Sounds difficult? Why phonological theory needs ‘ease of articulation’
David Shariatmadari
163289@soas.ac.uk

Introduction
In this paper I will try to show that theories of phonological structure must incorporate some measure of phonetic naturalness, specifically the idea that there exists a tendency to conserve energy in the use of the articulatory organs, with ‘easy’ sounds being those that require less physical effort to produce on the part of the speaker. I will call this the Ease of Articulation Hypothesis (EoA)\(^1\). A strong form of EoA would state that articulatory phonetics is the sole motivating factor for sound patterns including the structure of phonemic inventories, phonotactics and morphophonemic variation. This is clearly false. If it were the case, phonology would be indistinguishable from phonetics. There would be no reason for a given phonological process not to apply in all languages, since all human beings have the same vocal apparatus and what is phonetically natural for a speaker of Turkish must also be phonetically natural for a speaker of Japanese. And yet, there are clearly many phonological differences between these two, and among all other languages.

In contrast, a weak version of EoA might hold that articulation is one of a number of pressures competing for influence over the shape of the speech string. Some of the time it will win out. Whether or not it does depends, crucially, on the structure of the language concerned. Since every language has a different structure, the way in which phonetic influence is felt will be different for every language. The structuralist distinction between ‘substance’ and ‘form’ is relevant here. We can imagine the various substances of which language is moulded passing through a formal ‘filter’ out of which the sound system emerges. Phonetic substance is just one of the raw materials. It is joined by others, including morphosyntax and semantics, all of which have their own demands to make.

Opponents of EoA have made the mistake of using arguments which apply to the strong version of the hypothesis against all attempts to incorporate a measure of articulatory difficulty into phonological theory (Kaye 1989, Ploch 2003). A weak version of the EoA is to my mind, not only more credible than its strong counterpart, but essential in the analysis of sound patterns.

After setting out the theoretical justification and linguistic evidence for this argument, I will examine counterclaims. I will look specifically at Kaye’s critique of EoA, which he makes from the standpoint of Government Phonology (Kaye, Lowenstamm & Vergnaud 1985, Charette 1991, Kaye 1989, 1992, 1997). Kaye’s characterisation of this framework rejects all forms of EoA, opting instead for a system of principles and parameters in which all phonological patterns can be derived from the universal grammar. I hope to demonstrate that plausible explanations for these patterns exist externally.

---
\(^1\)There is a potential for confusion in the term ‘phonetic hypothesis’ as used by Jonathan Kaye to describe this idea (Kaye 1989). Phonetics can be regarded as comprising both the articulatory and acoustic/auditory analysis of sounds. It is primarily articulatory factors that concern me here.
Why minimum effort matters
The vocal tract is a motor system of the human body (though not a self-contained one) and we should expect it to be subject to the same laws as other motor organs involved in skilled movements. One of the laws appears to be a principle of minimum effort (Nelson 1983). The advantages of conserving energy for an organism that cannot rely on a steady supply of food are clear. It is these advantages that give rise to a tendency to economise, all other things being equal. “Most of the speech articulators are capable of generating higher forces (expending greater amounts of effort) than are common in speech” (Perkell 1997: 335, parentheses in the original). The fact that they do not do so, and that not every utterance is maximally clear and carefully articulated, is significant. A compromise is reached: a lot of energy is not expended unless the situation really demands it. In Björn Lindblom’s words, “unconstrained, a motor system tends to default to a low-cost form of behaviour.” (Lindblom 1989: 413).

So, what kind of things constrain us, in that we do not generally sit around doing nothing all day? Hunger, thirst and the instinct to mate act against minimum effort as far as the gross musculature is concerned. What are the analogous motivating factors for language? The complexity of the world around us, and the correspondingly complex system of sounds needed to describe it mean that we would be unlikely to get much benefit from speech if it consisted of grunts alone. This, coupled with a need to ensure that our system, once in operation, is clearly understood, act against economy to stimulate effort. Can a stable equilibrium between the opposing tendencies be reached? This is impossible for at least two reasons. Firstly, the communication channel is noisy: sounds may be imperfectly heard and reanalysed, before being passed on in their new form to others. Secondly, the linguistic system interacts with social systems in a way that forces it to change shape. New words are required for new objects; certain forms become associated with certain social groups; migration brings previously isolated dialects into contact with one another. All of this induces dynamism.

At the everyday level, the process of communication is constrained by context and familiarity. Helmut Lüdtke illustrates this with the example of a group of Germans in search of a missing object (Lüdtke 1989). If one of them finds it he is likely to alert his companions by using a reduced form that requires less energy, such as [xaps] for *Ich habe es*, ‘I have it’. In a noisy room, in which there is a greater risk of being misunderstood, he might revert to the longer [ixʔ habøʔes]. In [xaps] we have, among other changes, the loss of two empty onsets and the devoicing of the word final /b/. The low-cost behaviour that results in abbreviation of this type is a universal trait. If a weak version of EoA holds, we might expect it to be systematised, that is, to be reflected in the phonology. Indeed, cross-linguistically, empty onsets are rarer than filled onsets and word-final devoicing is an extremely common process. The latter is a classic minimum effort strategy. As Lindblom states, “devoicing of final consonant [occurs] in anticipation of the abducted state of the glottis associated with normal breathing and silence” (Lindblom 1981: 237).

Meillet’s celebrated example of the reduction of Latin *hoc die* ‘this day, today’ to Old French *hui* ‘today’ on account of its high frequency illustrates the same tendency (Meillet 1938: 131). There is rarely any need to articulate such a common word as *today* carefully; its occurrence is widely expected and an interlocutor’s desire to use the word can be gleaned from the merest of hints. Thus, the awkward transition from
voiceless to voiced consonant is elided to leave a diphthong. What we are seeing here is, in essence, the ‘phonologisation’ of phonetic ease of articulation phenomena².

To summarise: minimum effort can never be given free rein – the result would be silence. But where it can, it makes its influence felt.

**Seeing the phonetics in phonology**

Motor economy is an extra-linguistic constraint associated with the parts of the body used to create speech sounds³. It is almost inconceivable that this tendency would not have some reflex in the formal system in which the vocal apparatus are engaged during speech. It need not define it completely - as previously stated, this would result in a situation where there is no difference between the phonetic and phonological domains - but is likely to be detectable.

**Phoneme inventories**

John Ohala notes that the ease with which stops may be voiced depends on certain phonetic factors (Ohala 1997). These include the place of articulation of the stop. Voicing depends on subglottal air pressure being lower than supraglottal pressure. Otherwise, air cannot move outwards past the glottis. Supraglottal air pressure rises when a constriction is made in the oral cavity. If this constriction is made towards the front, it will take longer for the pressure to equalise as the volume of the space enclosed is greater. Thus, voiced velar or uvular stops are more difficult than voiced labial stops – more difficult since it would require additional effort to distend the oral cavity to the extent that it would be able to accommodate voicing at these places of articulation. These phonetic facts are reflected in the phonological inventories of the world’s languages. Where gaps in stop inventories exist, they tend to be for voiced phonemes at back places of articulation or voiceless phonemes at front places. For example, the Niger-Korfordanian language Efik has a voiced bilabial stop /b/ but no voiceless counterpart and a voiceless velar stop /k/ but no voiced counterpart, despite exhibiting a voiced-voiceless contrast for other stops. Thai, which also has a voiceless-voiceless contrast elsewhere, has a /k/ but no /g/, as does the Mexican language Chontal. The total distribution in the UCLA Phonological Segment Inventory Database (UPSID) of /g/, which is relatively harder to produce, is more limited than its voiceless counterpart /k/ - 175 tokens compared with 283 (Maddieson 1984).

Since voicing requires air to flow unimpeded across the glottis, voicing during any prolonged closure (for example, geminates or stop clusters) is difficult. Again, this is reflected cross-linguistically. In Japanese, only the voiceless consonants in the pairs /t-d/, /k-g/, /tʃ-dʒ/ /s-z/ are found long. In a similar way, the Caucasian language Lak makes a short-long contrast only among its voiceless consonants (1). In the UPSID database, the distribution of /tʃ/ compared with /dʒ/ is 7 and 3 tokens respectively. For /k:/ and /q:/ it is 9 and 4 tokens respectively.

---

² Both Lüdtke and Meillet’s examples are discussed in relation to the development of languages in Shariatmadari (2004).
³ I will not attempt here to deal with the complex issue of how minimum effort might be more precisely defined. A discussion of possible systems for the physical measurement of articulatory effort is provided in Chapter 2 of Kirchner (2001).
The consonants of Lak, which exhibit a short-long contrast only for voiceless segments (from Maddieson 1984).

<table>
<thead>
<tr>
<th>Place</th>
<th>Bilabial</th>
<th>Dental/alveolar</th>
<th>Dent./Alv. labialised</th>
<th>Palato-alveolar</th>
<th>Palatal</th>
<th>Velar</th>
<th>Uvular</th>
<th>Pal.-Alv. labialised</th>
<th>Pharyngeal</th>
<th>Glottal</th>
<th>Variable Place</th>
<th>Labial-Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless plosive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>long vl. plosive</td>
<td>p:</td>
<td>t:</td>
<td>k:</td>
<td>k′:</td>
<td>q:</td>
<td>q′:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vl. asp. plosive</td>
<td>pʰ</td>
<td>tʰ</td>
<td>kʰ</td>
<td>kʰ′</td>
<td>qʰ</td>
<td>qʰ′</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>voiced plosive</td>
<td>b</td>
<td>d</td>
<td>g</td>
<td>g</td>
<td>g′:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vl. ejec. stop</td>
<td>p′</td>
<td>t′</td>
<td>k′</td>
<td>k′:</td>
<td>q′</td>
<td>q′:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vl. asp. sib. aff.</td>
<td>t’sʰ</td>
<td>t’s</td>
<td>t’ʃ</td>
<td>t’ʃ:</td>
<td>t’ʃ′</td>
<td>t’ʃ:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>long vl. sib. aff.</td>
<td>t’s</td>
<td>t’s:</td>
<td>t’ʃ</td>
<td>t’ʃ:</td>
<td>t’ʃ′</td>
<td>t’ʃ:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vl. sib. ejec. aff.</td>
<td>t’s’</td>
<td>t’s′</td>
<td>t’ʃ′</td>
<td>t’ʃ′:</td>
<td>t’ʃ′:</td>
<td>t’ʃ′:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vl. nonsib. fric.</td>
<td>x</td>
<td>x’</td>
<td>x</td>
<td>x:</td>
<td>x:</td>
<td>x′:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>long -</td>
<td>x</td>
<td>x:</td>
<td>x′</td>
<td>x:</td>
<td>x:</td>
<td>x:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vl. sib. fric.</td>
<td>s</td>
<td>ʃ</td>
<td>ʃ</td>
<td>ʃ:</td>
<td>ʃ:</td>
<td>ʃ:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>long -</td>
<td>s:</td>
<td>ʃ:</td>
<td>ʃ</td>
<td>ʃ:</td>
<td>ʃ:</td>
<td>ʃ:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vd. sib. fric.</td>
<td>z</td>
<td>ʒ:</td>
<td>ʒ:</td>
<td>ʒ:</td>
<td>ʒ:</td>
<td>ʒ:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vd. Nasal</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vd. Trill</td>
<td>r</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vd. lat. approx.</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>vd. central approx.</td>
<td>j</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Phonotactics**

As far as sequences of sounds are concerned, the more significant the change in place or manner of articulation from one segment to the next, the more difficult it is to articulate. EoA predicts that this will be measurable to some extent in the phonology. In languages which allow clusters of consonants, certain patterns are more common than others. These patterns need not be stored in the universal grammar, and there is some reason to believe that they are phonetically motivated. Strong cross-linguistic evidence exists for the view that syllables tend to form a ‘sonority peak’, and the condition that governs the distribution of consonants either side of this peak is that they be less sonorous towards the syllables boundaries. Sonority can be said to correspond to a sound’s “loudness relative to that of other sounds of the same length, stress and pitch” (Ladefoged 2001: 227). The usefulness of oscillations in sonority in the speech string is clear when we consider the need to employ phonetic contrasts in order to maintain a capacity for semantic differentiation. But is there any communicative reason that the transition between high and low values embodied by the sonority slope be smooth? Perceptual distinctiveness would be better served by an abrupt change. In fact, only an explanation couched in terms of physiological constraints predicts a gradual modulation between the nucleus and its neighbours.

As follows from the principle of minimum effort, “extreme displacements and extreme velocities are avoided [by the articulatory system]” (Lindblom 1981: 231).
There is every reason to expect constraints such as these to be reflected in the concatenation of segments, and hence, in syllable structure. Lindblom conducted an analysis of Swedish consonants based on the jaw movement required to produce each one. The distribution of these consonants along a measure of jaw position conformed very closely to that of their sonority, as inferred from their phonotactic behaviour (2). These data support the view that the smooth transition characteristic of cluster-nucleus-cluster groupings is physiologically motivated. There is no jump between widely separated jaw positions. Instead, in the sequence $C_1C_2VC_2C_1$, occupants of the $C_2$ position must be intermediate, in terms of movement, between those of $C_1$ and $V$.


Assimilation

From phonotactics we move to the effect adjacent segments may have on one another over time - assimilation. This is a much studied historical process. Examples of segments becoming more alike are extremely numerous. If one effect of minimum effort is to smooth the transition from consonant to vowel as the preceding discussion showed, we can imagine that if this principle were to win out, we would have a situation in which there was no transition at all. Wherever this can happen within the constraints imposed by the need to communicate effectively, it does.

Data from Latin and its descendent, Italian, illustrate this point (2). In these etymologically related examples, the sequences /kt/ and /pt/ become /tt/ and the sequence /bs/ becomes /ss/. These changes favour minimum effort. As Lindblom puts it, “we see that assimilation, defined as reduced distance between two sequentially timed articulatory targets, implies less work per unit time” and thus, less effort (Lindblom 1981: 237).

---

It is interesting to note that sonorant clusters derived from Latin remain unassimilated in Italian, in words such as *paulcre* ‘to speak’, *ornamento* ‘decoration’, *permere* ‘to permeate’. Defined according consonant strength in (4), sonorants are less effortful than stops and the need to eliminate clusters in which they participate is correspondingly weaker.
(3) Assimilation of Latin obstruent clusters in Italian.

<table>
<thead>
<tr>
<th></th>
<th>/kt/ &gt; /tt/</th>
<th>/pt/ &gt; /tt/</th>
<th>/bs/ &gt; /ss/</th>
</tr>
</thead>
<tbody>
<tr>
<td>factum</td>
<td>‘deed’</td>
<td>acceptus</td>
<td>absens</td>
</tr>
<tr>
<td>doctor</td>
<td>‘teacher’</td>
<td>corruptus</td>
<td>subsulto</td>
</tr>
<tr>
<td>fatto</td>
<td>‘fact’</td>
<td>accetto</td>
<td>assente</td>
</tr>
<tr>
<td>dottore</td>
<td>‘doctor’</td>
<td>corrotto</td>
<td>sussulto</td>
</tr>
</tbody>
</table>

Lenition

The process of lenition is said to involve the ‘weakening’ of consonants. As the following table shows, weaker consonants involve less effortful articulations.

(4) Scalar values for consonant strength (adapted from Trask 1996: 56):

<table>
<thead>
<tr>
<th>Stronger &gt; weaker</th>
<th>Articulatory effect of weakening</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. geminate &gt; simplex</td>
<td>shorter duration</td>
</tr>
<tr>
<td>2. stop &gt; fricative &gt; approximant</td>
<td>less obstruction of airflow, gestural reduction</td>
</tr>
<tr>
<td>3. stop &gt; liquid &gt; approximant</td>
<td>less obstruction of airflow, gestural reduction</td>
</tr>
<tr>
<td>4. oral stop &gt; glottal stop</td>
<td>minimal use of articulators</td>
</tr>
<tr>
<td>5. non-nasal &gt; nasal</td>
<td>less obstruction of airflow through nasal cavity</td>
</tr>
<tr>
<td>6. voiceless &gt; voiced&lt;sup&gt;6&lt;/sup&gt;</td>
<td>easier transition to and from vocalic segments</td>
</tr>
</tbody>
</table>

The distribution of lenition prone consonants is not random, and has been the subject of much discussion. Any phonetic account must explain why it occurs predominantly in codas. Juliette Blevins views the environments in which lenition occurs as “weak prosodic positions” (Blevins 2004: 145<sup>7</sup>), a correlate of which is reduced duration. Articulatory undershoot, a minimum effort strategy, has been observed to occur when the time allocated to a segment is shortened, as will be discussed further in the following section. It may well be the underlying cause of consonant lenition.

The weakening of oral stops to glottal stops, called debuccalisation, is a common change. In certain varieties of English, /θ/ is substituted for /t/ medially and word finally resulting in forms such as /wɔθə/ and /wɔθ/ for water and what. Since total failure to pronounce a segment can be regarded as the ultimate in motor economy, the fact that debuccalisation is often a prelude to complete consonant loss (Blevins 2004: 121) suggests that it is indeed a process motivated by minimum effort.

In Brazilian Portuguese the lenition of the velarised /θ/ to the velar glide /w/ results from a reduction of the tongue-raising gesture associated with the former segment.

---

<sup>5</sup> Blevins provides an alternative, perceptual account of total assimilation in the environment VC<sub>1</sub>C<sub>2</sub>V. Briefly, in a sequence of stops, the second, occurring as it does before a vowel, is more perceptually salient. Its place features get associated with the entire cluster, resulting in an analysis of the sequence by listeners as VC<sub>1</sub>C<sub>2</sub>V (Blevins 2004).

<sup>6</sup> The difficulty of voicing cannot, perhaps, be reduced to a two place scale in this way, and must depend partly on environment, partly on the nature of the segment: intervocically, voicing should be easier. However, the voicing of stops, as we have seen, is relatively difficult, as is word-final voicing.

<sup>7</sup> Blevins admits that “prosodic conditions for fortition and lenition require much further study” (Blevins 2004: 145n).
Thus words such as *tribunal* ‘tribunal’ and *carnaval* ‘carnival’, which retain ‘l’ in the orthography, are pronounced with a final /w/. The same phenomenon can be observed in some English dialects, in which the /fl/ of *feel* and *animal* and the medial /fl/ of *also* emerge as /w/. The ‘silent’ *l* in the orthography of many English words reflects a velarised /fl/ that was lenited to /w/ before merging completely into the preceding vowel. Examples include *walk* /wɔk/ *talk* /tɔk/ *half* /hæfl/ and *yolk* /jɔuk/ (Trask 1996: 61).

As with assimilation, contrasting processes are observed which have the opposite effect to lenition. However, we find that they are “much less frequent” (Trask 1996: 60), indicating that the tendency to minimise effort, though not invincible, is a strong one.

**Morphophonemic variation**

We turn now to the interaction of phonology with the morphological level of analysis. As we have seen, there are certain phonetic constraints on the articulation of voiced stops. These constraints result in morphophonemic variation in some languages. For Nubian, the addition of a morpheme meaning ‘and’ has the effect of doubling the final consonant of the preceding word (5).


<table>
<thead>
<tr>
<th>Noun stem</th>
<th>Stem + and</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>/fab/</td>
<td>/fabːn/</td>
<td>father</td>
</tr>
<tr>
<td>/seged/</td>
<td>/segetːn/</td>
<td>scorpion</td>
</tr>
<tr>
<td>/kadʒ/</td>
<td>/katʃːn/</td>
<td>donkey</td>
</tr>
<tr>
<td>/mʊɡ/</td>
<td>/mʊkːn/</td>
<td>dog</td>
</tr>
</tbody>
</table>

As we noted in our discussion of the occurrence of geminates in phoneme inventories, prolonged voicing of stops is difficult as the air needed to vibrate the glottis has nowhere to escape. The further back in the oral cavity the occlusion is made, the stronger the effect. In these examples, the voiced consonants become voiceless when lengthened. Only /b/ has a geminate reflex, presumably, Ohala suggests, because it has the furthest forward articulation.8

Articulatory ‘undershoot’ is a concept used by Lindblom in his analysis of rapid speech. More effort is required to hit the same target when the rate of speech is increased. Unless the communicative act is likely to be jeopardised by some measure of articulatory laxing (for example, if the interlocutor is not a native speaker) this extra energy will not be expended, in line with the principle of minimum effort. The result is ‘hypoarticulation’ which Lindblom contrasts with careful speech or ‘hyperarticulation’ (Lindblom 1989). Hypoarticulation effects include the raising of F1 and F2 values for vowels. This reduces the salience of high-low and front-back oppositions. Vowel reduction in unstressed syllables is a process common to many

---

8 Ohala doesn’t refer to the fact that the phonemic inventory of Nubian lacks /p/, according to Maddieson (1984). This undermines somewhat the claim that /b/ ‘resists’ devoicing - the language in fact has no other option. However, the data still support the force of Ohala’s argument.
languages and has been linked to the shorter duration of unstressed portions of the speech string (Flemming 2005).

(6) The vowels of Italian in primary stressed syllables and elsewhere (adapted from Flemming 2005: 2).

<table>
<thead>
<tr>
<th>i</th>
<th>u</th>
<th>i</th>
<th>u</th>
<th>e</th>
<th>o</th>
<th>e</th>
<th>o</th>
<th>e</th>
<th>o</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{array}{ll}
/i/ & 'vino' \text{ ‘wine’} \\
/u/ & 'kura' \text{ ‘care’} \\
/e/ & 'peska' \text{ ‘fishing’} \\
/o/ & 'nome' \text{ ‘name’} \\
/e\text{-e}/ & 'be\text{\textsuperscript{E}}lo' \text{ ‘beautiful’} \\
/o\text{-o}/ & 'ma\text{\textsuperscript{E}}le' \text{ ‘soft’} \\
/a/ & 'mano' \text{ ‘hand’} \\
\end{array}
\]

The data in (6) show the vocalic inventory of Italian in both stressed and unstressed syllables. The unstressed inventory is smaller. This can be explained as the result of the combined demands of minimum effort and perceptual clarity. The former ensures that the extra work required to maintain vowel quality when duration is short is not done. Consequently, F1 and F2 contrasts are reduced with the effect that the central region of the vowel space becomes crowded. The latter, in the form of a constraint on the minimum distance between vowels, means that some in this crowded area must be eliminated. In this way, the tendency to save energy shapes the phonology of Italian.

Similar patterns are observed in languages such as Brazilian Portuguese and Slovene (op.cit.: 3).

Counterarguments

Many linguists accept the view that ease of articulation plays some role in shaping sound patterns (for example Harris & Lindsey 2000, see also Hayes, Kirchner & Steriade 2004). There exists nevertheless a spectrum of opinion which might be described in terms of an emphasis on external causes on the one hand and internal causes on the other. The view of phonology associated with Jonathan Kaye falls firmly into the latter category. Kaye (1989, 1997) rejects the possibility that ease of articulation plays any role in the development of phonological systems, seeing them as expressions of a cognitive faculty that just happens to use the phonetic medium.

Kaye’s central arguments against EoA (he does not distinguish between the strong and weak forms described above) is that if it were true, i) languages would, over time, converge on an optimally easy state and ii) difficult patterns would never arise in the first place (Kaye 1989). These views are stridently reiterated by Stefan Ploch in his discussions of the phonetics-phonology interface (Ploch 2003, 2004).

In his defence of thesis i), Kaye imagines several possible rebuttals that would enable EoA to survive and demonstrates how each one of them is incorrect. Firstly, he
considers the possibility that different languages have different scales of difficulty. That is to say, convergence cannot not occur “because what is difficult or costly for one group is easier for another”. This is easy to demolish: the claim that different races have different phonetic aptitudes is absurd, as children whose first language is different from that traditionally associated with their ethnic group demonstrate. They are not phonologically impaired by the supposed mismatch. Secondly, there is the chance languages are slowly converging, and there simply hasn’t been enough time for this to become obvious. Again, easily dismissed: we know that languages change fairly rapidly. Modern Italian, French and Spanish developed from Latin in a period of less than 2000 years. If convergence were to occur, it would have done so by now.

Turning to thesis ii), Kaye provides evidence against articulatory explanations of changes in one language by showing that the reverse process has taken place in another. If Italian geminates arose in order to make sequences like /pt/ and /kt/ easier to pronounce, how is it that Moroccan Arabic exhibits processes that generate /kt/ clusters? And why, if word final devoicing in German is phonetically motivated, does Yiddish, which should be subject to the same phonetic constraints, undergo word-final voicing?

I hope to show firstly that the rebuttals Kaye imagines in his defence of thesis i) are the wrong ones (others being harder to discount) and secondly that thesis ii) is flawed since it fails to take account of semantic and other factors which act against ease of articulation.

Before we enter into a discussion of the reason convergence is not predicted by EoA there is one further point made by Kaye which warrants discussion. It concerns the effects of states in which the ability to expend energy is compromised. He asks “Are phonological processes [resulting from ease of articulation] more frequent when one is tired? When one become old?” and continues, in a footnote, “A physical state such as being drunk may have an effect on linguistic production. I am unaware of any systematic study of the phonology of inebriation. My feeling is that many manifestations of this condition do not correspond to phonological phenomena found in more normal situations.” (Kaye 1989: 48). EoA argues that certain widely observed phonological processes are the result of a need to conserve energy. Kaye’s point is that if being tired, old or drunk - states in which energy is depleted or motor control made more difficult - do not produce changes in speech similar to these processes, then EoA must be invalid. I have not come across any research in phonology which examines old age or fatigue. As for inebriation, there are in fact many studies, particularly from the field of forensic linguistics, on the phonological effects of alcohol consumption. Most of them are reviewed in Chin & Pisoni (1997). In one relevant piece of research, Lester & Kousen observed that subjects who had consumed alcohol exhibited final-obstruent devoicing and the lenition of affricates to fricatives (Chin & Pisoni 1997: 145), both commonly attested historical processes. Others observed “articulatory laxing or insufficiency” (op.cit.: 158) and in a 1991 study, Angelika Braun reported cases of consonant cluster reduction under the influence of alcohol (op.cit.: 224). It should be pointed out that these effects were accompanied by the ‘slurring’ or extension of segments in many cases, something we would not expect to see were minimum effort the only principle involved (alcohol induces a variety of physiological changes). However, in at least one study this could be attributed to the subject, “apparently aware of his condition…trying to compensate through inappropriate lengthening or aspirated stops” (op.cit.: 157).
Why no convergence?

Despite the accumulation in recent years of a great deal of evidence of their shared properties, an obvious fact about languages is that they differ. A brief look at Maddieson’s compendium of phonemic inventories leaves the reader in little doubt as to the variety of systems that serve the same purpose for the peoples of the world (Maddieson 1984). Some make use of well over 100 segments, whereas others get by with just 11. Some allow consonant clusters, some don’t. Against this background, is it really possible to say that there exists a measure of articulatory difficulty that is common to all languages? If so, what is it about speakers of Polish, for example, that allow them to produce such difficult clusters? If ease of articulation were the only variable involved, this would be a good question. It is, however, the structure of languages that differ, not the abilities of speakers. Articulatory difficulty is the same wherever you go; variation in language structure means the extent to which phonetic effects are liable to be felt is language specific.

What is meant by variation in language structure? Whilst a language might not quite be un système où tout se tient, it remains a system of sorts, a network of inter-relationships that straddle the different levels of analysis. The precise arrangement of these relationships depends on several factors. As André Martinet, who dedicated much of his career to elucidating the notion of language structure, explains, “the shape of a language depends on the communicative needs of the group that uses it. These needs may change according to the social, intellectual and economic evolution of the group” (Martinet 1960: 164). To this list might be added the history of the language system itself, the extent to which it has been influenced by contact with other systems and, finally, universals of substance, including ease of articulation and perceptual distinctiveness.

For Martinet, language structure comprises paradigms and syntagms. Any utterance, such as Iris wanted to be a writer, is a syntagm, or string of units. Phonological, morphological, semantic and syntactic rules determine the ways in which these units can be concatenated. If we turn to look at the units themselves we can imagine how they might be substituted for others with similar properties. Thus /b/ can be legitimately substituted for another member of the class [+obstruent], /s/, changing be to see. Another noun, politician, can likewise take the place of writer. These kinds of alternations are paradigmatic.

Martinet takes the view, since echoed by Lindblom and others, that competition between minimum effort and the need to communicate effectively govern the evolution of languages. He further makes the distinction between paradigmatic and syntagmatic economy. The former represents a tendency to minimise the total number of items that must be stored in the lexicon, the latter, broadly speaking, a tendency to minimise the number of units in a speech string. A change in the environment, such as the development of a new technology, usually necessitates a change in the language. The use of old words, combined in a new way, satisfies the requirements of paradigmatic economy, whereas syntagmatic economy would be best served by the invention of a single new word. The results of a cost-benefit analysis in the choice of either strategy depend on how frequently the new technology must be referred to. In this way, when internet communications were in their infancy and their use limited to a small section of the population, electronic mail sufficed to describe messages sent from one computer to another. The shorter coinage, email, took over when they became ubiquitous.

Syntagmatic economy is simply ease of articulation by another name. The same principle of minimum effort that results in forms such as [xaps] for ich habe es and
discourages us from expending the additional energy needed to articulate a voiced
geminate is at work here. Paradigmatic economy is a cognitive constraint which is
important for us only insofar as it represents another possible check on the influence
of articulatory effort in shaping sound patterns.

The crucial fact that limits the possibilities of convergence is that what will be
amenable to syntagmatic reduction depends on (at the word level) frequency and (at
the phonemic level) semantic contrastivity, both of which differ from one language to
the next.

**Enter semantics**

The number of contrasts entered into by a phoneme has been described as its
functional load\(^9\). An opposition with a high functional load serves to distinguish many
words. In English, the number of minimal pairs associated with /b/ and /p/ is large. In
John Higgins’ database of 70,000 words (Higgins 2005) there are 612 (for /ð/ and /θ/,
in contrast, there are just 8). Sound change can result in the loss or addition of phonemes, through merger or the phonologisation of one or more allophones. An
example of the loss of a phoneme is the merger in English of /ʌ/ and /ʍ/ which used
to distinguish words like *whether* and *weather*. The collapse of the /b-/p/ opposition,
were it to occur, would create many homophones. The functional pressures which act
to maintain /b/ and /p/ are, because of this, extremely strong.

The level of functional load, which is language dependent (if a similar survey of
minimal pairs were conducted for Icelandic, the results would be different), may serve
to block minimum effort effects if the outcome of these would be massive homophony\(^10\). I will call this the Functional Load Hypothesis (FL). The advantages of
avoiding homophony are clear. The potential for misunderstanding, and therefore the
failure of the communicative act, are greatly increased when a difference in meaning
is not reflected in a difference in sound. The requirement is not absolute, and a
significant level of homophony may be tolerated (it is important to note that a
calculation of the riskiness of homophony is not based on phonetic identity alone; the
likelihood of co-occurrence in the speech string is also important – thus *torque* and
talk are homophonous but unlikely to be confused).

Language specific functional load is a factor that Kaye fails to consider when he
adduces anti-phonetic changes in some languages as evidence against phonetic
explanations for changes in others. Take his example of the assimilation of /kt/ in
Italian and the emergence (through vowel syncope) of this same cluster in Moroccan
Arabic. My view is that /kt/ is more difficult than /tt/. A change in place over two
segments requires more effort than maintaining the same place for a similar amount of
time. Let us imagine for a moment that the functional load of /kt/ in Italian is low,
whereas in Moroccan Arabic it is rather higher. If this were the case, we would expect
ease of articulation to be more likely to win out in the former language but be blocked
in the latter. For the functional load of /kt/ versus /tt/ in Italian to have been high

---

\(^9\) The term used by Martinet is *rendement fonctionnel* (Martinet 1955: 54). Hockett has “functional load
(or yield or burden)” (Hockett 1967: 300).

\(^10\) Hockett’s paper notes that changes such as the neutralisation of /t/ and /d/ to /ɹ/ word internally in
American English were ‘allowed’ despite resulting in many homophones. He therefore goes on to
sketch a mathematical apparatus for quantifying functional load in order to say, for a given language,
exactly what level it would need to be to block change (Hockett: 1967).
before the assimilatory change, it would have had to distinguish a large number of words. In fact, the geminate /rt/ did not exist in earlier stages of the language. There were no minimal pairs of the type nacto (meaning A) and natto (meaning B). /kt/ was free to change to /tt/ without the prospect of increasing homophony. The question of why Moroccan Arabic has /kt/ can be divided in two: a) why is an anti-phonetic change generated in the first place and b) once there, why does it remain? Though the complex morphological constraints of Semitic languages undoubtedly bear on the development of Moroccan Arabic, the reasons why a hetero-organic cluster might emerge are the same for any language. Regular sound change can occur as a result of reanalysis. In addition, processes of lenition, assimilation or loss elsewhere in the system have knock-on effects. Either of these pathways could generate an anti-phonetic change. What we then have to discover is why natural phonetic processes which would massage the results of this change away over time are blocked from applying. Here again, Arabic is perhaps a special case. Homophony may not be the impediment to change, though functional load, conceived of in a slightly different way, must play a part: because of the unique status of the triconsonantal root in Semitic languages, the loss of /k/ from a word would rob it of membership of a whole class of words that share the same root.

Kaye’s second example concerns word-final devoicing in German and the reversal of this process in Yiddish. Devoicing could cease to apply in Yiddish for the same reasons sketched above to explain the emergence of /kt/ in Arabic. According to the version of FL I have sketched, it would be maintained if changes elsewhere in the system made it useful to increase the number of contrasts available to the language. These changes might include, for example, an increase in the number of words that end in consonants which are otherwise distinguished by voice alone. However, Yiddish and German are very closely related. We would have to ask what had happened to Yiddish to increase the possibility of homophonic clash so dramatically. If we suppose an anti-phonetic change can only be maintained because of the need for more contrasts we are left with the problem of why contrastivity has become reduced in one dialect of a language but not another.

However, the failure of an anti-phonetic change to be eliminated by the tendency to minimise effort could have several causes. A need to increase the contrastivity of the language is one. Another might be the attachment of important social markers to one or other form. If word-final devoicing were assigned low prestige by a particular group this might well be enough to override ease of articulation and support voicing. It is indeed quite possible to imagine the cultural separateness of speakers of Yiddish being reflected in their language in this way.

‘A Cognitive View’
The framework for understanding sound patterns Kaye puts forward as an alternative to phonetic accounts of the same phenomena is government phonology, hereafter GP\(^\text{12}\). GP is an internalist approach, and was inspired partly by developments in

\(^{11}\) One might suppose that all members of the class would change together and its integrity be maintained. However, since root consonants may occur in different environments, for example, separated by vowels (kitab ‘book’ versus maktaba ‘library’, both of the root k-t-b), the sound change could affect some words from the class and not others, resulting in a serious disequilibrium.

\(^{12}\) As with any theory, since its inception GP has been taken up by different scholars and developed in idiosyncratic ways. The sketch of GP I present here is intended to reflect most closely the views of
syntactic analysis, in particular Chomsky’s theory of government and binding (Chomsky 1981). Kaye’s GP takes the view that fundamental phonological operations and units are specified by the universal grammar. Phonological variety occurs because of different settings of certain binary parameters among languages. Thus, in terms of syllable structure, there are eight logical possibilities. Either onsets branch or they don’t (parameter I), either nuclei branch or they don’t (parameter II) and either rhymes branch or they don’t (parameter III). English is a language with all three parameters set to YES whilst Persian has NO for parameter I and YES for parameters II and III. Initial clusters such as /tr/ or /fl/ are therefore permitted in English but impossible in Persian.

GP analyses phonemes in terms of acoustically motivated unary features, or elements. These are [I], [U], [A], [H], [L] and [?] representing frontness, roundness, backness, a high tone, a low tone and occlusion respectively. The precise signification of some features changes according to whether they are used to describe vowels or consonants. In subsequent examples I will give the elemental expression after the standard phonemic transcription where appropriate.

Used alone or together the elements generate all possible segments. In this way the vowel /ü/ can be represented as [I.U], a combination of frontness and roundness. Phonotactic rules are conceived of in terms of governing relations between expressions which may be headed or headless. Segments that are permitted at the beginning of onset clusters are conceived of as headed and govern their neighbour to the right. Headed expressions (which contain one underlined element – the head) may not occur together. Thus /b/ [U.?.L] can precede /l/ [A.?] in an onset since it is headed and therefore a good governor. /b/ may not, however, form an onset cluster with /t/ [A.?.H] since they are both headed. In rhyme-onset clusters, the governing relation is from right to left and government can only occur if the governee is as complex or less so than its governor. Thus /ld/ [A.?][A.?.L] is a legitimate rhyme onset sequence but /nl/ [A.?.L][A.?] isn’t.

A GP analysis of the cross-linguistic bias against voiced obstruents does not readily present itself. There might have been room within the theory for an account of minimum effort based on the simplicity of expression. However, if we compare the GP expression for the voiceless stop /k/ [?.H] with its harder counterpart /g/ [?.L] we find that they are equally complex.

Descriptions of lenition, assimilation and vowel reduction from a GP perspective have been made by John Harris and Geoff Lindsey (Harris 1990,1994, Harris & Lindsey 2000). For reasons of space I can only give a brief summary of this research. Harris sees lenition in terms of the loss of elements from a syllabic position. The more elements a segment contains, the stronger it is. Governed positions such as the coda are more prone to weakening as a result of a ‘complexity condition’ that states they be only as complex or less so than their governor. According to the principle of ‘licensing inheritance’ some positions are said less able to support complex segments because they receive depleted ‘licensing potential’. In this way, Harris explains commonly attested asymmetries in the distribution of lenition-prone consonants.

Vowel reduction is analysed by Harris & Lindsey as the vocalic equivalent of lenition, that is to say, the loss of elemental content. An articulatory account is rejected on the basis that it cannot explain so called ‘centripetal’ reduction, in which

Jonathan Kaye, whose assumptions about phonology are summarised in *Phonology: A Cognitive View* (1989). As we will see, not everyone working within the framework has such an uncompromising view of the ontogeny of phonological rules.
the vowels of the unstressed inventory, rather than shrinking towards the centre, occupy the corners of the vowel space. Thus, in Russian and Tamil “canonical five-term systems reduce to a, i, u in unstressed positions” (Harris & Lindsey 2000: 190). It is true that, for the model of vowel reduction presented by Flemming and reproduced here, centripetal systems appear to pose a problem. We would expect /a/ to move towards the centre when unstressed, since F1 values are raised in segments of shorter duration. Flemming does note however, that it is the shorter duration of stress, not stress per se that conditions vowel reduction. A closer look at the structure of the languages in question is needed13.

Harris & Lindsey are less adamant than Kaye that phonological systems must have their origin in the universal grammar. They end their discussion of vowel patterns with generous open mindedness: “It might be that [the vocalic elements] are components of the universal grammar…alternatively, it might be that elements form part of some more generalised genetic endowment, or that they are posited de novo by learners through some interaction between external signals and general receptive capabilities” (Harris & Lindsey 2000: 203). Characterised in this way, the elements, conditions and principles of government phonology represent a straightforward attempt to formalise empirical observations – a valuable exercise. To go further, as Kaye does, and claim that this formalisation has some kind of existence independent of the facts, makes theory vulnerable to accusations of circularity. One might ask: why do we have these rules? - because they account for the data. But why are the data the way they are? The answer here seems to be – because they follow the rules.

In contrast, analyses based on external constraints are more explanatory because they offer a tangible motivation for phonological patterns. The reason the system is shaped a particular way is a result of its interaction with other (linguistic and extra-linguistic) systems. We have been focussing on the vocal organs which, unlike cognition, are accessible. Their characteristics (the position of the velic port, the manipulability of the tongue) are not inferred, but directly observable.

**Evidence from sign language**

Central to my argument in this paper is that form is not merely expressed through substance, but is partly shaped by it. If this is the case, it should be so for all linguistic systems, spoken or otherwise.

Sign languages, since they differ from spoken languages in just one respect – the medium in which they are expressed - can be used to test which aspects of language are universal, and which are modality-specific. The production of sign languages is effected not orally, but manually. The way they are perceived is not auditory but visual. One important question for sign language research is to demonstrate whether or not there are in sign language processes analogous to the phonology of spoken languages. The answer seems to be yes. Hand shape, place of the hand in relation to the body and the movement of the hand can be seen as analogous to choice of articulator, place of articulation and manner of articulation respectively (van der Hulst & Mills 1996). When a difference in just one of these features results in a new word we have a minimal pair, some of which, in Sign Language of the Netherlands (SLN), are illustrated here (7).

---

13 In addition, there seems to be a discrepancy between Flemming’s description of the reduced inventory of Russian and Harris’. The former has [i ə u] while the latter has a, i, u.
(7) Minimal pairs in SLN differentiated by handshape…

verklikken ‘to tell tales’  tandarts ‘dentist’  instituut ‘institute’

…and by place of the hand in relation to the body (from van der Hulst & Mills 1996).

onschuldig ‘not guilty’  wonen ‘to live’  leren ‘to learn’

So, the phonemic principle is not unique to spoken language; the notion of the syllable and morphophonemic alternation also seem to be applicable to sign language (van der Hulst & Mills 1996: 12). The problem presented by these kind of data for systems such as GP is that the fundamental features which some practitioners of the theory attribute to the universal grammar - the elements - are conceived of as auditory-perceptual units. The acquisition of language is made possible, or greatly eased, by the presence of these values in the brain of the infant learner. If the infant learner is deaf, and has no access to the stimuli according to which these features become interpretable, we might expect his or her linguistic development to be retarded. Yet this is not the case. As van der Hulst and Mills put it, “If sign languages are taken seriously, no modern textbook or article ought to state that there is a universal set of phonological features out of which specific languages make a choice – where this ‘universal set’ contains features such as [voice], [coronal], [nasal], and so on. If this set were the one made available by universal grammar, deaf children would be pretty much on their own when they have to figure out the abstract phonological structure of their language and the relation between this structure and its phonetic exponents” (van der Hulst & Mills 1996). The set of features provided by GP is too highly specified to be part of a posited universal grammar. Any ‘blueprint’ for phonology must not contain units that exclude sign languages, since they are demonstrably linguistic systems. The temptation to use universal grammar to explain as much of what is observed in the phonological patterns of spoken language as possible is increased, in
the case of Kaye’s Government Phonology, by its reluctance to concede the possibility that some of these patterns are the reflexes of external constraints.

There is also evidence that ease of articulation as well as ease of perception effects are present in sign languages. If sign language phonology is analogous to spoken language phonology, EoA predicts that constraints on the movement of the hands and arms influence the relative frequency of certain signals in the inventory of signs and sign alternations. Jean Ann developed a system for classifying handshapes as ‘easy’ or ‘difficult’ based on rigorous anatomical analysis (Ann 1996). She examined both American Sign Language (ASL) and Taiwanese Sign Language (TSL), predicting that the same hard handshapes would occur less often and the same easy handshapes more often in both languages. The handshapes considered were those in which all the fingers were closed except one, which was either extended, bent or curved. The results were in line with Ann’s prediction. In TSL all five hard handshapes occurred less often and four of the six easy handshapes occurred more often than average. In ASL all five hard handshapes occurred less often and three of the six easy handshapes occurred more often than average. Of the easy handshapes that were not found to occur more frequently, three (one in ASL, two in TSL) appeared less often than average and two (in ASL) were of about average frequency. Ann suggests that two of the three easy handshapes that occurred less often than predicted did so for perceptual reasons. They both involved the thumb in a curved articulation. Due to the physiology of the thumb, curvature here is not as readily perceived as curvature in the fingers. The risk of misunderstanding associated with this lack of clarity might explain why handshapes involving this articulation were avoided.

Is EoA a respectable theory?
Articulatory explanations for sound patterns have been called “either false or untestable” (Kaye 1997: 213). Ploch suggests that they are “always set up in a circular and thus unfalsifiable and non-empirical manner” (Ploch 2003: 74) and that, as non-empirical judgments they are simply “psychological and/or socio-political phenomena[al]” (op.cit.: 78).

For some, the doctrine of falsifiability sets the standard for theoretical investigation. Ploch reproduces the question Popper demands of every scientist – “‘Can I describe any possible results of observation or experiment which, if actually reached, would refute my theory?’” (op.cit.: 74). I would question whether any modern linguistic theory, particularly those of a cognitive bent, can pass the test of falsifiability so starkly framed14. Let us take a moment to consider, however, whether any aspect of EoA is capable of being tested and refuted.

As characterised by Kaye, EoA is not only falsifiable but clearly false. The non-convergence of languages convincingly refutes it. However, I have set out the reasons why a weak version of EoA does not predict convergence. Articulatory constraints compete with other pressures including ease of perception, semantics and social factors and don’t always win out. There is, however, a weakness in my argument: whenever ease of articulation effects are not observed, I can put their absence down to

14 It is also perhaps worth mentioning that Ploch’s point about the non-identity of causation and correlation - that the phonetic naturalness of many phonological processes does not mean that phonetics is the ‘cause’ of phonology (Ploch 2003: 80) - applies equally well to his alternative ‘exclusively cognitive’ explanation. GP, to which this refers, accounts for many linguistic facts, but it does not follow that their cause is cognitive.
the greater need, in that instance, of semantic contrastivity or social differentiation. Who can prove me wrong?

An important objective for proponents of EoA, if they want to meet the empiricist challenge, is, then, to find ways of quantifying the competing pressures – no easy task. FL might be amenable to this kind of investigation. Hockett developed an algorithm for computing functional load (Hockett: 1967) that might conceivably be applied, using corpora, to earlier stages of a language. On examining that language’s subsequent development, if segments with very low functional load were preserved in environments where phonetic reduction could be expected to occur, the hypothesis would fail. Similarly, one way of testing whether the risk of homophony is responsible for the maintenance of word-final voicing would be to examine languages that do not permit voiced segments in this position. They should have few homophones ending in voiceless consonants (fewer than would be created by the removal of the devoicing rule in a language that currently has it), since otherwise, this articulatorily motivated change would have been blocked. If the number of homophones was higher than expected, the hypothesis would fail.

Regardless of the extent to which EoA might or might not be testable, a closer consideration at this point of how far falsifiability should be accepted as the sine qua non of linguistic theory would be relevant here. Unfortunately an analysis of the weighty issues involved is beyond the scope of this paper. Suffice it to say that a fuller response to Ploch’s characterisation of the hypothesis as invalid on epistemological grounds is undoubtedly called for.

Summary and Conclusion
Throughout this paper I have made it plain that I see phonological systems as emerging from the complex interplay of different factors, each of which have their own demands to make on the shape of the speech string. I have in this case chosen to focus on the contribution of articulatory phonetics. Beyond specifying what sounds are physically possible (a pharyngeal tap, for example, is ruled out for anatomical reasons), constraints on articulation come in the form of a general principle of minimum effort, such that sounds requiring greater effort are more difficult, and therefore less favoured. Phonological reduction occurs whenever it has the opportunity within the system of constraints that allow effective communication. This is the versions of the ease of articulation hypothesis which I adhere to. In no sense do I claim that minimum effort is the only influence on sound patterns; just that it does play a part, and that the traces it leaves are observable.

Among these traces are an asymmetry in the distribution of obstruents: voiceless obstruents are more common than voiced ones because the maintenance of voicing during occlusion requires greater effort. Voiced consonants at back places are harder than those at front places of articulation due to the smaller volume of the oral cavity. To illustrate these asymmetries I provided examples from phoneme inventories and morphophonemic variation in different languages.

I used Björn Lindblom’s measurements of jaw extension to demonstrate a possible articulatory account of phonotactic rules. Whereas purely perceptual demands would be satisfied by an abrupt modulation between consonant and vowel segments, articulatory constraints impose a rule of smooth transition from occlusion to openness.

Lindblom was also partly responsible for the idea that vowel systems are shaped by the principle of minimum effort. I used an example from Edward Flemming’s development of this notion to show why the vocalic inventory of Italian is smaller in
unstressed syllables. I went on to consider assimilation and lenition, and gave phonetic accounts of these processes.

That articulatory constraints may be ‘phonologised’ is an eminently plausible idea: the vocal tract is the means by which we produce speech. It is subject to certain constraints. Phonological systems are an expression of all the constraints that act on the speech string be they phonetic, perceptual, semantic, morphosyntactic, social. Cross-linguistic evidence seems broadly to support this hypothesis. It is further strengthened by the observation that the phonology of sign language, which uses a different communication channel and is different though analogous to the phonology of spoken language, exhibits ease of articulation effects. Where speech is influenced by the physiology of the vocal apparatus, sign language is influenced by physiology of the hands and arms.

This view is by no means at odds with mainstream phonology. Research into the influence of phonetics on phonology is alive and well. In contrast, Kaye’s version of GP explicitly rules out the possibility that sound patterns are influenced by ease of articulation. In Phonology: A Cognitive View (1989), a touchstone work for GP, he lists the reason why he considers this stance justified. I have tried to show that the reasons he gives for rejecting EoA apply to a version of the argument that is so extreme as to be easily defeated. As a result of competition between the additional pressures of perceptual clarity, semantic contrastivity and even social markedness, convergence isn’t the inevitable result of a preference, where possible, for easier articulations. Kaye needs to show, why, in spite of this, EoA still deserves to be thrown out. Since his account does not admit the possibility that articulatory constraints help shape sound patterns, and that the patterns are to a certain extent simply formalised reflexes of these constraints, it is important to ask why, in the context of that framework, we have phonology at all.

Phonology is not a logical requirement of a communication system. As Kaye points out, artificial programming languages do not exhibit correlates of phonotactic rules or word-final devoicing (Kaye 1989: 11). His answer is that it helps us parse. Phonology is an evolutionary adaptation that speeds up communication with all the obvious attendant benefits for the organism. The evidence for the phonology-as-parsing hypothesis comes mainly from syntactically or morphologically conditioned alternations. One example is affrication before the third person pronoun in English but not before nouns beginning with the same sound, so that while what you want is commonly pronounced [wɒtjuːwɒnt], what Eunice wants would emerge as [wɒtnjuːniswɒnts]. Boundary phenomena such as stress, devoicing and harmony processes help us distinguish the end and beginnings of morphemes and thus aid in the rapid decoding of the signal.

Phonology-as-parsing is a compelling hypothesis and merits further investigation. I see no reason, however, to accept it as the sole motivation for phonology. Syntax and morphology play their part shaping sound patterns since they are essential to the organisation of speech. But they are far from being the only factors making demands on the ultimate form of the speech string. Semantics goes into the mix, together with ease of articulation, ease of perception, patterns inherited by the system (language history), social markedness and contact with other languages. This may sound messy,
but language is an enormously complex system, comparable perhaps to an ecosystem such as a tropical rainforest. The number of variables acting on such a system are dizzying. It may be extremely difficult to quantify their effects. We should, however, resist the temptation to create a neat repository for these variables in the form of a universal internal blueprint, particularly when there are possible external causes which have the advantage of being relatively accessible, unlike the conceptual structure of the brain.

References


