final choice of tone.

10.3.4 Planning the Prenuclear Tune

The nucleus can be the only accented syllable in the intonational phrase. The default tune is then an about constant mid-level pitch. There are, of course, always small pitch variations. There is, in particular, systematic covariation with metrical stress. Stressed syllables tend to be somewhat higher in pitch level than unstressed ones. Or, more precisely: Metrical stress is, in part, realized through variation in pitch. The default tune requires no more lookahead than is required for assigning metrical stress. But a tune's course can be "improved" when there is lookahead. De Pijper (1983) found that when a tune without accents spans several syl-

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

lables, a reader will show *inclination*. That is, he will make a gradual rise of pitch up to the nucleus. Default tune and inclining tune are depicted in example 29, where the nucleus has tone 1.

(29)

The amount of lookahead needed for an inclining tune is, of course, rather limited. A speaker can incline almost "blindly"; the only risk is that he may come up against his pitch ceiling, as sportscasters sometimes do. In other words, the speaker should not incline too much as long as the nucleus is not yet in sight.

If an intonational phrase introduces a new topic in the conversation, its first stressed syllable and the subsequent tune can be quite high-pitched, at least in British English (Brown et al. 1980). More generally, a high-pitched tune signals a speaker's special involvement with what is said. Also, *wh*-questions and certain kinds of emphatic utterances (*Who will be at the party?; Do come to the party*) have high pitch-initial tunes. That high starting level can, of course, be set without specific lookahead.

A tune may also contain one or more additional pitch accents. Depending on the amount of lookahead, these pitch accents are realized individually or in conjunction. If there is one further pitch accent in the tune, these two cases can be illustrated as in example 30a and 30b, where the nuclear tone is a high-fall and a low-fall, respectively.

(30) (a) individual realization (b) "hat pattern"

These two patterns are alternative realizations for sentences such as *We mIght be able to gO*, with *might* and *go* accented. In example 30a, the tune falls back to mid-level immediately after pitch-accented *might*, and the level is maintained until the nuclear tone is made on *go*. Example 30b depicts a so-called hat pattern ('t Hart and Collier 1975); the speaker accents

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

might by way of a rise and *go* by way of a step down to mid-level, and in between these the tune maintains a high pitch level. Lookahead is a condition for producing the hat pattern. The speaker must know at the first pitch accent that another one is going to follow. There are no clear differences in intonational meaning between the two patterns in example 30, but the hat pattern is more euphonious. It constitutes a more aesthetically pleasing delivery of two subsequent accents. The hat pattern is a more likely pitch contour when two accents are to be made in close succession. But even then a speaker often makes individual realizations. Brown et al. (1980) found that in such cases the fall back to mid-level between the pitch peaks was often not fully realized; they called this phenomenon, in which the syllables between the peaks stayed somewhat raised in pitch, *tonal sandhi.*

The pitch-movement parameters that the Prosody Generator incrementally computes for prenuclear tune and nuclear tone affect the phonetic spellout procedures directly. Successive syllables are set to rise or to fall in pitch, or by default to stay level with the previous syllable. The sizes and slopes of the rises and falls are programmed, as well as their precise timing; it is quite critical where in the syllable a rise or fall is made ('t Hart and Collier 1975).

This completes the rather brief treatment of intonational planning. Metrical and intonational planning go hand in hand. The metrical peaks are the main loci of pitch movement, and the planning of intonational phrases is as much a metrical as an intonational affair. We have, moreover, seen that both forms of prosodic planning can be done incrementally. It is not necessary for a speaker to buffer long stretches of surface structure in order to program the prosody of connected speech. But if such buffering is possible, as it is in slow speech or in reading, the Prosody Generator can generate a more euphonious output, more rhythmic phrasing, and larger melodic lines.

10.4 The Generation of Word Forms in Connected Speech

The consideration of the generation of word forms in the previous chapter was limited to the spellout of stored ("citation") forms. The present section will reconsider word-form spellout as it occurs in connected speech. The present chapter began with the argument that in connected speech the syllabic and segmental composition of word forms are context dependent. Segmental and phonetic spellout depend on prosodic decisions. The domain of syllabification, in particular, is not the citation form

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

but the phonological word (see subsection 10.2.1). The Prosody Generator must, therefore, construct phonological words before complete phonetic spellout is possible. But the construction of phonological words is impossible without access to segmental information. This issue will be taken up first in the present section; we will then move to some aspects of phonetic spellout in context.

10.4.1 Segmental Spellout in Context

To account for context-dependent segmental spellout, such as occurs in assimilation (e.g., /tɛm bʊks/), in cliticization, and in reduction (e.g., *a bottle o'milk*), we must assume that segmental spellout is a two-step process.

The first step is basically as described in subsection 9.2.2. When the morphological slots of an item's address frame are filled with the appropriate morphemes, the item's stored syllabic and segmental composition is spelled out.

The second step involves the Prosody Generator. It receives, for each successive lexical item, the spellout from the first step. The metrical information is independently fed into the Prosody Generator from metrical-spellout and phrase-structural information, and a partitioning is made into phonological words. The most frequent case (in English) is that a single input item becomes a single phonological word. But with the limited lookahead discussed in subsection 10.2.1, phonological words consisting of two or even more lexical items can be built up. The segmental strings of these items are concatenated and modified, following quite general phonological rules. In addition, assimilations at phonological word boundaries are generated (as in/tɛm bʊks/).

This modified segmental output forms, for each phonological word, the input to the phonetic spellout procedures. Syllable frames are addressed in just the way proposed in subsection 9.2.3., with metrical and intonational parameters set for each of them.

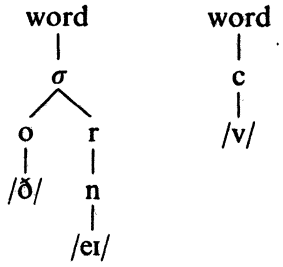
Two examples may help to clarify the two-step procedure at the level of segmental spellout. The first is a case of what was called *obligatory* cliticization in subsection 10.2.1; the second is an instance of *optional* cliticization.

Cliticization is obligatory if the "little" element is nonsyllabic. Auxiliary forms such as *'ve* and *'ll* are allomorphs of the full forms, and we have considered reasons to suppose that they are indicated as such in the surface-structure representation. So, if the speaker uses a casual speech register to generate the sentence *They have called*, the first two morphemes accessed will be *they* and *'ve*. The first step of segmental

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

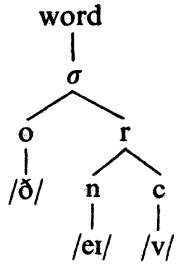
spellout will, for these elements, yield the following:

(31)



When the Prosody Generator receives this pair, it will initiate a second step that consists of attaching the nonsyllabic coda-branch /v/ of the second word to the rime of the first word. This gives the following:

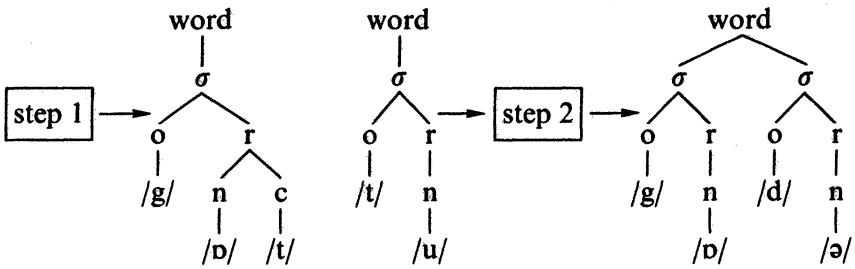
(32)



The triple of syllable constituents—onset, nucleus, and coda—then becomes input to phonetic spellout. Because this phonological word consists of one syllable, the Prosody Generator generates one syllable frame for the phonetic spellout procedure. That syllable frame is "enriched" with metrical and intonational information, as was discussed in subsections 10.2 and 10.3.

The Prosody Generator can impose an optional encliticization when the register is casual and the speech rate relatively high. When the speaker is generating the sentence *John got to swim in the morning*, the first step of spelling out the lexical items *got* and *to* produces their "citation" syllabic structures. In a second step, the Prosody Generator combines them into one phonological word, which involves resyllabification. These steps are shown in the following diagrams.

(33)



Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

Here we see both deletion (of /t/) and change (/t/ → /d/, /u/ → /ə/) of segments. This, together with the loss of one word boundary (i.e., a silent beat), will probably simplify the eventual phonetic pattern.

It was stated above that these phonological adaptations are quite regular. This is more a hypothesis than an established fact. But there is reason to make that hypothesis. If some form of accommodation or enclitization is irregular (i.e., specific to a particular word or a pair of words), it is probably lexicalized. That is, the encliticization is stored as such in the lexicon. Examples of this are the contraction of *do not* to *don't* and that of *will not* to *won't*. These are rather irregular encliticizations in English. In other words, if the Prosody Generator is to deal with one of these constructions, it must have stored its irregular shape. But such storage is precisely what the lexicon is for. The lexicon is the repository of forms: the Prosody Generator can adapt them to context. One should also not exclude the possibility that frequently occurring regular enclitizations are stored as such; /gɒdə/ may be such a case, and the even more frequent /wɒnə/ (for *want to*) is a good candidate.

We will not consider in any detail the rules governing step 2. They are different for different languages, and even for English they are not well known (but see Dogil 1984, Kaisse 1985, Nespor and Vogel 1986, and Pullum and Zwicky 1988). It should be repeated, however, that these rules are sensitive to local phrase structure. There will be no encliticization of the kind illustrated in diagram 33 when *got to* occurs in the sentence *John got, to be sure, to swim every morning*. The parenthetical phrase boundary following *got* blocks the formation of the phonological word /gɒdə/. Local phrase-structural conditions govern the applicability of encliticization and assimilation rules. Generally speaking, the phonological operations in step 2 prepare a segmental output string that can be pronounced with less articulatory effort than the input string.

The step 2 procedures compute the segmental consequences of metrical structure. The Prosody Generator receives an item's stored metrical information. In addition, it receives an item's pitch-accent feature, if any. The resulting metrical planning (see section 10.2) can also have consequences at the segmental level. For instance, a syllable's vowel may be a different segment when it is stressed than when it is unstressed. These consequences of metrical planning are presumably also implemented at step 2.

10.4.2 Phonetic Spellout in Context

The first input to phonetic spellout is a string of address frames, each containing three slots: one for onset, one for nucleus, and one for coda

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

(but see subsection 9.5.1). Each frame is "enriched" by prosodic information; it contains parameters for the syllable's duration, loudness, and pitch movement. Also, silent pauses between frames are indicated. The second input to phonetic spellout consists of triples of onset, nucleus, and coda, which are to fill the frames. These triples are produced by the above step 2 procedures and, when necessary, cluster composition (subsection 9.3.4). One may wonder again, as we did in section 9.6, whether it is strictly necessary that all segments and clusters be explicitly labeled in terms of syllabic function before they can be used as fillers. But we will not pursue this issue again.

Each address frame that is filled by the appropriate syllable constituents forms an address where the "standard" phonetic plan for that syllable can be found. The actual phonetic plan results from the imposition of the prosodic parameters that were attached to the address frame.

These parameter settings are responsible for some characteristic phenomena of fast speech. When a syllable's duration is set to be very short, as may happen in fast speech, vowel length will be reduced rather more than consonant length in the phonetic plan (see subsection 10.2.4). This vowel reduction can, in extreme cases, annihilate the syllable's syllabicity, as we observed in reductions such as *p'tato* and *t'mato*. When durations are short, the phonetic plan will also show more temporal overlap of adjacent phones. Take the word *cue*, where the vowel requires the articulatory feature of lip rounding. In slow speech, that lip rounding can be realized in the course of the diphthong. In fast speech, however, the lip rounding will have to start with the initial consonant in order to be realized at all. In other words, there is, to some extent, coarticulation of the onset consonant and the following vowel. A syllable's phonetic plan specifies at what moments various articulatory gestures are to be initiated. These temporal relations are crucially dependent on the syllable's duration.

It is quite probably the case that phonetic spellout is organized per phonological word. Though syllable frames are successively filled by their triples, as they become available, it is likely that the whole phonetic plan for a phonological word is collected before it is delivered to the Articulator. In other words, the Articulator cannot start pronouncing a phonological word's first syllable if it has not received the whole phonetic word plan. Meyer's (1988) evidence in support of this assumption was mentioned in chapter 9: If the Articulator were to begin as soon as a (phonological) word's first syllable is made available, a word's first syllable would be a good prime, but its first syllable plus part of its second syllable

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

would not. Meyer, however, found that a word like *pitfall* was primed more strongly by *pitf* than by *pit*. Similarly a word like *pedagogue* was more strongly primed by *peda* than by *pe*. If the phonetic spellout is not delivered until all of a phonological word's syllables have been planned, one would further expect that phonological encoding takes more time when there are more syllables in a word. This in fact seems to be the case. (The evidence, collected by Klapp and his co-workers, will be reviewed in section 11.1, where we will consider the interfacing of phonological encoding and articulation.)

This section has discussed some aspects of word-form planning in the context of connected speech. The obvious dependency of spoken words on context and rate made it necessary to consider mechanisms that adapt spelled-out stored forms to their contexts of occurrence. Such mechanisms help to create fluently pronounceable phonetic plans in connected speech. The Prosody Generator plays a central role in this adaptation by creating phonological words and by setting appropriate parameters for phonetic spellout. Phonetic spellout is probably made available to the Articulator per phonological word.

Summary

This chapter reviewed the speaker's phonological encoding of connected speech. It began with a sketch of the processing architecture underlying the phonetic planning of connected speech (figure 10.1). A Prosody Generator that interacts with the spellout procedures introduced in the previous chapter was introduced. It accepts various kinds of input. There is, first, surface phrase-structural and pitch-accent information, which is relevant for prosodic planning. There is, second, the metrical spellout, on which metrical planning is based. There is, third, executive, attitudinal, and emotional input, which affects such aspects of phonetic planning as speaking rate, intentional pausing, general loudness level, tune, and tone. Finally, there is input from segmental spellout, which can be modified by the Prosody Generator to create new phonological words.

The Prosody Generator produces two kinds of output. The first kind is a string of address frames for phonetic spellout. Each address frame is enriched by parameters for the syllable's set duration, its loudness, its pitch movement, and (may be) its precision of articulation. The second kind of output consists of the fillers for these address frames: segments that can fill onset, nucleus, or coda slots. Phonetic spellout then consists in the re-

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

trieval of stored phonetic syllable plans and their subsequent parametrization for duration, loudness, etc. As soon as all syllables for a phonological word have been planned, the Articulator can take over.

A major task for the Prosody Generator is the generation of rhythm. It generates phonological words, phonological phrases, and intonational phrases. I went a long way to show that the Prosody Generator needs little lookahead to create these structures. It can, mostly, be done incrementally. This means that it is not necessary to buffer more than one or two successive lexical elements in order to assign them their appropriate metrical weight. But if more buffering is possible—such as at lower speaking rates—the resulting metrical pattern can become more euphonious than if buffering is limited. I discussed, in particular, how new phonological words are generated by cliticization. I also outlined basic procedures for the generation of phonological and intonational phrases. And I considered these procedures in the alternative framework of metrical grids. The section on the planning of rhythm was then completed by reviewing what happens to segment and syllable durations in different word and phrase contexts, and by reviewing the shaky evidence for isochrony as it was originally defined. What there is in terms of isochrony, I argued, is due to metrical euphony rules rather than to stretching and shrinking of unstressed syllables.

Another main task of the Prosody Generator is to compute a pitch contour for the utterance. This can also be done without much preview, both for the more global and for the more local aspects of pitch. Among the more global aspects are declination, key, and register. I discussed, in particular, the counterintuitive suggestion in the literature that the speaker adapts his slope of declination to the length of the intonational phrase. There is no convincing ground for accepting this as a fact. For the more local phenomena of tune and tone, the state of affairs is very similar to metrical planning. Very little preview or buffering is required to assign an appropriate pitch curve to an intonational phrase. But with more lookahead more melodic lines can result, such as hat patterns and inclinations.

The final section returned to the connected speech phenomena of encliticization and assimilation. The generation of word forms in context causes striking deviations from the words' citation forms. New phonological words are formed out of pairs or triples of lexical items, and all sorts of accommodations can arise at boundaries between phonological words. The mechanism responsible for these contextual adaptations—a system of

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9

rules operating on the output of segmental spellout—was outlined. The rules can enforce the merging and resyllabification of items, depending on the phrase-structural relations between them. It is this restructured material that forms the input to phonetic spellout. The eventual result is a phonetic plan for connected speech.

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb.08442. Accessed 19 Aug 2020.
Downloaded on behalf of 76.119.101.9