Chapter 9

Generating Phonetic Plans for Words

How does a speaker generate the phonetic form of a word, given the developing surface structure? This chapter will characterize the process of phonetic planning as the spelling out of stored form representations and their projection on pronounceable syllables. The stored representations involve, in particular, the morphological, metrical, and segmental composition of words. A major task of phonological encoding is to generate a string of syllables that the Articulator can accept and pronounce. Syllables are basic units of articulatory execution. As was outlined in the preceding chapter, they consist of phones which, when executed, are complex and temporally overlapping articulatory gestures. The adult speaker, we conjecture, has an inventory of syllables. They need not be generated from scratch over and again. Rather, these stored articulatory patterns are addressed during phonological encoding on the basis of the spelled-out word representations. Also, certain free parameters are set, such as a syllable's duration, stress, and pitch. The eventual phonetic plan is a string of such specified syllables.

This chapter deals with the phonological encoding of single words. There is a certain drawback to this: It may seem as if phonological encoding is a wasteful process. Spelling out a word's segmental makeup also makes available the stored syllabification of the word, i.e., the segments' abstract grouping in syllables and syllable constituents. Why then should there be a second phase where strings of segments are used to address stored syllable representations? The main reason for this seemingly roundabout way of phonetic planning is to be found in the generation of connected speech. A word's stored syllabification is not sacrosanct. In connected speech, words often form coalitions with their neighbors that lead to so-called resyllabification. A phrase like *I gave it*, for instance, is easily resyllabified as *I ga-vit*. This enhances the fluency of articulation. To make an optimally pronounceable phonetic plan, the Phonological Encoder needs the segmental spellout of adjacent words. And it will often come up with a different syllabification than what is stored in each word's form code. In the present chapter, however, we will by necessity look only at cases in which the word's stored syllabification corresponds precisely to the syllable pattern in the eventual phonetic plan. Resyllabification in connected speech will be addressed in the next chapter.

The mission of the present chapter is to show that the phonetic plans for words are not stored and retrieved as ready-made wholes. Rather, they result from accessing and spelling out stored multi-level representations and using these to address syllable programs. This spelling out and addressing can occasionally become confused, leading to slips of the tongue, which are as revealing for the scientist as they can be painful for the speaker.

The chapter begins with a treatment of the tip-of-the-tongue phenomenon, the tantalizing situation of almost retrieving a much-wanted word. Bits and pieces of the word form become available, but not the whole thing. It is a highly slowed down version of word-form access, and it is often revealing of the underlying processes.

In section 9.2, phonological encoding is described as involving three major levels of processing: using lemmas to retrieve a word's morphemes and metrical structure, using morphemes to access a word's syllables and segments, and using segments and clusters to address stored phonetic syllable plans. These three levels will be called *morphological/metrical spellout*, *segmental spellout*, and *phonetic spellout*. And each of these levels can be a source of characteristic speech errors.

This is further elaborated in section 9.3, which considers in a systematic way what units can and what units cannot be involved in sublexical errors. This will lead to a further refinement of our three-level analysis of phonological encoding. We will then turn to a more systematic processing account of what can happen to such units. They can be added, omitted, exchanged, and so on, and each of these phenomena should eventually be understood as a derailment of processes that underlie normal, undisturbed phonological encoding.

There are two major accounts of the causation of sublexical form errors in speech. They will be reviewed in the final sections of this chapter. In most respects the two theories are complementary rather than competitive. The error account of Shattuck-Hufnagel's (1979) slots-and-fillers theory, on which the present chapter is based to a large extent, will be given in section 9.4. The activation-spreading account—in particular, Dell's (1986, 1988)—follows in section 9.5. A short final section addresses the issue of serial ordering in these two theories.

9.1 The Tip-of-the-Tongue Phenomenon

The form of a word is usually easily activated when its lemma is accessed, but there are comical or embarrassing cases of speech need where the transition from lemma to sound form is hampered. This is known as the *tip-of-the-tongue* (TOT) phenomenon. James (1893) considered it, but Brown and McNeill (1966) were the first to study TOT states experimentally. They gave their subjects dictionary definitions of moderately unusual objects—for example,

> a navigational instrument used in measuring angular distances, especially the altitude of sun, moon, and stars at sea.

The subjects had to retrieve the name of the object. Some subjects knew the instrument's name immediately; others could not remember it at all. But some felt that they knew it and that they were on the verge of producing the word. These subjects, who were in the TOT state, were asked to guess the initial letter and the number of syllables, to mention the words that had come to mind, and so on. For the above example, subjects tended to guess /s/ as the initial phoneme and two as the number of syllables, and sound-related words like *secant* and *sextet* had come to mind (meaning-related words, e.g., *compass*, also occurred). Apparently there is much lexical-form information available in the TOT state. (The target word was *sextant*.)

In the Brown-McNeill experiment, and in later, more extensive replications (Gardiner, Craik, Bleasdale 1973; Yarmey 1973; Koriat and Lieblich 1974, 1977; Rubin 1975; Browman 1978; Reason and Lucas 1984; Kohn, Wingfield, Menn, Goodglass, Berko-Gleason and Hyde 1987; Priller and Mittenecker 1988), it was found that in about 60–70 percent of the cases the first phoneme or cluster was correctly guessed, the middle part of the word was more error-prone, and subjects did better again on the final segment. The number of syllables was correctly guessed in 60–80 percent of the TOT states, and the subject usually knew which syllable was stressed. There is, in short, a partial activation of the word-form representation, involving the word's metrical structure as well as its initial and sometimes its final segment. But a full spelling out of the word's segments is blocked.

Jones and Langford (1987) were able to induce such blocking by giving subjects a "blocking word" after the definition. (If *sextant* were the target word, *secant* would be a good blocker.) When the blocker was given right after the definition, there was an increased chance of the subject's entering a TOT state. It was irrelevant whether the blocker was a high- or a low-frequency word; only its phonological similarity to the target word mattered. Also irrelevant were semantic blockers (e.g., *compass* if the target word were *sextant*). This supports the notion that in the TOT state there is no search for the lemma; it has already been retrieved on semantic grounds. What fails is full access to the form information. A phonological blocker further "misguides" this search.

These experiments show that the lexical-form representation is not all-or-none. A word's representation in memory consists of components that are relatively accessible, and there can be metrical information about the number and accents of syllables without these syllables' being available. Jones and Langford call this a "word sketch." Let us see what this initial sketch might look like.

9.2 Frames, Slots, Fillers, and Levels of Processing

Speech errors provide ample evidence for the independent availability of word sketches or frames and of the elements that are to fill them. Take segment exchanges, such as *I* s*ould be* sh*eeing him soon* (from Shattuck-Hufnagel 1979). When the speaker tried to access the first phoneme of *should*, /ʃ/, the first phoneme of *seeing*, /s/, intruded. The /s/ of *seeing* must already have been available as a word-initial segment candidate when *should* was being generated. After the mis-selection, however, the speaker did not say *I sould be eeing*. The fact that the initial phoneme of *seeing* had already been used has in no way removed the word-initial slot of *seeing*. It persists and is filled by the still-available word-initial candidate /ʃ/. There are frames with positions for morphemes, phonemes, or other elements; during speech these frames are filled with candidate elements. In the following it will be argued that the frame is an address template for a procedure or subroutine. Once a frame is filled, the address is complete and the procedure can be identified and executed. This is just another case of *productions* in Newell's sense, discussed in chapter 1: IF the filled address frame is such-and-such, THEN spell out the corresponding form information.

Let us consider three examples to clarify this notion. They are taken from three major levels of processing in word-form generation; together they give a first sketch of the system that subserves phonetic planning. The three levels of processing are *morphological/metrical spellout*, *segmental spellout*, and *phonetic spellout*.

9.2.1 Morphological/Metrical Spellout

Morphological/metrical spellout is a procedure that takes lemmas and their diacritical parameters or features as input and makes available both the morphological and the metrical composition of a word. The example involves addressing the correct inflectional form of a verb. Remember that, in English surface structure, verb lemmas have diacritical parameters for number, person, and tense, among other things. There is, for instance, one lexical entry for the verb *eat*, but it contains various lexical items, such as *eat*, *eats*, *ate*, and so on (see subsection 6.1.2). The diacritical parameters serve to select the correct item within the verb's lexical entry. How does this come about?

The form address, we suppose, is a frame consisting of slots: one slot for the lemma as it appears in surface structure, and further slots for the diacritical parameters. More precise, the slot for the lemma is the form address to which the lemma points (see subsection 5.1.2). In subsection 6.2.1 the lemma's form address was represented by an arbitrary number. Here, however, for ease of identification, we will put the lemma's name in the address slot instead of a number. That name is, of course, not the word form to be retrieved. Thus (with the diacritical parameters limited to the three ones mentioned), the morphological and metrical forms for *segmented* and *knew* are stored under addresses such as these:

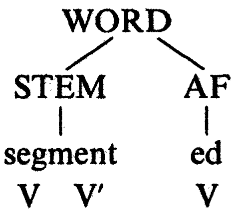
lemma	**number**	**person**	**tense**
segment	any	any	past

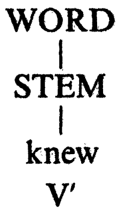
lemma	**number**	**person**	**tense**
know	any	any	past

In both cases the address *frame* is the following template:

lemma	**number**	**person**	**tense**

when a lemma of category V appears in surface structure, this address frame is automatically made available. The frame now has to be filled in order to become an address for a morphological/metrical spelling-out routine. When each slot is filled by an appropriate element, the lock is opened and the morphological and metrical spelling-out routines become available. The routines make available the morphological structure of the item and the number of syllabic peaks (if any) for each morpheme. Also, the peak that carries word accent is marked. For *segmented*, the procedures generate a stem (*segment*) and a suffix morpheme (*ed*) in that order, and they locate two syllable peaks in *segment* and one in *ed*. The second peak of *segment* is marked for word accent. For *knew*, the subroutines produce a single morpheme, with a single syllable peak which carries word accent. These results can be represented as follows:





In other words, the procedures unlock two sets of form information: the morphological representation and the basic metrical pattern of a word. It is, as yet, not strictly necessary to assume that these two kinds of information are always simultaneously retrieved. But that assumption is not critical for the sketch of phonological encoding to be developed in this chapter. The crucial point is that a word's metrical information becomes available at a very early stage, as is apparent from the tip-of-the-tongue studies. This metrical information is particularly important for the generation of connected speech, as will be seen in the next chapter.

The metrical information consists of the syllabicity status of each morpheme, i.e., the number of peaks it corresponds to at the skeletal tier, plus the stress distribution over peaks. Recall from subsection 8.1.6 that a word's basic metrical pattern can involve more than two levels of stress, as represented in the word's metrical grid. It is probably correct to assume that the full metrical pattern as it is stored in the lexicon becomes available at this stage. For ease of presentation, however, the present discussion will be limited to two levels of stress: one for the peak carrying word accent and one for all other peaks. Accented peaks will be indicated by V′, nonaccented ones by V.

At this level of spellout, only *stored* metrical information is made available. Pitch accent and contextually determined metrical properties of the word (subsection 8.2.2) are generated by what we will call the *Prosody Generator* in chapter 10.

No segmental information (subsection 8.1.5) is available at the present level of representation. It is only for ease of reference that the retrieved stems are written as *segment* and *knew*; we could as well have used numbers to refer to these entities. Similarly, the suffix is depicted as *ed*, but we could as well have written *pta* (for "past-tense affix").

When there are misselections in filling the address frame, the wrong address is composed and an inappropriate spellout routine is retrieved. In the speech error *that I'd hear one if I knew it* [intended was: *that I'd know one if I heard it*] (Garrett 1980b), the lemmas *hear* and *know* got exchanged by mechanisms discussed in chapter 7. The point here is that, in addressing the morphological/metrical spell-out of the first verb, the wrong lemma (*hear*) and the appropriate diacritical tense feature (present) were used as fillers. Together they precisely formed the address under which the morphology of the item *hear* was stored. Similarly, the pair of fillers (*know*, past) formed the key that opened the lock for the morphological/metrical spellout of *knew*, as in the above example.

The metrical spellout—i.e., the number of peaks for each word and their stress values—is crucial for the construction of address frames at the last, phonetic spellout level. It determines the number of syllables to be retrieved at that level (see subsection 9.2.3).

9.2.2 Segmental Spellout

The procedures for segmental spellout take the morphemic and metrical information as input and generate the segmental composition of the word. In the course of this, the segments' grouping in syllable constituents also becomes available.

The address frame of a syllabic spellout procedure contains morphological slots, such as for stem and affix. The fillers are morphemes together with their metrical information, as spelled out at the previous level of processing. Two examples are given below. They are the addresses and spellouts for the verb form *segmented* and for the plural noun *segments*.

address frame:

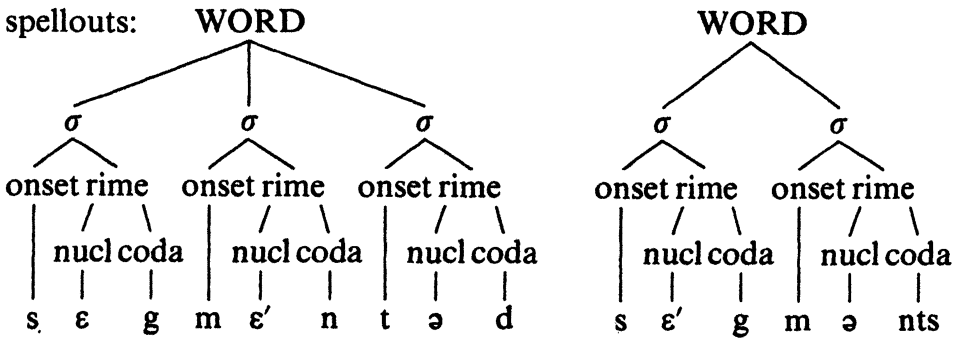
stem	**af**

addresses:

stem	**af**
segment V V′	ed V

stem	**af**
segment V′ V	s

spellouts:



The address frames at this level are made available on the basis of the spellouts of the previous level of processing. When the slots get filled with (respectively) the stem *segment* and the affix *ed*, the address is created for the spellout of *segmented.* The retrieved procedure yields three syllables, each parsed into onset, rime, nucleus, and coda—i.e., just the syllable tier representation of the word (subsection 8.1.4). In this example, most but not all of the onsets, nuclei, and codas are single phonemes. Only the last coda is a consonant cluster: /nts/. This needs further spelling out. There is, we will assume, a *cluster spellout procedure* that takes a cluster as input and yields the individual segments or phonemes as output. So, for the cluster /nts/ this can be depicted as follows:

$$/\mathrm{nts}/ \rightarrow \boxed{\begin{array}{c}\text{cluster}\\ \text{spellout}\end{array}} \rightarrow /\mathrm{n}/,\ /\mathrm{t}/,\ /\mathrm{s}/$$

This issue will be taken up in more detail in subsection 9.3.4.

It is probably the case that each spelled-out segment is categorized as onset, nucleus, or coda, or as part of a cluster (but see section 9.5). The reason, as we will see in subsection 9.2.4, is that in speech errors onsets can exchange with onsets, nuclei with nuclei, and codas with codas. But it is seldom that segments of different category exchange. Still, these categories can be revised in connected speech. In *I ga-vit*, the coda segment /v/ has become an onset segment.

Two further points should be noted. The first one is that the input morpheme boundaries are not preserved in the syllabic grouping of segments. The affixes *ed* and *s* are generated as parts of the syllables /təd/ and /mənts/. The second point is that the metrical information is now translated into nuclear phonemes. Compare the nuclei of the syllables /mɛn/ and /mənts/. They are different vowels at this level, /ɛ/ and /ə/, and this is a consequence of the difference in stress assignment.

What sort of addressing errors can be made at this level of processing? The critical fillers are different kinds of morphemes, such as stems and affixes. Errors occur when an inappropriate filler is used. In *take the freeze*s *out of the steak*er [*take the steaks out of the freezer*] (from Fromkin 1973), the stem of *steaks* combined with the affix of *freezer* in addressing a segmental spellout routine. As a result, *steaker* was produced. Subsequently, the leftover stem of *freezer* and affix of *steaks* were used to address a further segmental spellout routine. The output was the two-syllable *freezes*; i.e., the affix was given its context-dependent form /-ɪz/ instead of /-s/. Such a context-dependent form is called an *allomorph*. This shows that the spellout routine does make use of the fact that the misplaced *-s* affix of *steaks* is really nothing but an unspecified plural affix at the filler level (it is written as *s* only for convenience).

Another remarkable type of speech error that arises at the present level of processing is the stress error. Cutler (1980a) reports an extensive analysis of such errors. One of her examples is *I put things in that abstr*A*ct that I cannot justify*, where the noun *abstract* erroneously carries word accent on the second syllable. How could this error arise? Cutler showed that in such errors the lexical-stress placement is always that of a related word, another derivation of the same morpheme. For the erroneous noun *abstr*A*ct*, the related word is probably the verb *abstract*, which has stress on the second syllable; for the error *articul*A*tory*, the related word is *articulation*; and so on. This systematic relation can be explained by an erroneous choice of filler in addressing the segmental spellout routine. Take the above examples *segmEnted* and *sEgment.* Assume that a speaker is trying to say *Peter segmented two segments.* At some stage the morphological/metrical spellouts of *segmented* and of *segments* become available. To address the spellout routines for *segmented*, the speaker may erroneously choose the available morpheme

segment
V′ V

as the filler, together with the right past-tense affix. The spellout procedure then generates *sEg-men-ted.* It is not necessary for lexical-stress errors that the related word appear somewhere in the sentence. What is important is that the related word be somehow activated in the production process. How that can happen is the topic of subsection 9.5.2.

Before we turn to the next level of processing, notice that an English speaker may occasionally compose a well-formed address by completing a frame with the right kind of fillers and still fail to retrieve a word form. As was extensively discussed in subsection 6.1.3, English speakers do now and then produce new words. An example may have been the word *steaker* in the above speech error. This is surely a possible word in English, but it is unlikely that the speaker had it in store. Still, the speaker created a spellout on the spot. This is presumably done by analogy (Stemberger 1985b), an old, important, but still ill-understood notion in psychology and linguistics.

9.2.3 Phonetic Spellout

After retrieving a word's sequence of segments, the Phonological Encoder will use them to address phonetic plans for syllables. The phonetic plan for a syllable specifies the articulatory gesture to be executed by the Articula-

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb08442.0001.001.

tor. It can be characterized as a sequence of phones, but phones are not discrete nonoverlapping events. Rather, each phone is a temporal gesture itself, which typically overlaps in its execution with other phone gestures in the syllable (Browman and Goldstein 1986). Moreover, the dynamic properties of a phone depend substantially on where it appears in the syllable, and on the other phones the syllable is composed of. In other words, phones in a syllable's phonetic plan are always *allophones* (subsection 8.1.5), context-dependent realizations of phonological segments or phonemes.

It is very likely that the skilled language user has an inventory of syllable plans, a stock of frequently used motor programs. Phonetic spellout will, then, consist largely of retrieving these syllable programs. This subsection discusses how these programs can be addressed, following a notion developed by Crompton (1982). Before we turn to that, two things must be noted. The first is that stored syllable programs are not completely fixed. A syllable can be pronounced with more or less force, shorter or longer duration, different kinds of pitch movement, and so on. These are free parameters, which have to be set from case to case. This issue will be taken up in the next chapter. Second, it is probably not so that *all* of a language's possible syllables are stored in the speaker's mind. New formations are certainly possible, and we will return to that issue below.

Syllable plans have addresses, and the first step of phonetic spellout is to compose the appropriate address. An address, we assume, consists of three slots: one for an onset, one for a nucleus, and one for a coda. These slots will, one after another, have to be filled by appropriate subsequent segments as these become available from segmental spellout. Slots can also accept clusters. How clusters are formed to become fillers for slots will be taken up in subsection 9.3.4. Let us now work out these notions by way of an example.

The example concerns addressing the phonetic plans for the first syllables of the words *segmented* and *hemisphere*: *seg* and *hem*. Their addresses result from filling the following address frames:

onset	**nucl**	**coda**

onset	**nucl′**	**coda**

Where do these address frames come from? It was mentioned earlier that the metrical output of the first spellout level would be crucial input for the construction of addresses at the phonetic spellout level. The procedure can be quite simple: For each peak of metrical spellout, an address frame is triggered that contains precisely three slots—one for onset, one for nucleus, and one for coda. They come in two kinds, one for unstressed syllables and one for stressed syllables. This corresponds to the two kinds of peak in metrical spellout. The left address frame above is for an unstressed syllable, the right one for a stressed syllable.

In the segmental spellout of *seg* (see above), the first phoneme segment, /s/, was categorized as syllable onset; the second segment, /ɛ/, as unstressed nucleus; and the third segment, /g/, as coda. This is precisely the triple of fillers matching the slots in the left address frame:

onset	**nucl**	**coda**
/s/	/ɛ/	/g/

The address is now complete, and the syllable's phonetic plan can be retrieved. It consists of the syllable-specific allophones [s], [ɛ], and [g], and the whole phonetic plan for the syllable can be written as the spellout

[sɛg]

For *hem* of *hemisphere* the story is analogus. The segmental spellout gave three phonemes: /h/, /ɛ'/, and /m/, for onset, nucleus, and coda. In combination they are the adequate fillers for the right address frame above. The completed address for the syllable retrieved is

onset	**nucl'**	**coda**
/h/	/ɛ/'	/m/

and the phonetic spellout is

[hɛ'm].

Note that the nuclear phone is marked for accent. In articulatory terms this may mean longer duration, more amplitude, or pitch movement.

When, by chance, an inappropriate filler is made available, an addressing error may occur. This happened when the speech error h*eft* l*emisphere* [*left hemisphere*] was made (Fromkin 1973). Here the syllabic spell-outs for both lemmas had become available. When a syllable onset had to be specified for addressing the articulatory subroutine for the first syllable, the segment /h/ was apparently more strongly activated than the segment /l/. Both were of the correct category (onset). Together with the available fillers for nucleus and coda, the syllable *heft* was erroneously addressed and the phonetic syllable plan [hɛ'ft] was activated. Similarly, when the next syllable was programmed, an onset filler was needed. The /h/ had been used, but the /l/ was still available and of the right category. Together with /ɛ'/ and /m/, it filled the address frame to cue the syllable *lem*, releasing the articulatory plan [lɛ'm].

Also for this level of spelling-out one should wonder whether the native speaker of English will occasionally be productive. That is, will an adult speaker occasionally produce a well-formed but nonstored syllable, i.e., one he never uttered before? A rough count of the number of different syllables in English (with thanks to Hans Kerkman) yielded a number of about 6,600 over 38,000 word types. Hence, the number of English syllable types is relatively small, and could be easily stored in the speaker's lexicon. And other languages, such as Japanese, probably have much smaller numbers of different syllables. Still, some of these syllables may be quite infrequent, arising only in unusual morphemic combinations (such as in *infarct-s*, which involves the syllable [fɑrkts]). Speakers are probably able to produce new but well-formed kinds of syllables by analogy. This shouldn't surprise us, because this is presumably the way they acquired their syllable repertoire to start with. The mechanism of such new formations, however, is unknown.

A phonetic spellout mechanism, such as that proposed here (following Crompton 1982), will handle much but not all allophonic variation in language production. The different phonetic realizations of a segment or phoneme are, in large part, dependent on the different syllable environments in which it appears. But there is context beyond the syllable which may also affect a segment's phonetic realization. Moreover, a syllable's stress and pitch properties are largely contextually determined. We will defer discussion of these sources of variation to chapter 10, which deals with phonetic planning in connected speech.

9.2.4 The Unit-Similarity Constraint

All three of the examples above have shown how speech errors may arise when inappropriate fillers are made available in the addressing of spelling-out procedures. In the following we will speak of the *target* (i.e., the appropriate filler) and the *intrusion* (the inappropriate filler). What is a possible intrusion for an address slot?

The three levels of spellout—morphological/metrical, segmental, and phonetic—have provided us with different kinds of slot fillers: lemmas and diacritic features at the first level, morphemes (roots, stems, and affixes) at the second level, and syllable constituents (onsets, nuclei, and codas) at the level of phonetic spellout. Each slot required a filler of its own category. An affix cannot fill a stem slot, an onset cannot fill a coda slot, and so on. A filler must have the right password for a slot: the filler's category at the relevant level or tier of representation. As a consequence, targets and intrusions obey the following principle (Shattuck-Hufnagel 1979):

Unit-Similarity Constraint The intruding element is of the same level of representation and category as the target element.

Lemmas exchange with lemmas, stems with stems, affixes with affixes, syllable onsets with syllable onsets, and so forth. But stems or roots never exchange with affixes, affixes do not exchange with onsets, and onsets hardly ever exchange with codas.

The three levels of processing considered so far cover a major part of a word's form-generating mechanism. But much detail is still to be provided. Before turning to an analysis of the system's control structure, we will consider whether there are more or other target-intrusion pairs. The determination of what kinds of units are displaceable can help us find out what kinds of fillers are needed in addressing the form-generating routines. So far, some anecdotal evidence has been provided for lemmas, diacritical features, stems, affixes, and syllable constituents as fillers. But what about syllables, distinctive features, or other potential fillers? The next section reviews what can and what cannot be mislocated in speech errors.

9.3 Substitutable Sublexical Units

Fromkin (1971) conjectured that almost any linguistically defined sublexical unit or feature could be subject to substitution in speech errors. Later research, however, showed that, though this may be true as a general characterization, certain kinds of speech error are exceedingly rare (Shattuck-Hufnagel and Klatt 1979; Stemberger 1983a; Dell 1986). Other experimental work has added to a further specification of what is replaceable and not (Treiman 1983, 1984; Levitt and Healy 1985). Let us review some of the main sublexical units.

9.3.1 Morphemes

All the morpheme types discussed in subsection 8.1.1 can be displaced in speech errors. Stems and whole words are exchangeable, as in nam*ing a* wear *tag* [wearing a name tag]; or a stem and a root can exchange, as in *I hate* rain*ing on a* hitch*y day* [*I hate hitching on a rainy day*] (both from Shattuck-Hufnagel 1979). So, in view of the Unit-Similarity Constraint, whole words, stems, and roots are of similar filler category. But units of this category almost never exchange with affixes. Affixes are a displaceable category of their own. They can be anticipated, as in *people read the backs*es *of boxes* [*people read the backs of boxes*] (Shattuck-Hufnagel 1979). They can be persevered, as in *Ministers in the church*es [*Ministers in the church*] (Fromkin 1971). Notice in the latter example that perseveration of the affix does not create *church*s, but *churches*; i.e., the correct allomorph is produced. This shows that the shifted unit is an affix, not just the phoneme /s/.

Morpheme errors can be caused either at the level of morphological spellout or at the level of segmental spellout. The earlier example *I'd hear one if I knew it* [*I'd know one if I heard it*] is caused at the higher level, owing to an erroneous combination of lemma and diacritical feature in addressing the spellout routine. But it is unlikely that *raining on a hitchy day* arose at this level, since the two lemmas involved (*rain* and *hitch*) are of different grammatical categories. This violates the Unit-Similarity Constraint at that level of processing. Exchanging lemmas, we saw in chapter 7, are usually of the same syntactic category. It is more likely that the latter error is due to addressing failure at the level of segmental spellout, where stem and root are exchangeable. For *church* → *churches* the level is undecidable, but other affix errors are unambiguous. The perseveration *they needed to be made*d [*they needed to be made*] (Shattuck-Hufnagel 1979) can only be due to an addressing failure in segmental spellout. At the earlier level, *make* plus "past" had successfully triggered the routine that produces the stem *made*. In syllabic spellout, this stem combined with the persevered *-ed* affix of *needed* to produce the two-syllable form *maded*. It is therefore important in the analysis of speech errors (and of word-form production in general) to distinguish diacritical features such as "past", "plural", and "third person" from the affixes they induce.

9.3.2 Syllables

There are occasional reports of syllable replacements in the speech-error literature. Of course, all substitutions by monosyllabic morphemes are at the same time syllable replacements. The test case, however, is whether a single syllable of a multi-syllabic morpheme can be moved or replaced. Shattuck-Hufnagel (1979) presented as an example ca*ssy* pu*t* [*pussy cat*], where the syllable *pu* was moved. But she noted that such errors are highly exceptional. Dell (1986) made the same observation, and explained the exceptions in terms of a coincidence of replacements of smaller, subsyllabic units. (In the above example, this would have been syllable onset /p/ and syllable nucleus /u/.)

The sheer absence of pure syllable substitutions is especially remarkable since syllables clearly play an important role in speech errors and in fluent speech generally. MacKay (1972), for instance, observed that word blends often respect syllable boundaries, as in Wells's (1951) example *be-hortment* [*behavior/deportment*]. Fujimura and Lovins (1978) argued that syllables are the smallest relatively invariant articulatory units in speech production, and I concur with that view. But syllables are, apparently, never themselves fillers for address slots. This is in full agreement with the three-level model sketched above. At none of the three levels are fillers of the category "syllable" required to compose an address.

9.3.3 Syllable Constituents

In chapter 8 we distinguished between syllable onset and rime, and within rime between nucleus and coda. Onsets and rimes can be independently involved in speech errors. MacKay (1972) observed that in word blends breaks were more likely to occur before a syllable's vowel than after it; thus, *gr-astly* [*grizzly/ghastly*] should be more common than *mai-stly* [*mainly/mostly*] (both examples from Fromkin 1971). Syllable onsets do move as a whole in speech errors, whether they are single phonemes or consonant clusters; note f*ace* sp*ood* [*space food*] (from Fromkin 1971).

Rimes can move as well: *f*art *very h*ide [*fight very hard*] (from Fromkin 1971). The latter, however, is a rather infrequent type of speech error. The two rime parts, nucleus and coda, can also be replaced as units, although they have a stronger tendency to stick together (Shattuck-Hufnagel 1983; Stemberger 1983a). An example of nucleus movement is *cl*ea*p p*i*k* [*clip peak*]; coda substitution can be seen in *do a one ste*tch – *step switch* (both from Fromkin 1971).

An extensive analysis of more than 300 spontaneous speech errors involving the nucleus was reported in Shattuck-Hufnagel 1986. A major finding was that 79 percent of the errors occurred between vowels in stressed syllables, as in *the debote f*eik – *debate focuses on*, where both the target (here /eɪ/ and the intrusion (here /oʊ/) belonged to a stressed syllable. "Mixed" cases were rare, and the intruding vowel was always accommodated to the stress of the target syllable. Remember that the phonetic spellout mechanism created different addresses for stressed and unstressed syllables. A slot for a stressed nucleus will normally accept only a stressed vowel, and vice versa for a non-stress slot. This explains the low number of mixed cases in the data. It is just another demonstration of the Unit-Similarity Constraint. It does not, of course, explain the fact that errors involving two unstressed vowels are quite rare (as in *Buffo the – Byffy the Buffalo*, where unstressed /i/ and /oʊ/ are involved as target and intrusion). Stressed vowels are far more error-prone than unstressed ones.

A systematic experimental study of the movability of syllable constituents was done by Treiman (1983, 1984). She gave subjects different word games to play involving different types of syllables. In one game, subjects had to make two new syllables out of one. They would listen to four training examples and repeat each of them, e.g.

kig → kaz ig
buf → baz uf
tep → taz ep
nol → naz ol.

They were then asked to do the same to a set of test stimuli. A critical item could now be the syllable *skef*. What would a subject do? Either of two responses would be consistent with the training examples: *skef → skaz ef*, which separates onset and rime, and *skef → saz kef*, which separates the initial consonant from the rest. Only the first response type respects syllable constituents (onset, rime), and almost all responses were of that type, i.e., preserving the syllable's onset cluster.

Other syllable games tested the integrity of nucleus and coda. In one of these, subjects had to learn the syllable splitting again, but this time they were left no choice. The splitting had to go either as in

(1) isk → it ask

or as in

(2) isk → ist ak.

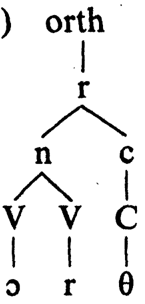
In type 1, nucleus and coda are left intact; however, in type 2 the break is made not between nucleus and coda but within the coda before the final consonant. Subjects made many more errors on type 2 than on type 1. Also, type 2 took many more trials to learn. It is apparently easier to leave the coda intact. Given this finding, Treiman could test cases where linguists disagree on where the boundary is between nucleus and coda. Remember, for instance, how *meter* was analyzed in subsection 8.1.4. There the so-called *liquid* segment/r/ behaved as a high-sonorous peak segment, i.e., as a V at the segmental tier. Now take a word like *bird* or *earth*. Is /r/ still a V, belonging to the nucleus, or is it rather a C-part of the coda? What would it do in Treiman's game? She tested such cases by examples of types 3 and 4.

(3) orth → ot irth

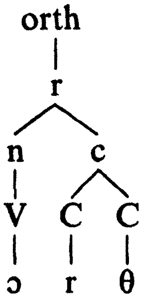
(4) orth → ort ith

Subjects found type 4 much easier than type 3. This pleads for the notion that the liquid is part of the nucleus rather than part of the coda; i.e., analysis i is preferred over ii:

(i)



(ii)



Games with nasals (e.g., *imf* → *iz amf*) gave results that were intermediary between these games with liquids and the original games 1 and 2 with obstruents (like /t/). These experimental results are in fine agreement with Stemberger's (1983b) and Shattuck-Hufnagel's (1986) finding that also in speech errors the liquids /r/ and /l/ tend to stick to the immediately preceding vowel, as in *cheeps 'n twirts* [for *chirps 'n tweets*]. In other words, they behave as parts of a displacing nucleus.

There is, finally, one syllable constituent that needs special attention: the *null-element*, be it a null-onset or a null-coda. On the surface, a syllable like *orth* or *art* doesn't seem to have an onset; it seems to have only a rime. Still, it has been suggested that there is an onset, though an empty one. Similarly, the syllable *spa* lacks a coda, but can be analyzed as having a null-coda. The main argument for these analyses is that the null-element can be anticipated or exchanged in speech errors. Shattuck-Hufnagel, discussing this matter (1979, 1983), gave examples such as *Doctor -inclair has emphasized* [*Sinclair*], where the empty onset of *emphasized* is already anticipated when the onset of *Sinclair* is prepared. Or is this simply an onset or phoneme deletion, without any relation to *emphasized*? Only careful statistical analysis can provide a decisive answer. Dell (1986) has presented statistical evidence that the null-element is not always involved in onset or coda omissions, but it is probably still needed for a complete account of the speech-error data.

In summary: There is good evidence that syllable constituents can function as fillers, and that their constituent category is crucial: Onsets exchange with onsets, rimes with rimes, nuclei with nuclei, and codas with codas, all in agreement with the Unit-Similarity Constraint.

9.3.4 Segments

Some two-thirds of sublexical speech errors involve single segments, either consonants or vowels (Shattuck-Hufnagel 1982). Many of these are, of course, just syllable onsets, nuclei, or codas. But there are still many errors where individual phonemes in onset or coda clusters are in trouble. This is, for instance, the case for the exchange error p*eel like* fl*aying* [*feel like playing*] (Shattuck-Hufnagel 1982), where the phoneme /p/, but not the whole onset cluster /pl/, is mislocated. Fromkin (1973) presents an impressive list of such divisions of consonant clusters. One of her cases is bl*ake* f*ruid*, where the liquids /r/ and /l/ are exchanged, leaving the rest of the syllable onsets in place.

These observations are, so far, not covered by the three-level model of section 9.2. It can handle phoneme exchanges as long as these phonemes are single syllable constituents themselves, as in h*eft* l*emisphere*. To explain errors of the above kind, however, a mechanism of *cluster composition* is required. Such a mechanism, moreover, is needed if phonetic spellout is to work. Remember that the address frames at that level contain three slots: one for syllable onset, one for the nucleus, and one for the coda. In order to retrieve a syllable like *brake*, one needs for the onset slot a cluster, /br/, not a single phoneme segment. Hence, there must be a way to build such clusters from consecutive segments spelled out at the previous level.

How can cluster composition be modeled? The present proposal is to treat it just like phonetic spellout. That is, we will assume that the speaker has an inventory of onset clusters. They are the small set of onset clusters that are phonotactically allowed in the language, such as /st/, /br/, /fl/, and /skr/ in English. Similarly, there will be a set of phonotactically possible coda clusters. A cluster can be addressed by filling the slots of a corresponding address frame. Let us work this out in some more detail for syllable-onset clusters, which are most prone to speech error.

Take the word *groom*. At the level of segmental spellout the sequence of segments /g/, /r/, /u/, /m/ is produced. At the next level they will be used to complete the onset, nucleus, and coda slots of a syllable address. What should go into the onset slot? Not just /g/, but the cluster /gr/. In subsection 8.1.4 we discussed a rule called "maximization of onset." The onset cluster will be made as large as phonotactic rules allow. Onset-cluster composition consists of collecting as many consecutive C-type ("consonantal") segments as possible and filling an address frame with the same number of slots. (When there is just one consonant, the resulting "cluster" will be a singleton.) In the example there are two consecutive consonants, /g/ and /r/. A two-slot address frame is made available and is filled with these elements. If the result is a phonotactically possible address, the corresponding cluster becomes available for insertion in the onset slot. This can be depicted as follows:

address frame:

onset C_1	**onset C_2**

address:

onset C_1	**onset C_2**
/g/	/r/

retrieved onset cluster: /gr/

For other pairs of segments no cluster will be found. For instance, if the slots were to be filled with /n/ and /t/, respectively, no cluster would be retrieved; /nt/ is not a possible onset cluster in English.

What happens if such a sequence arises by chance? Will it be pronounced? The answer is clearly No. Take Fromkin's (1971) example f*lay the* p*ictor* [*play the victor*]. The mislocated /v/ is devoiced when it appears in front of *-lay*, thus turning into /f/. There is a still-unresolved discussion in the literature on how this is caused (Fromkin 1971; Davidson-Nielson 1975; Garrett 1980a; Crompton 1982; Stemberger 1983a; Stemberger and Treiman 1986), but somewhere the segment pair (/v/, /l/) has to be turned into the phonotactically possible onset cluster /fl/, which can then be used as filler for an onset slot in phonetic spellout. A similar problem arises in explaining speech errors such as *we have a big* s*tellar down there* [*cellar*] (Stemberger 1983a). It is likely that here the /d/ of *down* in anticipated in pronouncing the target word *cellar*, and that this results in the onset cluster /st/ rather than /sd/. The generation of these cases may develop as follows.

address frame:

onset C_1	**onset C_2**

addresses:

onset C_1	**onset C_2**
/F/	/l/

onset C_1	**onset C_2**
/s/	/T/

retrieved onset cluster: /fl/ /st/

The /F/ in the leftmost address slot stands for either /v/ or /f/. When either of these is followed by /l/, a legal address is formed, namely that of the cluster /fl/. In other words, the feature of voicing is irrelevant for the potential filler of this slot if the next element is /l/. Similarly, the /T/ in the rightmost address slot means either /t/ or /d/. If either of these is preceded by /s/, a legal address is created for the cluster /st/. Similarly, the legal onset cluster /sp/ may be retrieved when the pair (/s/, /P/) is inserted in the address frame. Here /P/ stands for either /b/ or /p/. This may be at the base of speech errors such as *benefit spall – small businesses* (Stemberger 1983a). In this error the /b/ of *businesses* is anticipated and fills the C_2 slot of the cluster. However, the addresses (/s/, /b/) and (/s/, /p/) are not distinctive; they both refer to the cluster /sp/, /sb/ being phonotactically impossible as an onset cluster in English. Segment specifications such as /P/, /F/, /T/, and /K/ (for /k/ and /g/) are sometimes called *archiphonemes*. They are phonemes unspecified for at least one feature.

If cluster composition can involve archiphonemic segment specifications, the obvious next question is whether archiphonemes may arise in the *spellout* of clusters at the previous level. For instance, when the word *still* is segmentally spelled out, will the onset cluster be spelled out as /s/, /t/, or rather as /s/, /T/, where /T/ is the archiphoneme unspecified for voicing? The latter would normally suffice, because at the next level of cluster composition, where the cluster /st/ is to be retrieved, nothing more is required than the pair of fillers /s/, /T/ (at least, if the above proposal is correct).

Stemberger (1982, 1983a), who raised this question, gave an affirmative answer. There is evidence from speech errors to support the idea that archiphonemes arise in cluster spellout. One case involves the word *scruffy*. According to the above analysis, its onset cluster will be spelled out as /s/, /K/, /r/. Here /K/ is an archiphoneme with the voicing feature unspecified, i.e., it stands for both /k/ and /g/. At the next stage of cluster composition, this triple /s/, /K/, /r/ will suffice for the retrieval of /skr/; there is no possible onset cluster /sgr/ in English. But what happens if, by accident, the initial segment /s/ disappears? Both /kr/ and /gr/ are possible onset clusters in English, as in *crazy* and *grasp*. If indeed /K/ appears in the cluster spellout, not only can /k/ arise when the /s/ is lost, but /g/ is possible as well because the archisegment was not specified for the voicing feature. Stemberger (1983a) gave examples such as *in your really gruffy – scruffy clothes*, where indeed the /k/ of *scruffy* is realized as /g/ when the preceding /s/ happens to disappear. One wonders, of course, how perceivable such distinctions are.

Though the notion of archiphoneme is attractive for the analysis of such cases, it should be observed that an archiphoneme is just a phoneme on our definition in subsection 8.1.5. It is a segment specified for its *distinctive* features only. The segments /g/ and /k/ cannot be distinctively used in English when they follow /s/ in an onset cluster; hence, in that context they are the same phoneme.

The conclusion so far is that individual phonemes in onset clusters, whether or not they are syllable constituents themselves, can become misplaced in speech errors. The character of these misplacements has led us to assume a process of cluster composition that creates maximal onset clusters for syllables. These can then be inserted in the onset slots at the phonetic-spellout level. In the following, cluster composition will be considered as a preliminary or first stage of phonetic spellout.

The process of onset-cluster composition can err in several ways. An extensive analysis of word-initial cluster errors, both naturally occurring and experimentally elicited, was reported by Stemberger and Treiman (1986). A major finding of this study was that the second position in onset clusters is much more vulnerable to error than the first position. It is more likely to be lost or to be subject to substitution. Errors such as (5) and (6), which involve the second position, are more frequent than cases like (7) and (8), which involve the first position.

(5) They pace – place too little emphasis on their own results [loss of second consonant in onset cluster]

(6) prace – place Bresnan's arguments . . . [substitution of second consonant]

(7) Their attention pan – span is . . . [loss of first consonant]

(8) A crate – great quest . . . [substitution of first consonant]

Also, it is more likely that a second-position consonant becomes erroneously added than a first-position consonant. Example 9 is a more frequent type of error than example 10:

(9) Oh, so you have to bruy it with the T.V. [addition of second consonant]

(10) the same as the hit frate – hit rate for low-frequency items. [addition of first consonant]

We will not pursue Stemberger and Treiman's explanation of this asymmetry in detail here. But clearly, accurate filling of the C1 position in a cluster is given precedence in phonological encoding.

A similar mechanism of cluster composition can probably be suggested for codas. Here again, only phonotactically possible clusters are to be addressed. But the mechanism must differ in that there is no maximization rule for codas.

What about complex nuclei? Earlier we noted Stemberger's and Shattuck-Hufnagel's findings that a nucleus consisting of vowel plus liquid moves as a whole (an example was *cheeps and twirts*, where the whole complex nucleus /ɜr/ was displaced). Diphthongs, according to Fromkin (1971), are also rarely split into their component segments. Stemberger

(1983a) gives as an example

(11) They mooy – they may be moving back east again

but recognizes that these cases are exceptional. For the time being, there is no reason to assume the existence of a special spellout-and-composition mechanism for complex nuclei.

A final issue to be taken up here is whether all phonemic segments are equally "mislocation-prone." The answer is: to a substantial degree yes, but with certain qualifications. Shattuck-Hufnagel and Klatt (1979) counted the frequency of occurrence of the various phonemes in fluent speech (they restricted the count to occurrences in content words because function words are only infrequently involved in speech errors). If all phonemes are equally vulnerable, their frequency of occurrence in normal speech should predict their relative frequency of misplacement in speech errors. Correlations were computed for the consonants in the MIT speech-error corpus. The chance that a phoneme target would not be produced in the intended slot correlated 0.83 with frequency of occurrence of that phoneme, accounting for almost 70 percent of the variance. This means that, by and large, there are no "strong" intruding phonemes as opposed to "weak" error-prone ones. And indeed, each phoneme appeared about as often as target as it appeared as intrusion in the error data.

However, no such result was obtained in an experiment with elicited speech errors by Levitt and Healy (1985). They found that less frequent phonemes were indeed somewhat more error-prone, and that there was a tendency for more frequent phonemes to be more intruding or "stronger." In the study by Shattuck-Hufnagel and Klatt, too, certain phonemes were "weaker" or "stronger" than expected. In particular, /s/ and /t/ belonged to the "weaker" class. They tended to become "palatalized"—i.e., /s/ tended to turn into /ʃ/ or /tʃ/, and /t/ into /tʃ/. The reverse "depalatization" error, such as from /ʃ/ to /s/, was much less likely to occur. This asymmetry was, however, not reproduced in the Levitt-Healy study, and thus cannot explain their differential phoneme-strength effect.

The slight discrepancies between the two studies are probably largely due to the fact that the experimental study involved only single phonemes occurring in onsets of monosyllabic target items (such as *ra* and *li*), whereas the phonemes' syllable positions were not restricted in the observational study of Shattuck-Hufnagel and Klatt. In other words, a phoneme may, to some degree, be a "strong" contender for one syllable position but a "weak" contender for another position. Averaged over all positions in the syllable, however, these differences wash out, and all phonemes show an about equal average strength.

There is, moreover, a complex interaction between, on the one hand, the distributional properties of phonemes over segmental slots in the syllable and, on the other hand, differential vulnerability of syllabic slots to error. This catches the eye when one considers the following observation in the original study by Shattuck-Hufnagel and Klatt: Among the "stronger," seemingly less vulnerable phonemes was /ŋ/. The authors gave the obvious reason: This segment cannot appear in word-initial or syllable-initial position, which is the most error-prone position. Summarizing, therefore, one can say that all segments are about equally strong as contenders for segmental slots, but that there may be some dependency on position.

In conclusion: Phonemic segments are displaceable units. But when they are displaced, they are *accommodated* to their new environment. This means that they are phonetically realized in accordance with their position in the cluster or syllable. This allophonic accommodation is a natural consequence of the addressing mechanism proposed (following Crompton 1982). The same mechanism precludes the displacement of *allophones*, which is, in fact, never observed.

9.3.5 Distinctive Features

Can distinctive features, such as voicing and nasality, be moved individually? There are repeated claims in the literature that this is indeed possible. The classical examples are Fromkin's (1971). A shift of the voicing feature seems to occur in g*lear* p*lue sky* [*clear blue sky*], where the voicing feature of /b/ was anticipated in the initial segment of *clear*. Two new phonemes resulted: /g/ and /p/. A nasality shift is apparent in m*ity the* d*ue teacher* [*pity the new teacher*]. The /p/ is nasalized and thus turns into /m/; the/n/ is denasalized and turns into /d/. When two new phonemes are formed, one can be sure that a feature has been replaced.

One might conjecture that most phoneme replacements are, in actuality, replacements of single features or sets of features. For instance, the exchange *is* p*ade* m*ossible* [*is made possible*] might be due to a shift of the nasality feature only, rather than to an exchange of the phoneme segments /m/ and /p/ as a whole. In *you* g*etter stop for* b*as* [*you better stop for gas*], the apparent exchange of /g/ and /b/ may in fact be an exchange of a *pair* of distinctive features: anteriority and labiality. This, however, is very unlikely. Shattuck-Hufnagel and Klatt (1979) showed that multiple-feature shifts as in the latter example are far too frequent to be predicted by independent but coinciding single-feature shifts. The only really clear cases of feature replacements, i.e., where new phonemes are formed, occur with negligible frequency; only a few cases have been reported in the literature.

Levelt, W. J. M. *Speaking: From Intention to Articulation.*
E-book, Cambridge, Mass.: The MIT Press, 1989, https://hdl.handle.net/2027/heb08442.0001.001.

Apparently, distinctive features do not function as fillers in addressing subroutines. Their role in speech production is a different one, as will be discussed in subsections 9.4.2 and 9.5.2.

So far, we have considered the replaceability of word parts that are linguistically well defined: morphemes, syllables, syllable constituents, phonemes, allophones, and distinctive features. It turned out that syllables are hardly ever involved in substitutions, and allophones never. The displacement of features is possible but exceptional. This tells us something about the kinds of fillers that should figure in a slot/filler theory of phonetic planning: They are morphemes, syllable constituents, and phoneme segments. Allophones and distinctive features are probably computed only after the slots have been filled with the appropriate units. The spellout framework developed so far accounts for just these filler types. But is our present list of fillers complete? Shattuck-Hufnagel (1983, 1987) argues that there are still other displacement-prone word parts.

9.3.6 Word Onsets and Word Ends

Syllable onsets, we saw, are among the most frequent units involved in speech errors. But are they really always syllable onsets, or should they be characterized as word onsets? In monosyllabic words one cannot distinguish between syllable onset and word onset, but in polysyllabic words one can. In a word like *ferment*, is only the word-initial /f/ error-prone, or are both syllable onsets, /f/ and /m/, vulnerable? Shattuck-Hufnagel's (1987) statistics show that word onsets are more vulnerable than other syllable onsets in a word, by a factor of 4.5. Moreover, word onsets are far more prone to particular types of errors, especially exchanges, than other parts of the word. No less than 82 percent of the consonant-interaction errors in the MIT corpus occur in word onsets.

Still, one could argue that all this may be due to the tendency of word-onset syllables to be *stressed* syllables. It could be the case that onsets of stressed syllables are especially error-prone. So, in the verb *fermEnt* the /m/ would be error-prone, whereas in the noun *fErment* it would be the /f/. Shattuck-Hufnagel (1985, 1987) tested this in a so-called tongue-twister experiment in which a subject received a card with four words printed on it. (Examples are given in table 9.1.) The task was to read the card three times, then to turn it over and to recite it three times from memory. With examples like these, subjects made occasional errors, involving (for instance) misplacements of /p/ and /f/, as in *parade fad poot farole*. In the example for tongue-twister type i the /p/ is a word onset but it is not the onset of a stressed syllable; in the example for type ii /p/ is the onset of a stressed

Table 9.1
Effects of word position and syllable stress on segmental speech errors (after Shattuck-Hufnagel 1985).

			/p/ is	
Type	Example	Number of errors	word-initial	stressed-syllable-initial
(i)	parade fad foot parole	121	+	−
(ii)	repeat fad foot repaid	58	−	+
(iii)	peril fad foot parrot	178	+	+
(iv)	ripple fad foot rapid	8	−	−

syllable but not a word onset; in type iii it is both in a stressed syllable and a word onset; in type iv it is none of these.

The table presents the numbers of errors released in the experiment. The critical segment (/p/ in the examples) was almost never affected when it was neither word-initial nor stressed-syllable-initial (case iv). When it was stressed-syllable-initial only (case ii), there were substantially more errors. When it was word-initial only (case i), the error rate was even higher. The strongest effect resulted when the segment was both word-initial and stressed-syllable-initial (case iii). The statistics from the experiment show that the strong word-onset effect and the weaker stressed-syllable-onset effect were additive; they contributed independently to the chance of error.

Hence, word onsets do seem to have a special status as fillers, and this is in agreement with the tip-of-the-tongue results discussed in section 9.1. Still, syllable onsets are replaceable units themselves, independent of their word position. Both the experimental results and the spontaneous-error data show, however, that constituents of *stressed* syllables are especially error-prone.

It is interesting to compare this result with what Shattuck-Hufnagel (1986) found for vowel errors. There the most vulnerable position turned out to be in the syllable carrying main stress. It was far less relevant whether that syllable was word-initial or not. It would therefore be wrong to conclude that there is a special role for the word-initial syllable as a whole. Rather, it is solely the word onset (consonant or cluster) that has a special status. Shattuck-Hufnagel (1987) went a long way toward unraveling this special status. She compared consonantal errors where the intrusion came from the same planned utterance (such as exchanges) against errors where the intrusion has no obvious source in the planned utterance. These were called "interactional" and "non-interactional" errors (Dell 1986 called them "contextual" and "non-contextual" errors). An example of an interactional error is *a lung – a young lady*; a noninteractional error is *the* *inflation wate* [rate]. An analysis of errors in polysyllabic words, in which there are also non-word-initial syllable onsets, produced an interesting result: 77 percent of the interaction errors were word-initial, whereas only 28 percent of the non-interaction errors were word-initial. Is the special vulnerability of word onsets in some way related to phrasal planning (i.e., planning that involves two words from the same phonological or intonational phrase)?

To support this idea, Shattuck-Hufnagel (1987) performed another tongue-twister experiment. There were two kinds of twisters in this experiment: phrases and lists. For each phrasal twister (e.g. *From the leap of the note to the nap of the lute*), there was a list twister (e.g. *leap note nap lute*). It turned out that in the phrasal twisters 77 percent of the interactional errors were word-initial. In list twisters, however, only 44 percent of the interactional errors were word-initial. Word-onset vulnerability is apparently a consequence of generating connected phrasal speech. These findings have been replicated and extended by Wilshire (1985), but the connected speech mechanism responsible for the robust effect is as yet unknown.

The special status of word onsets is complemented by the replaceability of word ends in speech errors—in particular, word ends that are larger than a syllable. Shattuck-Hufnagel (1983) gave as an example a case where *Howard and Claire* was delivered as *H*aire *and Cl*oward. There is a shift here of the word-final *oward* (or was it a word-onset exchange where the intended order was *Claire and Howard*?). However, it is extremely rare that larger-than-syllable word endings move as units in speech errors.

This review of substitutable sublexical units has led to three main conclusions: (i) Only those sublexical-unit types for which there are address slots in the spellout mechanism are susceptible to error. (ii) Consonantal clusters are spelled out into their constituent segments, but this is not so for complex nuclei. (iii) Word-onset consonants are especially error-prone when connected phrasal speech is produced. The next section will summarize the now-expanded spellout model, and will then discuss how Shattuck-Hufnagel's slots-and-fillers theory accounts for a variety of errors.

9.4 The Slots-and-Fillers Theory and the Causation of Errors

9.4.1 Processing Levels

At the level of grammatical encoding, lemmas are released one by one with their diacritical parameters as the surface structure develops. Each of these constitutes the next bit of input for phonological encoding. According to the version of the slots-and-fillers theory presented here, this planning involves three levels of processing: morphological/metrical spellout, segmental spellout, and phonetic spellout (including cluster composition). These levels of processing are illustrated in figure 9.1, which shows the course of phonetic planning for the plural noun *crampons.*

A lemma (or rather its lexical pointer) and its diacritical features constitute the address where the item's morphological and metrical information is stored. The address is the IF-statement for a production that retrieves that information. The lexical entry *crampon* contains a singular item and a plural item. The diacritic feature "plural" directs the search to the appropriate lexical item. It is assumed that all inflectional word forms that are frequently used by the speaker are stored items in the form lexicon. Also, each frequently used derivational form constitutes an independent lexical entry (Cutler 1983a; Stemberger and MacWhinney 1986). It does not take more time or effort for a speaker to access complex inflectional forms than to access simple ones. Complex forms, inflectional or derivational, don't have to be *composed*; they are as available as simple forms.

The spellout at this level consists of a string of morphemes (stems, roots, affixes), as well as metrical information such as the number of peaks for each morpheme and the stored stress distribution relating to these peaks. For *crampons* there is a stem and an affix, and there are two peaks, of which the first one has primary stress. Each peak will trigger the generation of a syllable address frame at the phonetic spellout level. Syllable address frames probably come in two kinds: stressed and unstressed.

The segmental spellout routines take the metricized root-and-affix strings as input, and produce a string of syllables, each spelled out in terms of syllable constituents: onsets, rimes, nuclei, and codas. Onset and coda clusters are further spelled out as individual segments. The eventual result is a full segmental spellout of the item. Each segment is probably labeled as onset, peak, or coda item (but see section 9.6). If a segment pertains to a cluster, it may be further labeled as, e.g., "onset C_1" (for instance, /k/ in the example) or "coda C_1" (/n/ in the example).

The phonetic spellout routines, finally, are there to find phonetic syllable plans for strings of segments. These articulatory programs for syllables are largely stored; they only have to be addressed by way of the right key. An address consists of a triple of onset, nucleus, and coda. These are filled, in turn, by appropriate segments or segment clusters. The first syllable of *crampons* is found by inserting /kr/, /æ'/, and /m/ in the slots, respectively. Cluster composition must precede this process of insertion. The mechanism of cluster composition prevents phonotactically illegal clusters from arising.

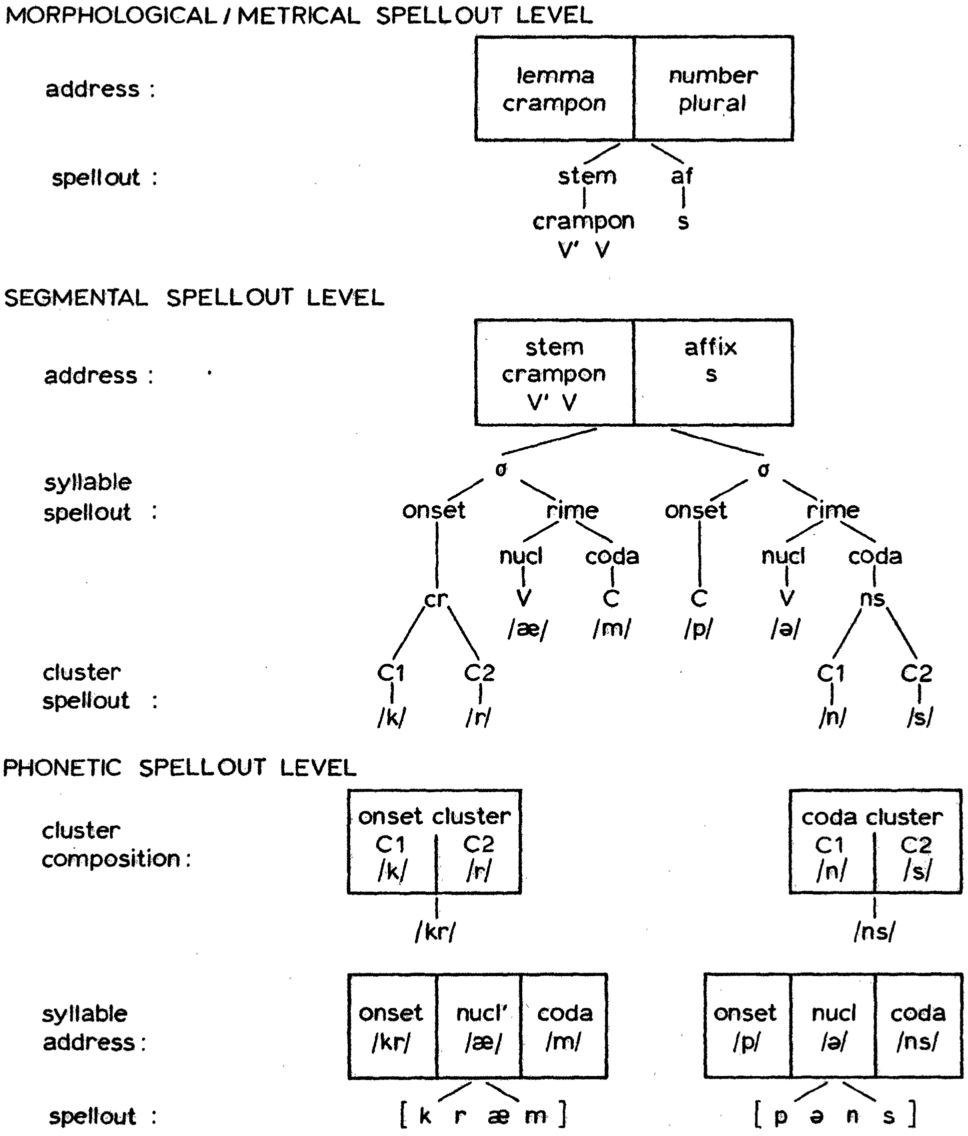


Figure 9.1
Levels of processing in generation of phonetic plan for *crampons*.

The retrieved phonetic plan for a syllable is a sequence of phones. When executed, a phone is an articulatory gesture over time. A syllable's phones are overlapping and interacting gestures. Each gesture consists of dynamic phonetic features, such as voicing and nasalizing.

There may be independently stored routines for stressed and unstressed syllables, as Crompton (1982) suggests. In *crampons*, [kræm] is a stressed syllable; [pəns] is an unstressed one. The nuclear slot of a stressed frame cannot accept an unstressed vowel such as /ə/. The appropriately filled address frames point to the correct stressed or unstressed syllable plans.

However, not everything can be stored. The syllable's stored phonetic plan is modifiable, depending on the prosodic context in which it appears. Also, there must be a mechanism for creating new, rare syllables; we will take up these issues in chapter 10. Let us now turn to the control structure proposed in the slots-and-fillers theory, and to its account of the main kinds of sublexical-form errors.

9.4.2 The Causation of Errors

On the slots-and-fillers account, errors of word form are due to failures in addressing. This is most easily seen at the phonetic-spellout level. Addressing at this level requires making triples of onset, nucleus, and coda available to complete the characteristic address frame for a syllable. Sometimes more than one appropriate filler is available for a slot, or the target filler is made available too late. The slots are then filled by a nonintended triple, and a similar but erroneous syllable is accessed.

Shattuck-Hufnagel's (1979) account of how different types of speech errors, exchanges, substitutions, additions, omissions, and shifts can arise under these circumstances involves a two-step control structure, with a *selection* step followed by a *checkoff* step. Let us consider how this is involved in the causation of the various kinds of errors.

Exchanges

Example: *a* b*ut-gusting meal* [*a gut-busting meal*]

There are two fillers available when a filler is requested for the first onset slot: /g/ and /b/. This is due to speedy segmental spellout at the previous level of processing. The most highly activated item is now selected to be copied in the onset slot. This happens to be /b/, which becomes the intrusion. After insertion, a filler is normally "checked off"; in other words, its activation is reduced to zero, and it is no longer available. This happens to /b/, but not to /g/ (which was not used). The nucleus and coda slots are correctly filled, and the syllable retrieved is [bʌt] instead of [gʌt]. At a later stage, when the next syllable-onset element is requested, /b/ is no longer available as a filler—it was checked off. But /g/ is still available and of the right kind. It fills the slot, and together with /ʌ/ and /s/ it forms the triple that accesses the syllable [gʌs]. In short: Exchanges can result when misselection is followed by normal checkoff. The original target will stay available and will fill the slot that was meant for the checked-off element.

Substitutions

Examples: *if you can change the* pirst *part* [*if you can change the first part*] (anticipation)
a phonological fool [*a phonological rule*] (perseveration)

A substitution combines a misselection with a checkoff error. In the anticipation error above, when an onset element was requested, the filler /p/, to be used later, was already available, and was more strongly activated than the target filler /f/. It was copied in the onset slot. So far, the situation is the same as in exchanges. But now, /p/ was not checked off, and stayed available as onset filler. (On the activation-spreading account to be discussed in the next section, /p/ is quickly reactivated after insertion.) It could therefore be used again as onset filler when the next syllable, [pɑrt], was addressed.

In perseverations the same processing errors are made, but in reverse order. In the second example above, /f/ is correctly inserted when the onset is requested for the first syllable of *phonological*; however, it is not checked off, and it stays available to fill a later syllable-onset slot. This leads to a misselection when, for a new word, a new syllable onset is requested; the still-available /f/ is more activated than the target /r/ at that moment. As a result, the syllable [fʊl] is addressed instead of the target [rʊl].

Omissions

Example: *Doctor -inclair has emphasized* [*Doctor Sinclair has emphasized*]

As was discussed above, this type of error suggests the existence of a null filler. The first syllable of *emphasized* has the null element as syllable onset. When that null element is early available as a filler, it may be misselected to fill the onset slot of *Sinclair*'s first syllable. It is, however, not checked off, and it stays available for repeated insertion at the appropriate occasion. On this account, omissions are nothing but substitutions involving a null filler.

However, other kinds of omissions require a different account. They are the so-called *haplologies*. Fromkin (1971) calls them *telescopic errors* because the utterance becomes contracted, as in *rigous* [*rigorous*] and *tremenly* [*tremendously*]. There are also much wilder cases, such as *I have a spart for* *him* [*I have a spot in my heart for him*]. Crompton (1982) observed that the jumps (over the omitted part) go from one syllable-constituent boundary to another identical constituent boundary. In *rig-ous* the same transition is at issue as in *rig-orous*, namely from syllable onset to syllable nucleus. In *tremen-ly* the transition is from coda to onset, as it is in *tremen-dously*. This shows that the regular pattern of address frames is followed: onset-nucleus-coda/onset-nucleus-coda, etc. Onset *clusters* or coda *clusters* are never broken open in haplologies. The situation is highly comparable to the way in which transitions are handled in blends (see subsection 9.3.2). According to Cutler (1980b), it is often the case that the two ends meet at a common segment. An example from Fromkin (1973) is *nitness* for *Nixon witness*, where the common element is /ɪ/.

It is, as yet, unclear how whole strings of fillers get lost in the phase of phonetic spellout. One cause could be that not enough syllable frames are set up, i.e., that a peak at the level of metrical spellout fails to trigger an onset/nucleus/coda frame. The available fillers then have to compete for too few slots. Let us call this a *frame-generation error*. Another cause could be that the lost elements are so similar to nearby elements that they are not recognized as different and they are put into the same slot. Stemberger and MacWhinney (1986) induced "no-marking errors"—errors where an inflection is not pronounced when the stem displays a sound form that could be a realization of that inflection. For instance, when presented with the word *lifting* or *yielding* and asked to pronounce the past-tense form, subjects occasionally answered *lift* instead of *lifted*, or *yield* instead of *yielded*. (Such errors are also observed in spontaneous speech.) This almost never happened for verbs like *bake* or *grab*, where there is no similarity between the past-tense inflection and the sound form of the stem.

Additions

Examples: *has slides sloping in* [*has sides sloping in*]
Glod bless you [*God bless you*]

The first kind of example is the most frequent one. The addition of /l/ to *sides*, however, is only apparently an addition. The most straightforward account is an anticipation of the onset cluster /sl/ of *sloping*. Thus, it is a substitution error. (This may also be so for example 10 above.)

The second kind of example (which is similar to example 9) is harder to account for. It is most probably due to misselection in cluster composition. When a syllable-onset filler was requested, a cluster /gl/ was inserted. How could this have come into existence? It probably started with a frame-generation error. An onset-cluster address was set up with two slots. C_1 was then correctly filled with /g/. C_2 accepted /l/, the C_2 element of the next onset cluster. This /l/ was not checked off after insertion, so it could be used a second time as C_2 filler, for the composition of /bl/. So, there was—on this account—a concatenation of an error in frame generation, a misselection, and a checkoff error. The origin of a frame-generation error is always difficult to locate. In the present case, an onset-cluster frame was correctly generated for *bless*; but it may have been generated too early.

Shifts

Examples: *Walter Conkrite* [*Walter Cronkite*]
Frish Gotto [*Fish Grotto*]

The apparently shifted /r/ in the first example is probably not a shift at all. The error can be accounted for as an exchange of two syllable onsets:/kr/ and /k/. The second example (from Fromkin 1971) is due to a frame-generation error followed by a misselection during cluster composition. The C_1/C_2 frame set up for *Grotto* appeared too early. It then accepted (erroneously) /f/ in its C_1 slot and (correctly) /r/ in C_2. This created the onset cluster /fr/, and both /f/ and /r/ were checked off. The cluster then filled the first syllable-onset slot. For the second syllable-onset slot, only /g/ remained as a filler.

All these types of errors can also occur at higher levels of processing. There can, for instance, be exchanges, substitutions, omissions, additions, and shifts of morphemes. (Several examples were presented above.) And the mechanisms are probably quite similar, involving failures of selection, of checking off, and/or of frame generation. Frame-generation failures are least understood, however.

The development of a complete slots-and-fillers theory requires an account of how the address frames are set up to start with. This is fairly simple at the phonetic-spellout level, where the sequence of onset-nucleus-coda frames was triggered by the sequence of peaks in metrical spellout. It is also relatively simple at the cluster-composition level. There should be a frame for each cluster spelled out at the previous level. But segmental spellout has a more highly structured sequence of address frames. They can be pairs of stem and affix, or pairs of roots, or roots with several affixes, and so on. The sequence of address frames for segmental spellout must be formed on the basis of the earlier morphological spellout results. When a lemma is spelled out as stem + affix, a stem/affix frame is set up, and similarly for other spellouts. It is not impossible that certain errors result from the setting up of deviant address frames at the level of segmental spellout, as well. That will, however, not be pursued here; instead we will consider a final factor in the causation of errors: phonemic similarity.

Phonemic similarity

Segmental errors are subject to a quite general constraint: Target and intrusion tend to be similar in distinctive feature composition. An exchange such as p*aid* m*ossible* is more likely than one such as *a two-*sen pet [*two-pen set*]. The more features on which the target and a potential intruder differ, the smaller the chance of error. This was shown initially by Nooteboom (1967) in an analysis of Dutch speech errors, and later by MacKay (1970) for German and by Fromkin (1971) for English errors. Shattuck-Hufnagel and Klatt (1979) demonstrated it for the consonantal errors of the MIT corpus, and Shattuck-Hufnagel (1986) showed the constraint to hold as well for the vowel errors of that corpus. Levitt and Healy (1985) confirmed the feature-similarity constraint in experimentally elicited errors.

Though the constraint is evidently correct, it is less clear which common distinctive features are the main determinants of segment confusion. Van den Broecke and Goldstein (1980) performed, for consonant errors, extensive multidimensional analyses on two American-English corpora, a Dutch corpus, and a German corpus. They found a clear confirmation of the feature-similarity constraint, though some phonological feature systems were better predictors of segment confusability than others. Certain features contribute more to exchangeability of segments than others, and again, the ordering differs somewhat for different feature systems. A fair summary, however, is this: Among the most affected features in consonantal errors are the place features. Target and intrusion frequently differ in the place where the main constriction is made in the vocal tract; a /p/ easily exchanges for a /k/, for instance. In other words, place is not a great contributor to similarity. Somewhat less affected is the voicing feature; that is, it is a more important determinant of (dis)similarity. At the other end of the scale is the manner feature nasality. That feature tends to be maintained in speech errors (but not in the above example *paid mossible*); to put it differently, nasal consonants are mutually quite confusable.

Shattuck-Hufnagel (1986) performed a similar analysis for vowel errors and found an interesting parallel to the just-mentioned observations. The feature that was most easily changed in vowel errors was the place feature: back versus front. On the other hand, manner features were far less vulnerable. In particular, tense vowels (such as /i/ in *beat*) tended to replace tense vowels, and lax vowels (such as /ɪ/ in *bit*) tended to replace lax vowels. The general pattern, therefore, seems to be that place features are vulnerable in speech errors, whereas manner features are more stable. There is no ready explanation for the similarity constraint within the slots-and-filler model; however, we will return to it in the next section, where an activation-spreading account is discussed.

This subsection has reviewed the causation of errors in the slots-and-fillers model. The main characteristic of the model is the thorough separation between the setting up of structural frames and the filling of these frames with appropriate, independently spelled-out units. The model accounts for the important and empirically well-supported Unit-Similarity Constraint. It also gives a principled account of which errors are possible and which are impossible. And it explains the causation of errors by failures of two control processes: selection and checkoff. In addition, there may be failure in the mechanism that governs the setting up of the address frames.

The model was not designed to make precise quantitative predictions of various error types. In particular, it has little to say about the spellout mechanisms, i.e., the ways in which sublexical units are activated and retrieved in order to become available as fillers. Also, the checkoff mechanism—the deactivation of units after insertion—may need further scrutiny. What precisely is a checkoff error? Is it a failure to deactivate, or is it, rather, a speedy reactivation process? These and similar issues are central to the activation-spreading model. The most detailed version of that model (Dell 1986, 1988) is quite compatible with the slots-and-fillers model, and it is in many respects complementary to it.

9.5 Activation-Spreading Theory

The enormous speed with which a speaker can access the form representations in long-term memory requires an efficient control structure. In this section we will consider further how, at the three main processing levels, the stored forms are accessed in order to make them available as slot fillers for the next level. The most elaborate accounts of these processes are to be found in the so-called *spreading activation* (better: *activation-spreading*) theories, which were introduced in section 1.3 and which were considered in subsection 6.3.5 in connection with the accessing of lemmas. Word-form access has been more of a hunting ground for activation-spreading theorists than lemma access, as is apparent from publications such as Kozhevnikov and Chistovich 1965, Dell and Reich 1977, 1980, and 1981, Dell 1984, 1986, and 1988, MacKay 1982 and 1987, Meyer and Gordon 1985, and Stemberger 1985a. The most elaborate treatment is that of Dell 1986; the present section reflects his notions to a large degree.

9.5.1 The Stratified Structure of the Word-Form Lexicon

Dell's (1986) activation-spreading model is organized in four strata or levels of nodes: the semantic, the syntactic, the morphological, and the phonological. These strata consist of nodes that are permanently available in long-term memory. Nodes can be activated, and when active they spread their activation to connected nodes at other levels. At the syntactic level the nodes stand for lemmas and their diacritical features. At the morphological level they stand for stems and affixes. At the phonological level they stand for syllable constituents and phonemes. In other words, the node levels correspond rather precisely to the inputs and the outputs of the processing levels discussed in the previous section. Missing, however, is a level at which phones are represented. There is, as yet, no activation-spreading account of the generation of (allo)phones.

Spellingout, i.e., going from level to level, is done by activation spreading or priming. Take, for example, the words *reset* and *resell.* The representations at the syntactic, the morphological, and the phonological strata are presented in figure 9.2. The web of connections leading down from the two lemmas at the syntactic level are precisely their long-term form representa-

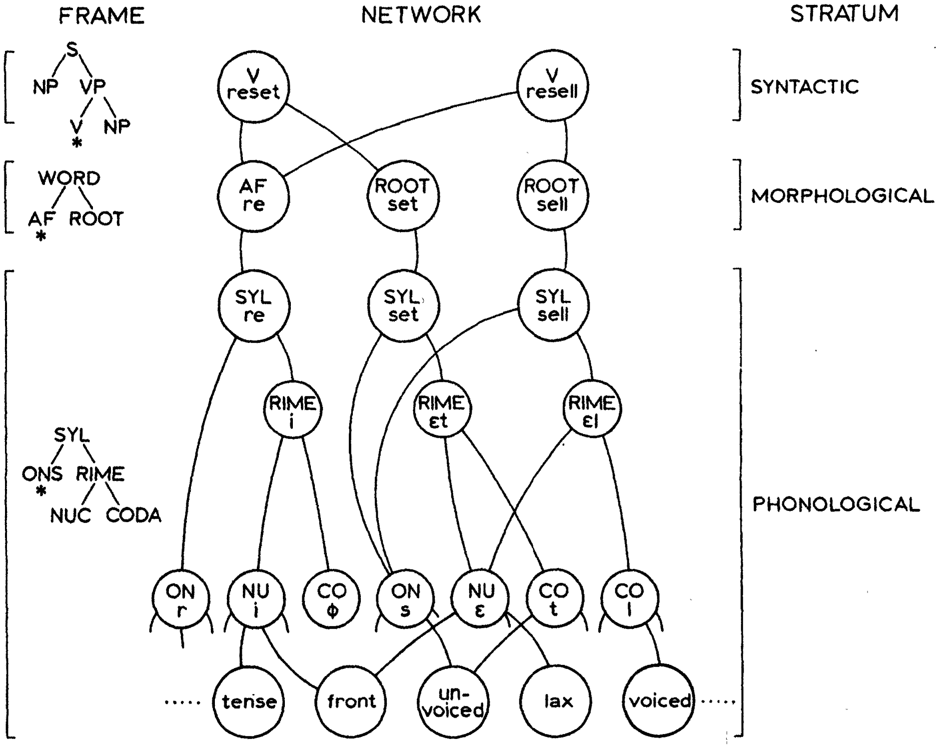


Figure 9.2
Example of strata, frames, and connected network in activation-spreading theory.

tions in the mental lexicon. The two representations share an affix at the morphological level as well as a syllable and several phonemes at the phonological level.

At each level of representation and each discrete moment in time there is one and only one *current node*. Assume that at the syntactic level the verb lemma *reset* is the current node in the developing surface structure. It is marked by an asterisk in the syntactic frame. This means that its state of activation is increased by some constant but substantial amount. At the next moment in time the increased activation is spread to the morpheme nodes *re* and *set* at the morphological stratum. This is equivalent to morphological spellout. The activation is also spread further to the connected nodes at the phonological level.

There must now be an independent structural reason for promoting a prefix node to the current node at the morphological level. In the slots-and-fillers theory this reason is the current availability of a particular slot in the developing address frame. Neither of the theories is well developed with respect to how these structural frames are set up, though Dell (1985) and MacKay (1982, 1987; see also subsection 12.1.4 below) have presented suggestions as to how an activation-spreading theory can promote a node to the current node. At any rate, a particular *kind* of node should be requested—for instance, an affix node. The requested kind of node, or the slot to be filled, is again marked by an asterisk, but now in the morphological frame. Of all the available prefix nodes in memory (i.e., in the network), the most highly activated one will be selected. Since the node for the prefix *re* was just primed, it will most probably become the current node. This increases its state of activation even more, by the fixed extra amount of activation allotted to current nodes. At the next moment, the morpheme's additional activation is spread to the phonological stratum—i.e., to the syllable and syllable-constituent nodes that are connected to the affix node. The syllable frame at this level successively requests an onset, a nucleus, and a coda.

A digression is in order here: That a syllable's onset, nucleus, and coda are requested in serial order deviates from Dell's (1986) original proposal, which says that "to simplify matters, it is assumed that onset, nucleus, and coda for a given syllable are selected simultaneously." But Dell (1988) revised this position largely on the basis of an experimental study by Meyer (1988), which makes it likely that a syllable's slots are filled in serial order. Since she found that, in accordance with Dell (1986), a word's syllables are also addressed in serial order, she could conclude that a word's phonetic plan as a whole is normally built up serially—"from left to right," so to speak.

The evidence is that one can reduce a subject's response latency in pronouncing a word by giving an appropriate prime. But Meyer found that the only effective primes were those that started at word onset. So, for the word *pitfall*, *p* is a good prime but *it* or *fall* has no effect. The priming effect, moreover, increases monotonically as longer word-initial stretches are given as primes. So, for *pitfall* the following primes are increasingly effective: *p*, *pit*, and *pitf*. Meyer developed an elegant technique to present such phonological primes.

Let us now return to the phonological encoding of *reset*. At the moment the onset of its first syllable is requested (a state of affairs indicated by an asterisk in figure 9.2), the most highly activated onset node is the phoneme /r/. It becomes the current node, and it spreads its additional current-node activation to its distinctive feature nodes. The syllable frame then requests a nucleus (and receives /i/), and a coda (the null element /∅/), which completes the first syllable. Dell (1988) allows for other syllable frames than onset-nucleus-coda, for instance a CV-frame that could accommodate the syllable /ri/ without recourse to a null element.

After a current node has been selected and its activation has been boosted and spread, the activation falls back to zero. This is equivalent to the check off mechanism in the slots-and-filler theory. It makes it unlikely that the same node is immediately available as current node again. It can be quickly reactivated, however.

A very similar story can be told about the promotion of the root morpheme *set* to current node. When after the prefix a stem node is sought at the morphological level, *set* is the most strongly activated one (*sell*, for instance, is not very active). It becomes the current node, and its activation level is boosted. The node spreads its activation to the phonological-level nodes. At that level, the current nodes required by the syllable frame are, successively, onset, nucleus, coda, onset, nucleus, coda, and so on. Each time, the most activated node of the type is selected.

An important property of the activation-spreading model is that a node's activation spreads not only to lower levels but also to higher ones. The arcs or connections in the network are perfectly bidirectional channels for the spreading of activation. Take the prefix *re*. When it becomes activated (first by priming from *reset*, then by being promoted to current node), it sends a quantum of activation back up to *reset*, but also to the lemma node *resell* and to all lemmas whose morphological spellouts begin with the prefix *re*. The slightly activated node *resell*, in turn, spreads some fraction of its newly acquired activation to the stem morpheme *sell*, and from there the activation perpetuates to the phonological level. The system would, of course, become steadily more excited in this way, and to prevent this the assumption is made that all nodes show an exponential decay of activation over time.

The control structure of spreading activation is one of parallel processing. There is simultaneous activation of whole sets of nodes at a given level of representation, and there is simultaneous activity at all levels; at each level there is always a "current" node. There are no long waiting lines in accessing the form lexicon, and that is just what is needed to account for the high-speed performance of speech generation.

9.5.2 Activation-Spreading and Speech Errors

The upward spreading of activation can now be used to explain several speech-error phenomena that are otherwise hard to understand. Here are some of them.

Malapropisms

A malapropism is the replacement of a word by another existing word that is related in form but not in meaning. This type of speech error was first analyzed (and named after Sheridan's character Mrs. Malaprop, who excelled in using wrong words) by Fay and Cutler (1977). The intruding word tends to have the same number of syllables, the same beginning, and the same stress pattern—much like a tip-of-the-tongue guess. Fay and Cutler list as examples *week* for *work*, *open* for *over*, *constructed* for *corrected*, and so on. Not all malapropisms are speech errors, of course; it happens occasionally that a speaker really doesn't know which word means what.

The activation-spreading account of real malapropisms is straightforward. We saw that *resell* is primed by the affix *re* (see figure 9.2). If there is enough time for *resell* to spread a fraction of its activation to the morpheme node *sell*, there is some minimal chance that it becomes selected as the current node instead of *set*. This chance is even higher if the activation that flowed from *reset* to *set*, and further down to the /s/ and /ɛ/ nodes, has enough time to flow back up to the root node *sell* before a new current node is selected at the morphological level. If indeed *sell* is selected as the current node, a malapropism is born: The speaker will say *resell* instead of *reset*. For these errors to occur, the rate of speaking should be low; otherwise there is not sufficient time for the backward spreading to take effect.

Lexical bias

There is a tendency in sublexical errors to create words rather than nonword strings. Speech errors such as h*old* c*ard cash* [*cold hard cash*], where the newly formed units are words, are more likely to occur than errors such as *I* s*ould be* sh*eeing him soon* [*I should be seeing him soon*], where nonsense strings result.

Dell and Reich (1981) proved this statistically for a large corpus of collected speech errors, and this found reconfirmation in the work of Stemberger (1984). Baars, Motley, and MacKay (1975) gave an experimental demonstration of lexical bias. These authors were the first to generate speech errors experimentally, and their technique has been much used since. It consists of asking subjects to read a list of word pairs. In this list there are target pairs, such as *darn bore* or *dart board*. A target pair is preceded by three bias pairs in the list. A bias pair contains at least the initial phonemes of the desired error outcome. So, in order to induce the error *barn door* for *darn bore*, bias pairs such as *ball dome* are given. Under these conditions, readers produce 10–15 percent spoonerisms on the target items, saying *barn door* [for *darn bore*] or *bart doard* [for *dart board*]. What Baars et al. found in their 1975 study is that there is much more slipping for target pairs that create real words when spoonerized than for those that create nonsense words. The error *darn bore* → *barn door* is a much more frequent type of slip than *dart board* → *bart doard*. There is a lexical bias in slips of the tongue.

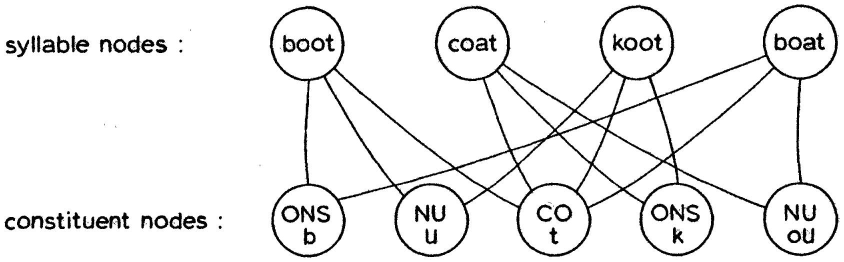
The activation-spreading account of lexical bias is, again, based on the flowing back of activation from lower to higher levels—in particular, from the phonological to the morphological stratum. This feedback can only prime the nodes of really existing morphemes in the language; there are no other nodes at the morphological level. There are nodes for *darn* and *bore* but not for *bart* and *doart*. In figure 9.2, there are nodes for *set* and *sell* but not for *sef*. This makes the error *resell* more likely than the error *resef*: a lexical bias. At the same time, the activation-spreading theory predicts that the lexical-bias effect should decrease at higher speaking rates. Backward spreading needs time to develop. This prediction was substantiated in an experiment by Dell (1985), who used a modified version of the technique of Baars et al. in which he varied the time available for a subject to respond. When the subject had to speak quickly, the lexical-bias effect disappeared.

The lexical-bias effect is also at the basis of the stress errors discussed in subsection 9.2.2. When the verb *segmEnt* is activated it will prime the noun *sEgment*, because they share most of the nodes at the phonological level. This explains how a related but differently stressed word can come to interfere in the generation of word accent.

The discussion about the origins of lexical bias in speech errors is still unsettled. An alternative to the activation-spreading accound is an editing account: the speaker intercepts nonwords just before uttering them. These issues will be further discussed in section 12.1.

The repeated-phoneme effect
It has often been observed that exchanges of word or syllable onsets are more likely to occur when the following phoneme is the same in each of the two syllables (Wickelgren 1969, 1976; Nooteboom 1973; MacKay 1970; Shattuck-Hufnagel 1985). Two examples are k*it to* f*ill* [*fit to kill*], where both exchanged consonants are followed by /ɪ/, and h*eft* l*emisphere* [*left hemisphere*], where the following vowel is /ɛ/ in both words. Dell (1984) used the technique of Baars et al. to study whether word-onset exchanges are more likely to occur when the following vowel is the same in each word than when they are different. This is indeed what he found. In addition, Dell found that the repeated phoneme need not be the vowel adjacent to the word onset; it can also be the word-final consonant. In the pair *boot coat*, for instance, the codas /t/are the same but the nuclei are not. Still, there is an increased chance that the initial /b/ and /k/ phonemes will exchange, and this must be due to their syllables' ending on the same coda.

Dell's explanation of these effects in the spreading activation model (see also Stemberger 1985a) makes use of syllable nodes at the phonological level, which mediate between morpheme nodes and nodes for syllable constituents (see figure 9.2). For *boot* and *coat*, for instance, the situation looks like this:



Here *boot* primes /b/, /u/, and /t/. The /t/ is, however, doubly primed; it also receives activation from *coat*, which moreover primes /k/ and /oʊ/. Together, /k/, /ʊ/, and the doubly primed /t/ feed back to the syllable node *koot*. Similarly, the syllable *boat* receives double activation from /t/, which increases the likelihood that it will become the current node in a slip. The repeated-phoneme effect should be dependent on the rate of speaking. At high rates, there is not enough time for lower-level nodes to feed back to higher-level ones. Dell (1986) found evidence of a diminished effect in exchanges when the speaking rate was increased.

Checkoff failure

Remember that Shattuck-Hufnagel (1979) explained the occurrence of substitution errors, in particular anticipations and perseverations, as resulting from a combination of a misselection and checkoff failure. A misplaced filler, once used, remains available as a filler. Its activation is, apparently, not turned back to zero. The activation-spreading model explains this, again, by means of feedback from lower to higher levels. When in the above diagram the syllable node *boot* is the current node, it spreads its activation to /b/, /u/, and /t/. The *boot*'s activation resets to zero. In their turn, however, the activated phonemes /b/, /ʊ/, and /t/ return part of their gained activation to *boot*. As a result, *boot* is reactivated shortly after it fired. This makes it a candidate for a subsequent choice as current node. The speaker may say *boot boot* instead of *boot coat*. Also, the syllable node *boat* will be activated by /b/ and /t/. This makes *boot boat*, where there is perseveration of /b/, a possible slip of the tongue. In view of this explanation, anticipations and perseverations should occur especially at low speaking rates, because then there is time to (re)activate a higher-level node from a lower-level node. At higher rates, exchanges (which require normal checkoff of used fillers) are more likely. This is exactly what Dell (1986) found when he applied the technique of Baars et al. at varying speaking rates.

Phonemic similarity

The intruding phoneme tends to be similar to the target in terms of its distinctive feature composition. This was called the phonemic similarity constraint in subsection 9.4.2. The activation-spreading theory can account for this phenomenon if a level of distinctive feature nodes is introduced. This level is depicted as the bottom level of nodes in figure 9.2. The explanation goes like this: Each activated phoneme node will spread its activation to the corresponding set of feature nodes, i.e., to nodes representing features such as voicing, place of articulation, and nasality. In their turn, the activated feature nodes will return activation to all phoneme nodes that share these features. Hence /s/ will, through its distinctive feature node "unvoiced", activate /t/; /i/ will, via the feature node "front", activate /ɛ/; and so on. In this way, similar phonemes increase their mutual availability as candidate fillers. This enhances the chance that an intrusion will be similar to a target.

However, the functional significance of a feature stratum is not altogether clear. The sets of feature nodes connected to a phoneme node cannot represent the allophonic feature composition of the segment in its specific environment. If that were the case, the segmental dislocations that the model describes would be replacements of (allo)phones by (allo)phones. But we know that phones do not exchange in speech errors; only phonemes do, and they become accommodated to their new environment. Hence, the feature sets represent phonemes; they are not phonetic output parameters. But this means that the activation-spreading theory in its present state does not generate a phonetic plan; at least the lowest level of planning, the phonetic specification of syllables, is still missing.

The present review of the activation-spreading theory (especially Dell's) doesn't do justice to one of its main strengths: its quantitative formal nature. Dell (1986) reported two computer simulations of the theory. There are parameters for the rate of activation spreading and for decay of activation. At each (discrete) moment in time, each node summates the incoming activation from all nodes it is connected to. There is a constant extra quantum of activation (called *signaling activation*) for the node that becomes the "current node." This is a fairly limited and fixed set of parameters. Independent of these is a speaking-rate parameter: the time (or the number of discrete time units) allotted for the generation of a syllable. Dell made the fruitful assumption that the rate of activation spreading is constant and independent of speaking rate. This independence made it possible to predict the nontrivial result that lexical bias would increase with slower speaking rate, and that high speaking rates would favor exchanges over perseverations and anticipations. As we saw above, these model predictions were experimentally verified.

Dell has been very careful not to overstate the power of his model. For instance, some of the properties of errors reviewed above do not fall out naturally from the model's assumptions. In particular, the model does not account for the special status of word onsets in speech errors, nor does it predict the extra vulnerability of stressed syllables and of syllable onsets. It is, of course, relatively easy to state these facts in terms of the model. One can manipulate the parameters in such a way that more activation is spread from a lemma to its word onset, to its accented syllable, and to syllable-initial consonants than to its other phonological constituents; but then the model is used only as a formal language to *describe* the data, not as their *explanation*. On these issues, see especially Dell 1988.

9.6 Serial Order in Phonological Encoding

According to the slots-and-fillers model, a word's segmental spellout by and large preserves its linear order; i.e., there is an ordered string of segments (Shattuck-Hufnagel 1979). When, at the next level, segments (or clusters) are required to fill successive syllable slots (i.e., onset, nucleus, and coda slots), the string of segments can be used in the given order. If the word is *pitfall*, the segmental spellout is the string /p/, /ɪ/, /t/, /f/, /ɔ/, /l/. These segments will then be used in the same order to fill the slots of two successive syllable frames, and the correct syllable plans [pit] and [fɔl] will be addressed. For this procedure to work, the segments need not be categorized in terms of onset, nucleus, or coda. The segment /t/ will never end up in the onset slot, because of its serial position. It is not necessary to label it as "coda segment".

If this procedure were followed strictly, speech errors wouldn't arise. The error *fitpall*, for instance, couldn't occur. Clearly, the ordering is not always fully specified (as Shattuck-Hufnagel of course recognized), since such errors do occur. This, however, requires a measure to prevent errors (such as *piftall*) where an onset and a coda segment are exchanged. Such errors, which violate the Unit-Similarity Constraint, are exceedingly rare. Are spelled-out segments categorized in terms of their function in the syllable (onset, nucleus, or coda)?

So far we have followed Dell's suggestion that segments are indeed labeled in terms of their syllabic function. And this is an obvious requirement in the activation-spreading theory. There is no guarantee in that theory that the order of maximally activated phonemic nodes corresponds precisely to the word's order of phonemes. When, for instance, the current node must be an onset segment, it could well be the case that the strongest activated phoneme is one that should end up as coda. To prevent this, Dell labeled phonemes according to syllabic function (see figure 9.2). When an onset is required, the most highly activated onset node is selected as the current node. This guarantees that the Unit-Similarity Constraint applies at this level. It is impossible to misselect an onset for a coda, or inversely; an onset intrusion must itself be an onset segment, and similarly for nucleus and coda intrusions.

Meyer (1988), however, suggests that this may be overkill. First, her experiments show that there is rather strict linear ordering in the making available of a word's segments. Dell's 1986 model may have to be modified to account for this finding. Second, syllabic slots may accept only certain *kinds* of segments. This is clearest for the nucleus slot. It will accept only phonemes of high sonority value (in particular, vowels). A spelled-out vowel need not be labeled as being of category "nucleus" to be recognized as a possible nucleus filler. Also, certain phonemes can never become onsets (e.g., /ŋ/ in English), and others will never be codas, depending on the language. Finally, cluster composition can play a distinct role. Subsection 8.1.3 mentioned Selkirk's Sonority Sequencing Generalization: A syllable's sonority slopes down from the peak in both directions. This means that an onset cluster, such as /sm/ or /skr/, has its phonemes in increasing order of sonority. Similarly, a coda cluster, such as /ld/ or /nt/, consists of phonemes that are decreasing in sonority value. To the extent that this property holds, it distinguishes onset clusters from coda clusters. If the onset slot of a syllable address accepts only sonority-increasing clusters whereas a coda slot takes only clusters that are sonority-decreasing, that will suffice to prevent violations of the Unit-Similarity Constraint as far as onset and coda clusters are concerned. It is not necessary, then, to categorize phonemes as "onset cluster element" or "coda cluster element".

This chapter began with an expression of concern that the phonological encoding of words may appear to be a wasteful process. At the level of segmental spellout, a word's syllabic composition becomes available. At the next level of processing, the spelled-out segments are used to address syllable plans of the same composition. Why can't there be a short-cut? The main answer to this is that in connected speech a word's spelled-out syllabic composition is often not preserved in the resulting string of syllable plans. Strings like *gave it him* are resyllabified in fluent speech (*ga-vi-tim*). They become new, so called phonological words. And this enhances the ease and the fluency of articulation. But we can now add a second point: If it is not strictly necessary to label each spelled-out segment with respect to its function in the syllable (i.e., as onset, nucleus, or coda element), phonological encoding may not require a full spellout of a word's syllabic composition at the segmental-spellout level to start with. On that view, syllables will appear only at the final stage of phonological encoding: the phonetic-spellout level. Further research is needed to settle this point.

Summary

Phonological encoding is a process by which the phonological specifications of lexical items are retrieved and mapped onto a fluently pronounceable string of syllables. Unpacking a word's phonological specifications and using them to retrieve the appropriate syllable programs involves various levels of processing. Studies of the tip-of-the-tongue phenomenon, in which this process of phonological unpacking is blocked or slowed, support this view. An initial sketch of the word can be available while further segmental details are still lacking.

There are two rather complementary accounts of phonological encoding: the slots-and-fillers theory, with Shattuck-Hufnagel as its mother, and the activation-spreading theory, with Dell as one of its fathers. Our review started from the slots-and-fillers perspective. It was proposed that there are three levels of processing in phonological encoding. At the first level, lemmas and their diacritical features are the fillers for addressing and spelling out the stored morphological and metrical composition of words. At the next level, this information is used to address and spell out the word's segmental composition. At the third level, a word's string of segments is used to address one or more phonetic syllable programs. These syllable programs are specifications of articulatory gestures, built up out of consecutive but mutually overlapping phone gestures.

At each level, independently defined address frames are set up that "request" slot fillers of particular types. For morphological/metrical spell-out, the address frames consist of slots for lemmas (or, rather, their lexical pointers) and their diacritical features. A frame is set up for each successive lexical category in surface structure. At the level of segmental spellout, frames are set up that contain slots for morphemes, such as stems, prefixes, and affixes. At the level of phonetic spellout, the frames for addressing syllable programs consist of onset-nucleus-coda triples. One such frame is set up for each peak in metrical spell-out. A preliminary process at this level is cluster composition, in which segments are combined to make phonotactically acceptable onset or coda clusters.

Each level, therefore, requests its own types of fillers: lemmas and diacritical features for morphological/metrical spellout, morphemes for segmental spellout, phonemic segments and syllable constituents for phonetic spellout. Neither syllables nor distinctive features or allophones are used as fillers for the addressing of stored form representations.

The slots-and-fillers theory accounts for the occurrence of speech errors by assuming occasional failures of two control processes: filler selection and filler checkoff. There may, in addition, be errors in the generation of address frames. The theory gives a natural account of the Unit-Similarity Constraint, the observation that target and intrusion in errors are almost always of the same structural level and category. And, closely related to this, it correctly distinguishes between possible and impossible speech errors.

How are fillers made available? By spellout in the slots-and-filler approach. More elaborate accounts of filler activation at the different levels of processing are proposed in the connectionist or activation-spreading theories. Dell's account, in particular, is formal and quantitative enough to predict the kinds and (to some extent) the relative frequencies of word-form errors, both in observational data and in experimentally induced slips of the tongue. It also provides explanations for various other speech-error phenomena, such as lexical bias effects and effects of speaking rate. In this theory, the different kinds of fillers are represented as nodes at different strata or levels of representation—among them a level for word or lemma nodes, a stratum for morpheme nodes, and a phonological stratum where there are nodes for syllables, syllable constituents, and phonemes. Between levels, nodes are connected by arcs along which the activation of a node is spread to nodes at a lower level, but also to nodes at a higher level. This layered network is a theory of the structure of the word-form lexicon and of the way in which it is accessed. Essential for its operation is the parallel activation of structural frames that control the order in which activated nodes are boosted, a central concept from the slots-and-fillers theory. As it stands, the activation-spreading theory does not yet account for the generation of (allo)phones as they appear in the final phonetic plans for syllables.

The final section of this chapter reconsidered the mechanism of serial ordering in phonological encoding. The more ordering there is of segments at the segmental-spellout level, the less need there is for segments to be labeled in terms of their syllabic functions. Such labeling can also be made superfluous by taking the sonority of segments and clusters into account. Maybe a word's syllabic composition appears only at the final, phonetic spellout level.

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.