Chapter 2

The Phonetic Framework

Of course, to understand the phonetic foundation of the Welsh mutation system within the fortis-lenis scale just presented, we must first gain some appreciation of modern phonetics. Indeed, the key to the system can only be found in the most modern approaches of "dynamic phonetics."

To the student who is not familiar with the nature of phonetics – and how that nature has changed in the past half century – just how the fortis-lenis scale operates within the mutation system may seem quite strange. The reason for this disorientation lies in the alphabetic biases with which we have come to view the sounds of a language.

The alphabet is perhaps the most economical and useful system of writing that has ever evolved. However, we must never forget that the alphabet is designed only for writing – for informing a reader what it was that the writer intended to say. It was never devised for the analysis of a sound system. Nevertheless, it has slipped into this role – a role for which it is quite simply not suited at all.

2.1 The Segmental Fallacy

Early in the twentieth-century linguistic tradition, Ferdinand de Saussure admonished linguists not to confuse sounds with letters. Writing, he concluded, "is not a guise for language but a disguise."[1]

Early structuralists took this warning to heart. While the word *right* contains a *gh* sequence in the orthography, the word was analyzed simply as [rait], without the slightest regard for the silent spelling sequence.

Far more serious, however, is the very notion of segmentation – the isolation of letters in sequence. With the invention of experimental acoustic phonetic equipment in the 1940's,[2] it was discovered that consonants and vowels are not produced in sequence with one another, but rather in a process termed "dynamic coarticulatory constraint" – certain "features" and feature types constrain the production of other features and feature types with which they are simultaneously produced.[3] Features (such as the aspiration and voicing we have already encountered) are characteristics associated with sounds, and a list of features used in this work is found in table 2.1.

Table 2.1: Main Welsh Features (with Letters)
I. Vowel Features A. Depth 1. Front (*i, e*) 2. Central ("obscure" *y, a*) 3. Back (*w, o*) B. Height 1. High (*i,*"clear" North Welsh *y/u, w*) 2. Mid (*e, y, o*) 3. Low (*a*) C. Rounded/Protruded (*u, o*) II. Consonant Features A. Position 1. Labial (*p, b, m, f, ff*) 2. Dental/Alveolar (*t, d, n, s*) 3. Alveopalatal (*si, tsi*) 4. Velar (*c, g, ng*) B. Manner – Degree 1. Stop (*p, t, c*) 2. Fricative (*ff, si, th, ch, f, dd*) 3. Affricate (*tsi*) C. Nasality (*m, n, ng, mh, nh, ngh*) 1. Voice 2. Voiced (*b, d, g, f, dd*) 3. Voiceless (*p, t, c, ff, th, ch, tsi*) D. Aspiration 1. Aspirated (initial *p, t, c, mh, nh, ngh)* 2. Unaspirated (*b, d, g*)

For example, if we say the words *feel* [fi:l] and *fool* [fu:l] very slowly and carefully, we will notice that the tongue is forward and the lips drawn back for the first word at the very same time that the lower teeth are engaging the upper lip in the "sound" [f]. Likewise, the tongue is back and the lips protruded (or rounded) for the second word at the very same time that the lower teeth connect with the upper lip for the "sound" [f].

Are the tongue position and the lip protrusion thus features of the [f]? Of course, they are not. Rather, they are features of the vowels – the [i:] and the [u:], respectively. They would be present if any other consonant were uttered in place of the [f] or if no consonant were there at all. They are perceived with the [f] simply because they are "dynamically coarticulated" with the [f] (they occur at the same time as the [f]).

Let us now compare spectrographic analyses of the words in figure 2.1.

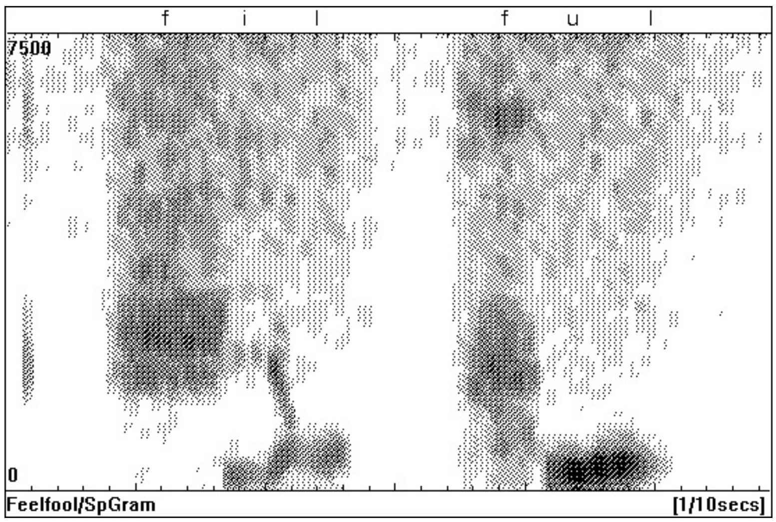

Figure 2.1: Spectrogram of *feel/fool* [fi:l/fu:l]

The spectrogram is an electronic record of speech sound. Time is represented from left to right (*feel* uttered before *fool*), amplitude (loudness) is represented by the relative darkness of the lines, and the "frequencies" (the crests of the sound waves in cycles per second) are mapped out vertically. Thus, as we proceed through time (through the utterance of the word), the graphic record shows us the distribution of frequencies during the various stages in the continuum.

As we shall see in greater detail in chapter 4, the lower frequencies give the patterns of the vowels, particularly the lower two horizontal lines – the first and second "formants" (the crests of the sound waves forming horizontal lines on the spectrogram). These are constrained by the higher frequency "noise" (an undistinguished blur) of the [f].

The question is: How shall we read these spectrograms? One way is simply to segment them cross-sectionally into three parts each, corresponding to the [f] the [i:] or [u:] and the [l]. This is done for *feel* [fi:l] in figure 2.2.

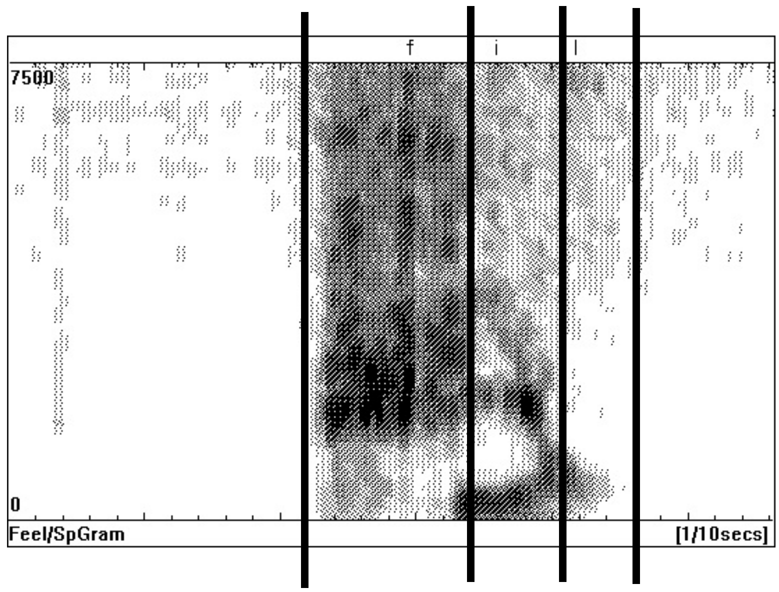

Figure 2.2: Cross-sectional Segmentation of *feel* [fi:l]

When we perform a cross-sectional segmentation, we commit a fallacy. On the lower end of the [f] are vowel formant frequencies that identify the [i:] as opposed to the [u:]. Typically, the structural phonologist claims that these are "allophonic variants" of the phoneme /f/, a notion used to bolster the alphabetic approach to sound analysis. In any event, we must somehow separate the vocalic features coarticulated with the consonantal features in order to define the consonantality of the [f].

This cross-sectional segmentation can be represented schematically as in figure 2.3, in which the vocalic features are necessarily included with the consonantal, pending further rules to get rid of them.

f	i:	l

Figure 2.3: Schematic Representation
of Cross-sectional Segmentation

While the English examples may make this approach seem rather benign, there are quite prominent cases in which this process has led to major problems. For instance, the "dorsal" fricative (produced by raising the back of the tongue) in German is realized as [ç] with front vowels – as in *ich* [iç] 'I' – and as [χ] with back vowels – as in *ach* [aχ] 'oh'. The [ç] thus maintains the front feature, while the [χ] maintains the back feature. As demonstrated in a thorough dynamic analysis, however, the front/back feature has nothing to do with the consonant, but is simply swept into it from the vowel. This inclusion of vocalic features with consonantal is necessary in the traditional approach because of the practice of cross-sectional segmentation.[4]

One reason for accepting this segmentation is that it appears to mimic the phonic approach to reading. However, readers do not produce one letter, stop, and then produce the next. Rather, in a word such as *feel*, they see the *ee* coming after the *f* and adjust their pronunciation accordingly, and they see the *l* coming after the *ee* and do the same.

If we did not anticipate in this way, the result would be unnatural and perhaps even unintelligible. A simple experiment with uttering sounds into a tape recorder and then cutting and splicing them together will provide us with a rather graphic proof: If we splice off the [f] of *feel* [fi:l] and attach it to the [u:l] of *fool* [fu:l], the result will be heard quite distinctly as *fuel* [fju:l].

Returning to phonics, in anticipating the consonant, we put those speech organs that produce the vowel into position before we even begin the consonant. Thus, we do not produce the *f* of *feel* and then produce the *ee*, but rather we produce the *f* and the *ee* together in such a way that the contact of the *f* ends while the *ee* is still being produced. Likewise, we do not complete the *ee* and then produce the *l*, but rather we maintain the *ee* as we move our tongue into position for the *l* and finally effect contact, ending the *ee* and the *l* at the same time.[5]

From a phonetic point of view, rather than using a hatchet to cross-sectionally segment the spectrogram, what if we were to use a scalpel to excise the [f] – or more precisely, the features of the [f]? The result would be something like that in figure 2.4 (albeit also an oversimplified representation).

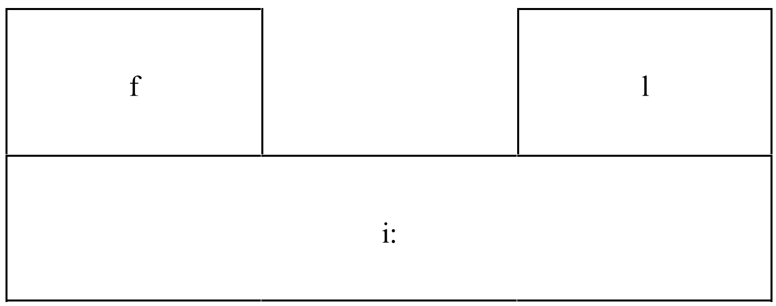

Figure 2.4: Excision of Consonantal Features from Syllable

To be sure, when we divide the spectrogram in this manner, we still include some vocalic features in with the consonants and some consonantal features in with the vowels – especially in the production of the “grave” labial fricative [f] (with energy concentrations in the lower frequencies) coarticulated with the grave vowel [u:]. Nonetheless, the division is far more accurate, and it provides us with a reasoned and patterned basis for the further proper classification of all features; for when we place ourselves in the position simply of shifting a feature “up” or “down” into the proper “box,” we maintain the dynamic time relationship as a constant. For our purposes in analyzing the fortis-lenis scale, moreover, we shall indeed find that the pertinent consonantal feature parameter is marked by higher frequency noise (see section 4.1 below).

When we separate the consonantal features from the vocalic features in this “dynamic” representation then, the consonantal features are assigned dynamically only to the consonant and the vocalic features are assigned only to the vowel. Schematically, we could represent the distribution of features as in figure 2.5.

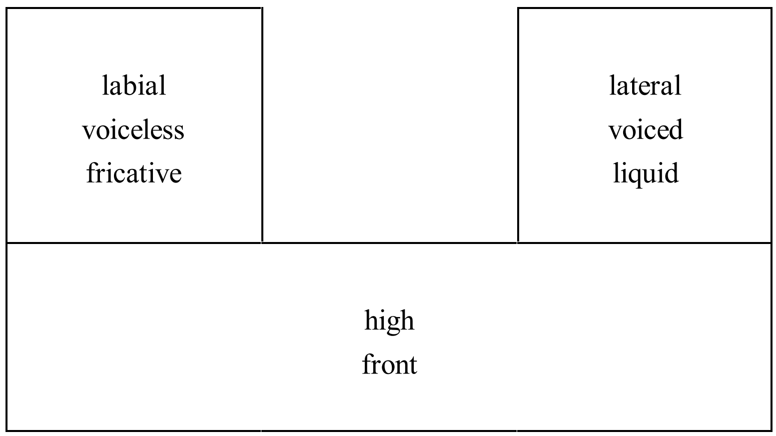

Figure 2.5: Incorporation of Features

This procedure yields the basic organization of features in the syllable. The syllable provides not only a convenient frame for analysis, but also a scientifically reliable one. The speech flow can indeed be cross-sectionally segmented into syllables not only impressionistically (where we can allow a pause), but even automatically. It is thus the shortest verifiably sequential unit of speech – and as a science must rely upon things that can be verified by the data as observed (not as theoretically idealized), the syllable must indeed take precedence over the segment.[6]

Figure 2.5 is thus the most basic representation of the "syllabic frame" used for phonetic and phonological analysis free from the constraints imposed by the segment. While the segment cannot be verified in the phonetic data, the only two elements of the syllabic frame – the syllable and features – are indeed verifiable.

By proceeding within such a framework, we can be assured that our analyses will be capable of describing the phenomena – subject only to our limitations as analysts and not to limitations imposed by a scheme of presupposed alphabetic representation.

2.2 The Dynamic Syllabic Frame

Returning to the preliminary syllabic frame, we should note that in figure 2.5, there is no structural reason to place labial above voiceless and fricative in the box for the initial consonantal constraint. We could just as easily arrange them the other way around. Indeed, if we wanted to, we could certainly put the consonantal boxes below the vocalic.

The reason we place the consonantal over the vocalic derives from the layout of the sound spectrogram. Consonants – especially in the aspect that characterizes the fortis-lenis scale[7] – are typically characterized by higher frequency emissions (energy higher up on the spectrogram), and vowels by lower; and we conventionally conceptualize higher frequency with higher elevation. There is no structural reason in the data, however, why we could not reverse the order.

As long as the occurrence of the features demonstrates "dynamic coarticulatory constraint" (from section 2.1), we can arrange them any way we choose.[8] This allows us the liberty of introducing the element of function into our framework. To refine our framework, we should examine the speech event in greater detail and see how we might organize the features to reflect not only the physical event itself, but also the functioning of the features within the system of speech production.[9]

If speech is not produced by the stringing of segmental sounds together, then how is it produced? To keep things simple, let us restrict ourselves to the "normal" articulation of the languages under study. Throughout the event, there will be two components – the physiological production and the acoustic perception.

2.2.a The Laryngeal Division. Physiologically, if there is no airflow to support speech, there will be no speech produced; and acoustically, if this airflow does not maintain some fundamental frequency, there will be no speech perceived. The first and most basic element of the speech event is therefore the forcing of air up from the lungs and through the larynx, where it passes through the glottis (the area between the vocal cords). This sets up a vibration of the vocal cords known as phonation, realized acoustically as the fundamental frequency.

We can hear this phonation if we simply utter any vowel sound, such as *ah* [a:]. Remaining with the same vowel, we find that we can alter the configuration of the larynx in such a way as to raise and lower the pitch. This pitch functions in the language in as short an utterance as a syllable, or as long as a sentence. If someone asks us a question and we answer [a:] with falling pitch, our interlocutor interprets our utterance as assent; while if we answer with rising pitch, our interlocutor interprets our utterance as a question, leading to the repetition of the original question.

The manipulation of pitch on the phonation/fundamental frequency is thus a constraint on the basic feature of the laryngeal division of speech. We refer to this use of pitch as intonation. At this point, we can represent the laryngeal division of the speech event as in figure 2.6.

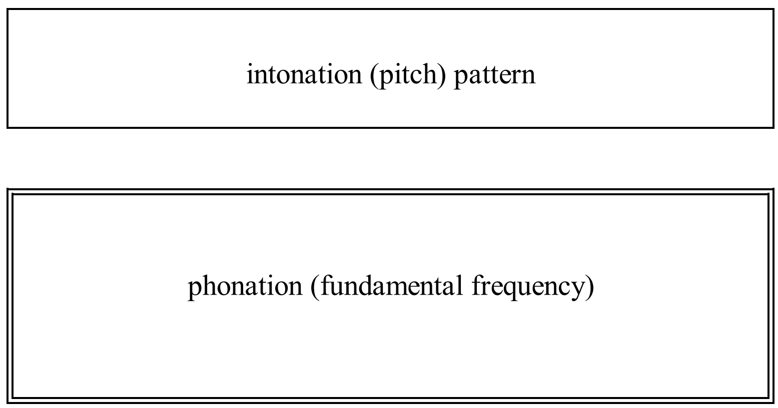

Figure 2.6: The Laryngeal Division of the Speech Event

In figure 2.6, we should note that we have created a functional convention. The lower, "basic" feature of the division is simple phonation. Constraining this phonation – and accordingly placed over it – is the intonation pattern. Our convention is thus to place the basic feature in the larger, lower box and the constraining feature in the smaller, upper box. Again, there is no reason inherent to the data themselves to place one feature over or under the other – we do so as a means of showing the function. By tradition (particularly within the London School of prosodic analysis), we may call the intonation/pitch feature of the constraining feature box the laryngeal "prosody."[10]

The laryngeal division extends beyond the syllable, for intonation is pertinent to phrases, clauses, sentences, and whole discourses. It is realized concurrently with the syllable for two reasons: First, the only way we have of uttering the phonation is to place our tongues in the position to utter a vowel, the basic element of a syllable. The only way around this restriction is to decapitate the speaker above the larynx – a rather self-defeating practice indeed. Second, the phonation pattern is the foundation upon which the syllable (with its syllabic vowel) is based.

2.2.b The Syllable Division. The syllable division of speech may thus be seen as a constraint on the laryngeal division. To be consistent with our framework, then, we should place the syllable division above the laryngeal to illustrate the function of constraint.

Within the syllable division, the basic features are those of the vowel. The vowel is defined physiologically by the tongue-body position (high/low, front/back; and in some languages retroflex, lateral, etc.), the degree of lip protrusion (round/unround or protruded/nonprotruded), and the jaw height (up/down).[11] Thus, the basic syllabic features should define the position of articulation. This is also parallel to the laryngeal division, in which the phonation is uttered at the larynx and the constraint is "superimposed" upon the phonation.

The features superimposed upon the syllabic vowel include such prosodies as syllabically pertinent pitch, nasality, and voicelessness.[12] As these features function to constrain the basic vocalic features, we can place them in the box over the basic vocalic box of the syllable division, as in figure 2.7

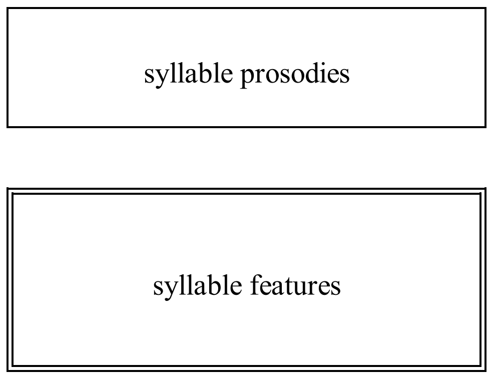

Figure 2.7: The Syllable of the Speech Event

Now we can definitely see a pattern emerging. Once again, however, this pattern is a functional one that is allowed by the data because the placement of the features in the vertical dimension makes no difference to these data.[13] As always, the basic box contains the position features, and the constraining box the prosodic features.

Physiologically, the position features are clear enough – front/back, high/low. Once again, in some languages we may need to include other configurational features (retroflex, lateral, etc.), but such features will always demonstrate how the mouth is divided into acoustic chambers – areas that produce different sounds depending upon their relative size (as the sound produced by blowing over the opening of a bottle varies with the amount of liquid in the bottle).

These acoustic chambers provide the fundamental frequency with a pattern of "harmonics" – frequencies based upon the fundamental features that occur higher. These "formant frequencies" may be compared with the notes played in a musical chord on a piano; and their appearance on the spectrogram does in fact resemble the manner in which the notes are arranged on the musical staff, with the pattern defining the chord, up from bass to treble. A high formant may be considered "acute" and a low one "grave," and so forth.

2.2.c The Obstruction Division. As we cannot utter a vowel unless it is constraining the phonation, neither can we utter a consonant without its constraining a vowel. If we attempt to utter a [p], for example, “in isolation,” we will notice that our tongue body is indeed in some configuration, our lips at some degree of protrusion, and our jaw at some height. The only way to utter this [p] truly in isolation from a vowel is thus to remove the tongue, lips, and jaw. This eliminates any hope of articulating the [p] as surely as the removal of the larynx eliminates all hope of uttering a vowel (again, in “normal” articulation).

From an acoustic standpoint, the consonant is realized on the spectrogram by the perturbation of the formant frequencies of the vowel to or from a point on the spectrogram that is typical for that consonant constraining that vowel. We call this point the “locus,” and it cannot be defined on its own, apart from the formant frequencies of the vowel it constrains. To be sure, from our rough division of the spectrogram into a syllabic frame, the locus is often in that part of the spectrogram associated with the lower vowel formants, and this fact does show that our divisions ought not to be intended to coincide completely with all details of the spectrogram. Nonetheless, as we shall see in section 4.4, the consonantality is indeed determined at the higher frequencies, insofar as the fortis-lenis scale is concerned.

Thus, from both a physiological and an acoustic standpoint, we treat the consonant as a constraint or “obstruction” on the vowel.

Indeed, to be precise, it is best to refer to it as an obstruction as in figure 2.8 and to place the obstruction division of the speech event over the syllable division.

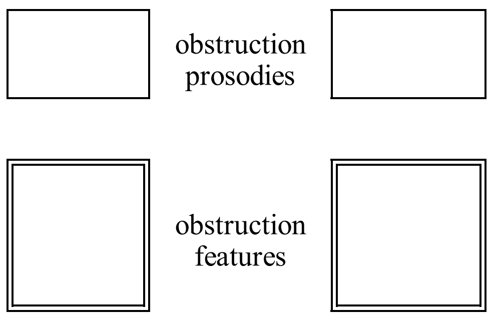

Figure 2.8: The Obstruction Division of the Speech Event

Once again, the basic obstruction features are those that define the position of obstruction – the position of the obstructing speech organs during the obstruction. These are further constrained by obstruction prosodies such as nasality, aspiration, etc., depending upon the functioning of the phonological system.

2.2.d The Pattern of Dynamic Coarticulatory Constraint. Thus we see that the speech event is a pattern of constraint. In the laryngeal division of speech, the phonation is constrained by its intonation prosodies; the laryngeal division is constrained by the syllable division of speech, in which the vocalic/syllabic features are constrained by the syllable prosodies; and the syllable division is functionally in turn (horizontally) and dynamically simultaneously (vertically) constrained by the obstruction division, in which the obstruction features are constrained by the obstruction prosodies. We can summarize the entire patterning of dynamic coarticulatory constraint as in figure 2.9.

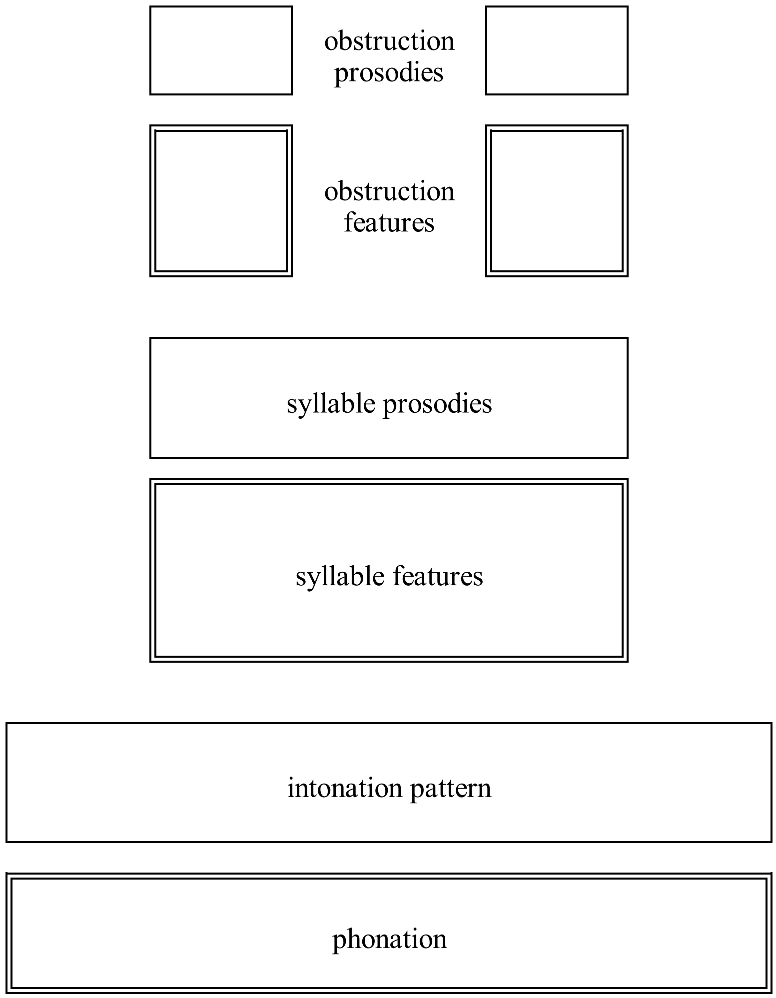

Figure 2.9: Dynamic Coarticulatory Constraint

2.2.e The Syllabic Frame. For the analysis of syllables and words, the laryngeal division will not be necessary, as this division is needed only in the analysis of sequences of syllables within an actual utterance (where intonation is pertinent). We can thus view figure 2.10 as the syllabic frame to be used in the analysis of the Welsh mutation system.

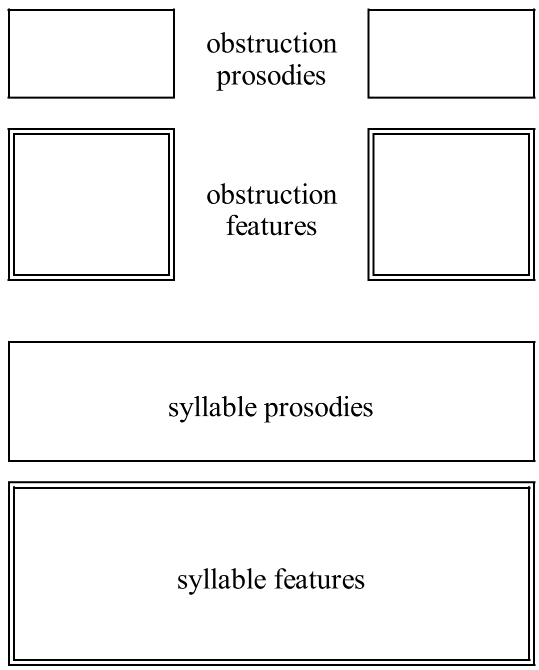

Figure 2.10: The Syllabic Frame

Within the frame, we should note that the features may be realized in more than one division, depending upon the functioning of the system. More than a product of function alone, however, such cooccurrences reflect the reality of the data, for a feature realized in a constraining division is produced and perceived to a heightened degree. This allows us, for example, to perceive a nasal consonant (obstruction division) constraining a nasalized vowel (syllable division) in the naturally nasalized speech of a particular dialect (laryngeal division).

We should also note very carefully that the features realized in the boxes are not at all limited to the boxes themselves. Such features as aspiration as an obstruction prosody often extend well beyond the point that the consonantal articulators leave their position of obstruction.[14] In earlier versions of the dynamic model, this was illustrated by maintaining jagged lines on either side of the boxes.[15] For phonology (as opposed to phonetics), the jagged lines were probably superfluous. Suffice it to say that the boxes represent positions within the syllabic frame, and the features are associated with (not inherent to!) these positions.

Notes to Chapter 2

1. Ferdinand de Saussure, *Course in General Linguistics*, trans. by Wade Baskin (New York: McGraw-Hill, 1959), p. 30.

2. Particularly the sound spectrograph – W. Koenig, H.K. Dunn, and L.Y. Lacy, "The Sound Spectrograph," *Journal of the Acoustical Society of America* 17 (1946), 19-49 [Rpt in: *Readings in Acoustic Phonetics*, ed. by Ilse Lehiste, 3-33 (Cambridge, MA: MIT Press)].

3. These concepts came to fruition as early as 1973 in Paul Mermelstein, "Articulatory Model for the Study of Speech Production." *Journal of the Acoustical Society of America* 53 (1973), 1070-83.

4. See Toby D. Griffen, "German [x]," *Lingua* 43 (1977), 375-90 both for a discussion of the structural and functional linguistic literature behind the controversy and for a dynamic analysis which renders the controversy vacuous.

5. This is, once again, the process of dynamic coarticulatory constraint central to the articulatory model of Mermelstein, "Articulatory Model for the Study of Speech Production."

6. Indeed, while the speech flow cannot be segmented automatically into phones, it can be (and has been) segmented automatically into syllables. See Paul Mermelstein, "Automatic Segmentation of

Speech into Syllables," *Journal of the Acoustical Society of America* 58 (1975), 880-83.

7. Once again, the consonantal and vocalic features do overlap in the vertical dimension of the spectrogram. However, as we shall see in section 4.4, the features that define consonantality and vocality within the scale do indeed divide rather neatly along the vertical plane.

8. In the prosodic analysis of the London school, the order was merely conventional. Such works as Firth, "Sounds and Prosodies;" R.H. Robins, "Aspects of Prosodic Analysis," *Proceedings of the University of Durham Philosophical Society* 1 (1957), 1-12; and Terence Mitchell, *Principles of Firthian Linguistics* (London: Longman, 1975) have contributed significantly to the development of the dynamic phonological model used here, and this contribution has been recognized and properly attributed in such works as Toby D. Griffen, *Aspects of Dynamic Phonology xxxxx*.

9. That insights should be gained by imparting a greater degree of functional linguistic methodology to Firthian phonology was predicted by André Martinet, *Phonology as Functional Linguistics* (Oxford: Clarendon Press, 1949).

10. Compare, once again, Firth, "Sounds and Prosodies." Of course, the major departure here from the Firthian approach is the removal of the "sounds," which actually represented segments.

11. Of course, different languages may add to these distinctions through configuration. For example, a retroflex tongue configuration (curled back as in the American English pronunciation of *r* [ɹ]) may

well serve to function alongside the simple positional features of the vowel. In our terminology, we accommodate such possibilities by referring to a *position* of articulation, rather than simply to a place of articulation.

12. Voicing is assumed as the natural state of the vowel unless it is suppressed in the "whispered" vowel. Such voiceless, whispered vowels may indeed function in the phonology of the language – see, for example, Toby D. Griffen, "The Swabian Voiceless Vowel," *Word* 34 (1983), 145-73.

13. Here again, we find a contrast with cross-sectional segmentation, which requires not simply an arrangement in the vertical dimension, but a *re*arrangement in the horizontal. Thus, while dynamic analysis can account for consonantal features that occur "below" the consonantal constraint in the spectrogram to show that they do function as obstructions, the traditional segmental analysis must shift features back and forth along the horizontal time continuum. This is the basic difference between the functional arrangement of the former, which does not alter the data, and the structural rearrangement of the latter, which does in fact alter the data.

14. This observation actually helped lead into the development of dynamic phonology – see Toby D. Griffen, "On Describing the Cluster Prosody," *LACUS Forum* 1 (1975), 140-47.

15. See, for example, Toby D. Griffen, "Toward a Nonsegmental Phonology," *Lingua* 40 (1976), 1-20; and *Aspects of Dynamic Phonology*.