Syllable weight, stress patterns and nativisation

Susan Fitt
sue@cstr.ed.ac.uk

Abstract

This paper looks at stress assignation for unknown foreign words read aloud by Scottish subjects. The syllable structure and syllable weight of the written words is examined, to determine where stress is most likely to be placed according to English stress rules, and the stress rules for the languages of origin of the words are also described. The stress patterns in the spoken responses are then analysed to determine the influence of syllable weight (both written and spoken), and the perceived language of origin of the word. Applicability of the findings to a wider context is discussed.

1 Introduction

'The stress rules of English are first of all riddled with exceptions and secondly fail the simplest attempts to confirm them directly, in many cases.' (Kahn 1980:134)

This paper focuses on stress assignment in unknown foreign names. In a series of experiments, 24 subjects (schoolchildren from Edinburgh, aged 13-17) were asked to read aloud a total of 60 town names from 6 European countries. They were also asked to say which of the six countries they thought each town was in.

2 Stress patterns and syllable weight

Hyman (1985) groups languages according to their stress patterns (Table 1 below). However, as regards syllable weight, English partly conforms to a group of languages which 'treats a syllable whose rime consists solely of a short (or lax) vowel as light, whereas a syllable whose rime has either a long (or tense) vowel and/or a final consonant (or more) is treated as heavy' (1985:5), and this can affect stress assignment.

In seeking to examine the issue of syllable weight and stress patterns, we must first note that application of the notion of 'syllable weight' depends first of all on defining the syllable structure. If we applied the syllable structure, for instance, of Wells, in which 'consonants are syllabified with whichever of the two adjacent vowels is more strongly stressed' (1990:xx), we would have a circular argument. Vowels would be stressed because the syllable ended in a consonant and was therefore strong, and the syllable would end

Language Type	No. of Languages
Languages with dominant initial stress (inc. German)	114
Languages with dominant second-syllable stress	12
Languages with dominant penultimate stress	77
Languages with dominant final syllable stress (inc. French)	97
Languages with no dominant stress placement (inc. English, Modern Greek)	113
Syllable weight stress	9
Antepenultimate stress	6
No stress or tone	16
Total languages included	444

Table 1: *Stress patterns in different languages.*

in a consonant because the vowel was stressed. (Wells' view, however, is phonetic rather than phonological, and for our purposes it is preferable to take phonological approaches to syllable definition.) There are, additionally, non-syllable-based approaches. Some theories of phonological weight, and the domain over which it applies, are briefly examined below. Only stress patterns for nouns are taken into account here, as it is assumed that town names fall into the class of nouns.

Chomsky and Halle (1968) allow all consonants following a vowel to be grouped with that vowel. In certain cases, however, they form a 'weak cluster', i.e. the group consists of 'a lax vowel followed by [r] or [w]' (1968:82). (There are also other kinds of weak clusters.) So, for example, the 'nd' affects the stressed second syllable of 'veranda', but the 'cr' of 'ludicrous' forms a weak cluster and so is unstressed. According to this theory, words such as 'vanilla' contain a geminate 'll', which forms a strong cluster and so is stressed. We can summarise as follows:

weak cluster	lax vowel followed by one or more consonants (where the second consonant may not be [r] or [w])
strong cluster	tense vowel followed by no consonants; or vowel followed by two consonants

For a noun, then, a vowel takes primary stress in the following cases (Main Stress Rule):

(1) $$V \longrightarrow [1stress/[x\text{---}(\begin{bmatrix} -tense \\ V \end{bmatrix} C_0^1 \begin{bmatrix} \alpha voc \\ \alpha cons \\ -ant \end{bmatrix}_0)\text{---}/$$
$$<_1<_2 +C_0 >_2 \begin{bmatrix} -tense \\ V \end{bmatrix} C_0 >_1] <_{1N}<_{2A}>_2>_1$$

In brief, stress is assigned to the antepenultimate syllable of a string ending in two weak syllables, or to a strong penultimate syllable.[1]

According to other theories, syllable structure is a factor in determining stress placement. In one framework, consonants followed by vowels are always syllable initial — 'the numerous syllabic frameworks have invoked principles that automatically result in a CV sequence being assigned to the same syllable' (Hyman 1985:2). Thus, VCCV can be VC.CV or V.CCV, but not VCC.V. If we used this syllable-structure theory to define the scope of the stress rules, we would be unable to allow Chomsky and Halle's analysis of 'veranda'. In this type of theory, the important notion is not weak vs. strong clusters, but closed vs. open syllables, with tense or lax vowels. 'Veranda' would be syllabified as 've-ran-da', and the penultimate syllable would be stressed as it is closed; if open, the antepenultimate syllable could be stressed (as in 'algebra'). Tense, or long, penultimate vowels are also stressed. As Kahn notes, the stress rule (2) taken from Main Stress Rule (1) 'will fail not only when C2 is a possible initial cluster of the type called 'weak' by SPE (viz., clusters consisting of C{r, w}) but when C2 is *any* cluster which is possible syllable initially' (1980:136):

(2) —— $C_2VC_0]_N$

Some theories, however, explicitly allow ambisyllabic consonants or clusters, for example Anderson and Jones (1974). In this account, 'all syllables have maximal initial and final clusters, regardless of the amount of overlapping entailed' (quoted in Kahn 1980:177). 'Boston', for example, consists of the syllables /bost/ and /ston/. However, Kahn claims that

> English stress assignment differs from many other phonological rules in applying at a level at which ambisyllabicity, necessary for accurate phonetic representations and relevant to the conditioning of certain low-level rules, is not yet appropriate. (1980:178-9)

If the rule above is taken to be true, then words such as 'cinema' will also fail the test, since the penultimate syllable is closed and should be stressed.

For others, syllable weight is related to the position in the word. Thus, for Giegerich on German,

[1] Chomsky and Halle do not use the term 'syllable', but 'string' and 'cluster' – 'syllable' is used here in order to give a brief summary.

> word medially a CVC sequence constitutes a heavy syllable and a CV sequence a light one, while word-finally CVCC sequences are heavy and CVC ones light. A CVC sequence is thus heavy in word medial position, whereas it should be interpreted as light in word-final position.
>
> (described in Kaminska 1995:115)

Hayes (also noted in Kaminska) goes further and claims that for English 'word final consonants simply do not count in the calculation of syllable weight; they are also not part of the syllable structure and therefore extrametrical'.

The rest of this paper concentrates on Chomsky and Halle's approach, and the syllable framework with initial pre-vocalic consonants.

2.1 Application of syllable weight to Scots

'Syllable weight is ...necessarily tied to the existence of a vowel length (or vowel tenseness) opposition.' (Hyman 1985:6). (No languages have been found in which open vs. closed syllables are the sole determiner for stress placement.) Unlike nearly all other varieties of English (Wells 1982), Scots has no phonological vowel length distinctions, (I am here discounting lengthening in certain morphological environments, such as 'brood' [brud] vs. 'brewed' [bruːd], and possible phonetic or allophonic lengthening before certain consonants). Diphthongs are then the only potentially long vowels in opposition to short vowels, and such an opposition is perhaps not the optimal description.

It might be preferable to use the feature [+tense] to distinguish vowels in Scots. However, 'the issue of formulating a definition of (the feature [±tense]) and specifying precisely its empirical correlates is complex and controversial ... Moreover, it seems to be the case that the postulation of the feature [±tense] can be more or less motivated (or necessary) depending on the phonological behaviour of segments in a specific language.' (Kaminska 1995:56.) In particular, the vowel /ɛ/ causes problems, as it would normally be classified as lax, but should perhaps be defined as tense due to its phonological and phonetic behaviour, such as appearing in open syllables, and a tendency to lengthen in certain environments. Kaminska, then, classifies Scots vowels as follows:

(3) tense /i/ /e/ /a/ /ɔ/ / o/ /u/
lax /ɪ/ /ʌ/ /ə/

/ɜ/, while occurring in middle-class Edinburgh speech, is not general in Scotland, but as it replaces the lax vowels of other accents it should perhaps be considered as lax. Diphthongs would also have to be classified as tense, as they are free vowels, leaving us four lax vowels.

(The definition and application of the feature 'tense' to our data is by no means straightforward — Chomsky and Halle (1968:75) explain the unstressed tense prevocalic [i] of 'various' by claiming that it derives from a phonologically lax vowel (/ɪ/); similarly, they claim that the final vowel of 'fiasco' is nontense, despite the fact that 'a nontense /o/ does not appear phonetically in the utterances of the dialect we are describing'. They 'specifically

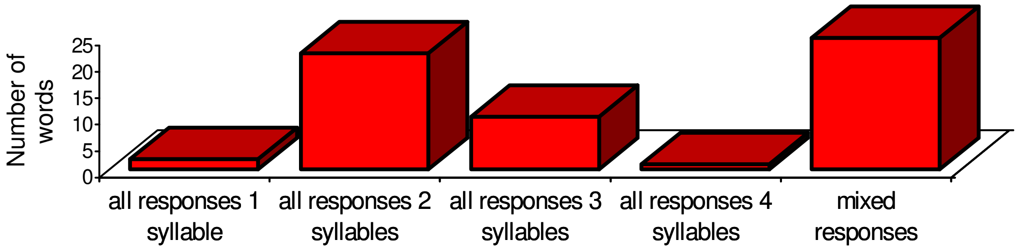

Figure 1: *Grouping of responses.*

reject the assumption that there must be a one-one relationship between the underlying lexical or phonological representation and the phonetic output, and we see no reason to suppose that underlying representations will be restricted to segments that appear in phonetic representations. Such a requirement would, in fact, be quite artificial and ad hoc.' However, application of such a theory needs to be very careful in order not to be 'artificial and ad hoc' itself.)

3 Results of experiments

A number of responses are not included in the results here. All responses for which the response string did not 'match the prompt' were rejected, including epenthesis, metathesis, stumbling, and so on, unless the epenthesis or omission occurred to resolve a non-English grapheme/phoneme sequence (e.g. <#ps> → [pʰə.s], [s] etc.) 'Matching the prompt' refers to a possible pronunciation of the orthography, where a liberal interpretation is made of 'possible' grapheme-segment correspondences, i.e. they have to be conceivable in the languages in the study, rather than extant. Examples of rejected responses are:

(4) Snåsa [ˈsan.sə]
Loano [loˈna.no]

The number of syllables in responses varied from 1 to 4, and can be grouped as in Figure 1. (As the words did not have similar structures or lengths, and were not random selections, Figure 1 does not necessarily show any general tendencies.)

Of the responses which had a consistent number of syllables, (the first four columns in Figure 1) and in which the stress could be confidently determined[2] the stress patterns fall as in Figure 2, in which it can be seen that there is a clear preference for penultimate stress.

[2]In some cases, and for some subjects in particular, stress was unclear.

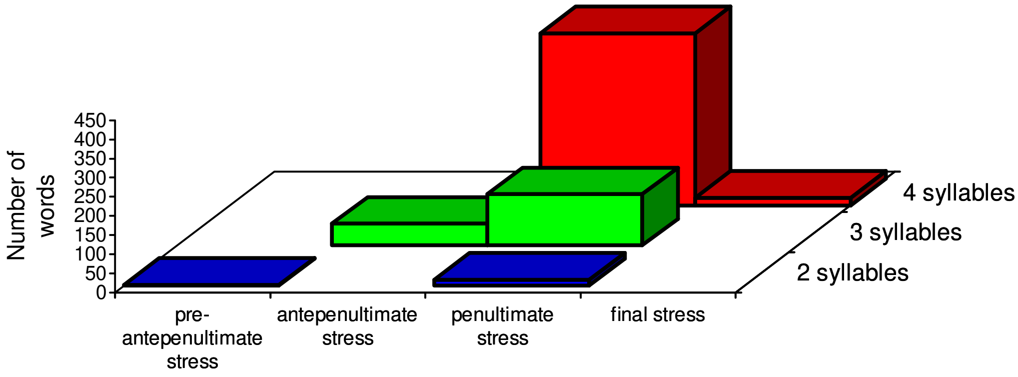

Figure 2: *Stress patterns.*

Town	Penultimate stress	Final stress
Velen	19	4
Acri	17	2
Slattocks	20	2
Sollom	19	4

Table 2: *Final stresses in bisyllabic words.*

3.1 Bisyllabic words

Of the 22 words which had only bisyllabic responses, all had a clear preference for penultimate stress.[3] A number of words elicited final stress from one subject, but this can be discounted. The words which elicited final stresses from more than one subject are as in Table 2. This still shows a low incidence of final stresses, but, in the case of 'Velen' and 'Sollom', is not unimportant. Unfortunately for any of the above theories of syllable weight and stress, however, the application of final stress cannot be explained by syllable structure, since final syllables of nouns do not take stress. They must therefore be treated as exceptions. But, why should a number of subjects treat an unknown word as an exception? (It should also be noted that phonologically and graphemically, these words do not form a separate class from others in the set, such as 'Tallard', 'Dokka' or 'Schapen'.) Also, the phonemic strings assigned by the subjects differ only in reduction of unstressed vowels, which would occur after stress assignment.

Two explanations suggest themselves. One is the operation of analogy, and another is that the perceived origin of the words influenced the stress patterns. 'Shalom', with

[3]This is described as 'penultimate' rather than 'initial' as it fits the overall description better.

	Sollom		Velen	
Perception of origin	**Penultimate stress**	**Final stress**	**Penultimate stress**	**Final stress**
Britain	4	1	1	0
France	4	1	3	2
Germany	5	0	4	0
Greece	3	1	4	0
Italy	1	1	4	1
Norway	2	0	3	1

Table 3: *Perceptions of origin for 'Sollom' and 'Velen'.*

final stress, is an obvious candidate for analogy with 'Sollom', but 'Solemn' is still more obvious (and is in fact homophonous with 'Sollom', which is an English town)? It could be the case, however, that some subjects did think of 'Shalom'. As for 'Velen', there are no obvious analogies which would lead to [vəˈlen].

As for perceptions of origin (see Table 3), it might be thought that towns which were thought to be French might be given final stress, but this is not supported by the data, which have a fairly random distribution.

Another possibility is that towns which do not have an overall consensus as to origin behave differently; further analysis is necessary to determine this.

3.2 Trisyllabic words

Stress patterns assigned to trisyllabic words are less homogeneous, but more interesting. They can be grouped as in Table 4 below. The pre-penultimate stress of 'Bobbio' could be explained by Chomsky and Halle by proposing that the [i] derives from a lax vowel (c.f. 'various' above). In syllabic analyses, as the second syllable is open, it is potentially weak and therefore unstressed, but this ignores the tense/lax distinction. In all but one case of pre-penultimate stress, the vowel of the second syllable was [i], whereas in both cases of penultimate stress it was given a diphthong. By Kaminska's analysis above, /i/ is a tense vowel and so no distinction is possible between the two. An analysis of [i] as deriving from a lax vowel might account for this, but a fuller analysis would be necessary to justify this supposition. It is possible that it cannot be justified, as external evidence, such as morphologically-related words, is lacking. A length distinction would also give an opposition of [i] to [ʌɪ] to provide a possible explanation.

'Livorno' and 'Valençay' can be explained by any of the above accounts. For Chomsky and Halle, the second syllables form strong clusters, while for other accounts the second syllable is closed and therefore stressed. 'Bolkesjö' may also be explained in this way, since subjects all pronounced <sj> as a sequence of segments, such as [sʤ] or [zj], rather than a

	Pre-Penultimate stress	Penultimate stress	Final stress
Mostly penultimate stress			
Livorno	1	20	0
Valençay	3	16	0
Bolkesjö	4	10	0
Larisa	1	21	0
Megara	4	18	0
Novoli	1	19	0
Copparo	3	19	0
Ekhinos	7	13	0
Mostly pre-penultimate stress			
Bobbio	20	2	0
Osimo	12	7	0

Table 4: *Stress patterns of trisyllabic words.*

single segment such as [ʃ]. Disappointingly, though, there was no difference in the syllabic structure of the four responses which gave pre-penultimate stress.

A problem arises from the grouping of 'Larisa', 'Megara' and 'Novoli', with penultimate stress favoured, and 'Osimo' with a tendency to pre-penultimate stress. There is no obvious difference in phonemic or graphemic structure, except the lack of an initial consonant, which is universally held to be of no account. All subjects giving penultimate stress to 'Osimo' used [i], and most of those giving pre-penultimate stress used [ɪ]; the two can be distinguished by the feature 'tense', in Chomsky and Halle's framework. However, there were some pre-penultimate examples with [i], and some penultimate examples of 'Larisa' with [ɪ], which being lax should not be stressed.

Perceived origin of the words may have some effect, if some of the subjects who thought 'Osimo' was in Norway associated it with 'Oslo', and so stressed 'Os' (see Table 5).

'Copparo' and 'Ekhinos' fall into the same category as the 'Megara' group above. Although they have geminate consonants in the first syllable in the graphemic string, which is potentially important in a Chomskyan framework, the stress-assignment rules apply from right to left and so the first syllable would be of no account. Penultimate stress assignment in all these words would be adequately accounted for by Chomsky and Halle, as the penultimate vowels are tense and followed by consonants, forming strong clusters. It could also be explained by syllabic structure plus the feature tense (all the penultimate syllables are open, so the vowel itself determines the syllable weight). However, this time it is the

	Britain	France	Germany	Greece	Italy	Norway
Penultimate stress						
Megara	0	1	2	8	3	4
Novoli	0	0	0	1	15	3
Larisa	0	5	0	3	12	2
Osimo	0	0	1	3	1	2
Pre-penultimate stress						
Megara	0	0	0	1	3	0
Novoli	0	1	0	0	0	0
Larisa	0	0	0	0	1	1
Osimo	0	1	0	4	1	6

Table 5: *Stress patterns and perception of origin for 'Megara', 'Novoli', 'Larisa' and 'Osimo'.*

length feature as an indicant of syllable weight, rather than the tense feature, which fails to account for the data.

Also worth noting are the phonetic strings assigned to 'Osimo'. All subjects who used penultimate stress used the vowel [i] in the second syllable, while most of those giving pre-penultimate stress used [ɪ] (though some did use [i]). Consonants fell into two clear categories: those using penultimate stress pronounced <s> as [s], i.e. [ɔˈsi.mo], while those using pre-penultimate stress pronounced <s> as [z], i.e. [ˈɔ.zɪ.mo]. Unfortunately this could only precede stress assignment if we assumed that <s> is voiced due to it being syllable final, and that this syllable structure preceded stress assignment. It is more likely that it is the stress which causes <s> to precede an unstressed syllable and have voicing assimilation to [z].

3.3 Quadrisyllabic words

Only one word was universally quadrisyllabic, 'Katerini'. All but two subjects gave this penultimate stress; the remaining two used initial stress. Both of these had [i] in the penultimate syllable, rather than [ɪ], which could be thought in some accounts to reject stress by virtue of being lax. These two answers are therefore not accounted for by the above theories (unless we can postulate an underlying lax vowel).

3.4 Mixed answers

Seven words ('Rötz', 'Rede', 'Dreve', 'Nahe', 'Greve' 'Aire, and 'Glinde') had either monosyllabic or bisyllabic responses In all cases the bisyllabic responses were given penultimate stress, as would be expected. Responses to other names form more complex patterns.

3.4.1 Two/three syllables

Loano: In the majority of cases, was pronounced as either [ˈlo.no] or [loˈa.no]; in all responses, penultimate stress was given. The trisyllabic responses fall into the 'Megara' group above.

Lechlade: This word was given a variety of consonants to represent 'ch', including [ʃ], [k], [x] and [ʧ], and a variety of vowels, but responses basically fell into two categories: [ˈlɛx.led] and [ˈlɛx.la.də] / [lɛxˈla.də]. In this case, country of origin does provide a possible explanation of the [ˈlɛx.la.də] pattern, as all 5 of these responses were assigned to Germany and follow German stress patterns (i.e. dominant initial stress, see Hyman 1985).

Hellesylt, Ålesund: Trisyllabic cases all had pre-penultimate stress. This is easily explained by the lax vowel [ə] in the second syllable (unless we have to assume an underlying representation of, say, [ɛ]).

Schwenke, Pfinztal: All responses had penultimate stress, as would be expected in all analyses. Trisyllabic responses were due to epenthetic vowels inserted in the word-initial consonant clusters.

3.4.2 Other responses

Maule: Monosyllabic, bisyllabic and trisyllabic answers were given. All multi-syllabic answers stressed 'Ma' - the trisyllabic answer [ˈma.u.lə], with a succession of vowels rather than a diphthong, is slightly problematic as the [u] vowel is tense and so should be stressed (unless it is assumed to derive from a lax vowel).

Psakhna: Most answers were bisyllabic, with penultimate stress, but three other responses (of 3-4 syllables, with epenthetic vowels in the consonant clusters) were also elicited. [pʰa.səˈkʰa.nə] and [p͡saˈkʰa.na] are explainable, but not [ˈpʰa.kʰa.na], which has three syllables of identical structure (or two equal sequences, VC, in Chomsky and Halle's framework), yet the penultimate syllable is not stressed.

Elatia, Evje: For 'Elatia', trisyllabic responses (such as [ɪˈle.ʃə]) all had stress on the penultimate syllable. Quadrisyllabic responses ([ɪˈla.tʰi.ə]) all stressed the pre-penultimate syllable, if the schwa is thought to form a separate syllable. 'Evje' falls into a similar pattern, but with bi/trisyllabic responses (and some monosyllabic). The quadrisyllabic responses for 'Elatia', and the trisyllabic for 'Evje', are similar to 'Bobbio' above.

Laragne, Savigne, Manosque: Bisyllabic responses generally had final stress, and trisyllabic responses penultimate stress. For these examples it is the bisyllabic responses which are problematic for the description. The obvious answer is that they are perceived to be French and given French final stress; all six 'Laragne' with final stress were given French responses, though the figures are only 2/5 for the finally-stressed 'Savigne', and 5/12 for 'Manosque'. (It is of course possible, that even for those subjects who did not think 'Savigne' was French, they were subconsciously aware of the group of words they belong to, and are similar to.)

Tsamandas: This had a variety of trisyllabic and quadrisyllabic responses, with the general consensus (9/14) falling on trisyllabic responses with penultimate stress, as would be expected. The Greek original is in fact stressed on the final syllable, but subjects are unlikely to know this.

Kvernes, Karousadhes: These elicited a variety of syllable structures and stress patterns, but with penultimate stress predominant.

4 Conclusions

Linguistic theory should be able to explain the pronunciation of unknown words, particularly in cases where the subjects agree on the pronunciation. In cases where they disagree, we also need to determine *why* they disagree. The application of the notion of syllable weight for determining stress placement in these names is problematic, not least because of the possibility of rule ordering. Also, for unknown words, with no obvious morphological structure, and only one token provided by each subject, it is difficult to justify an analysis resting on the interpretation of underlying segments, unless we say these are present by analogy with similar words.

Definition of weight is a further problem. Notions of tense vs. lax, or long vs. short vowels do not seem to describe the data adequately, either in a syllabic or a non-syllabic framework. While application of rules to strings of phonemes describes the data better than appealing to the notion of syllables (at least if all CV strings belong to the same syllable), neither is adequate. Further work is needed to see if the word-position of the syllable affects stress assignment in the data here (for example, whether CVC syllables are heavy word-initially but light word-finally).

Additional explanations of analogy and use of foreign language knowledge help to explain some of the responses, but a quantitative study of the data is needed to see how generally applicable this is.

References

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Hyman, Larry M. 1977. On the nature of linguistic stress. In Larry M. Hyman (ed). *Studies in Stress and Accent*. Los Angeles: University of Southern California, Department of Linguistics, pp.37–82.

——. 1985. *A Theory of Phonological Weight*. Dordrecht: Foris Publications.

Kahn, Daniel. 1980. *Syllable-based Generalizations in English Phonology*. New York: Garland Publishing.

Kaminska, Tatiana Ewa. 1995. *Problems in Scottish English Phonology*. Tübingen: Niemeyer.

Laver, John. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.

Sloat, Clarence, Sharon Henderson Taylor & James E. Hoard. 1978. *Introduction to Phonology*. Englewood Cliffs, NJ: Prentice-Hall.

Wells, John C. 1990. *Pronunciation Dictionary*. Longman.