1

The structure and content of phonological primitives

MARIJN VAN ‘T VEER, BERT BOTMA, FLORIAN BREIT AND MARC VAN OOSTENDORP

1.1 Introduction

The primitives of phonological theory – whether we call them features, elements, gestures, or by some other name – represent the smallest units in phonology and stand in some relation to phonetic reality. At least this much most phonologists would agree on; but once we begin to ask further questions concerning the precise nature of the primitives, we find a significant divergence of views. What kind of primitives are there? How are they structured? And what is their phonetic exponence? This chapter presents an overview of our current understanding of the nature of phonological primitives, and of the key issues that play a role in the debate on this topic. We do this as a way of framing the contributions to this volume, which approach the topic of phonological primitives from different theoretical and experimental perspectives.

This chapter is divided into three parts. We first consider issues which relate to the structure of primitives. Are primitives merely attributes, or do they also have a value? (And, in the latter case, are these values restricted to binary valuations such as ‘plus’ and ‘minus’, or is the range of possible values more extensive?) How do primitives interact, and which implications does this have for the traditional units *phoneme* and *segment*? Should such units also be viewed as primitives (though at some higher level of organization), or are they no longer necessary? The second part of the chapter considers the issue of the phonetic content of features. Do phonological features relate to phonetics? And if so, which aspect of phonetics do they relate to – articulation, acoustics, both? What, if any, status do the phonetic correlates of features have within the phonology? Finally, the third part is concerned with the developmental origins of the primitives. Do primitives play any significant role in early linguistic development? Are they innate or acquired? To what degree, if at all, does the origin of the primitives make a difference to ‘mature’ phonological competence?

1.2 Formal structure

The discovery that the primitives of phonology (in the sense of the smallest building blocks in the phonological structure) are smaller than phones (or segments) is usually attributed to Roman Jakobson and Nikolaj Trubetzkoy (see esp. Trubetzkoy 1939; Jakobson 1941, 1949); for a recent overview of this period in the history of the field, see Dresher (2021). Some of the 19th-century students of phonology and phonetics, among them Sweet (1877), had noticed the central role of contrast in phonology, as is echoed in de Saussure’s (1916) adage ‘in the language system there are only differences’. However, at the time, these differences were seen as relations between phonemes, and not as things that were encoded in the phonemes themselves.

This changed with the insight that phonemes had a ‘phonological make-up’ (Jakobson) or ‘phonological content’ (Trubetzkoy), which, in Trubetzkoy’s view, involved a distinction between privative and equipollent features. The first formalisation of this insight was undertaken by Jakobson in the 1940s, but it did not make use of privative and equipollent feature contrasts. Instead, Jakobson’s work eventually led to the now mainstream format of binary features. These features consisted of two elements: an attribute and a value. For example, the feature [+voice] was said to have the attribute *voice* and the value +, with the square brackets signalling that the two constitute a formal entity, the feature. This conception of features was also taken up in Chomsky and Halle’s (1968) *The Sound Pattern of English* (SPE), a publication that was so influential that to this day binary features continue to be part of the formal *lingua franca* of phonology.

It should be noted that SPE did in fact distinguish between different kinds of features, since Chomsky and Halle assumed that features had a slightly different structure in the lexicon than in the phonology, and yet another structure on the phonetic surface. In the lexicon, features could have the values *u* and *m* (unmarked and marked), in addition to + and −. The idea was that the rules of the phonological component would first change these values into + and −, depending on the context. For example, an [*u* round] specification of a vowel would change to [+round] in the context of [+back], and to [−round] in the context of [−back]). So, although features had four underlying values in theory, in practice this addition gave the system very little extra expressive power (see Hall and Mielke 2011).

On the phonetic level, SPE assumed that feature values were scalar. For instance, the slight nasalisation of a vowel preceding a nasal consonant in English was represented by assigning a (small) numerical value to the [+nasal] feature of the vowel (say, [2 nasal]). It was assumed that these numerical values were integers, although this seems to have been mostly a matter of convenience. A more logical choice may have been to assume that these values are all real numbers between, say, 0 (= [−]) and 1 (= [+]). But as far as we are aware, the implications of the SPE view regarding the use of scalar features have not really been explored. In later work, the use of scalar features has mostly been restricted to two domains: stress (where they distinguished between different levels of stress) and sonority (where they distinguished between different levels in the sonority hierarchy); we will return to this briefly in Section 1.2.2.

Below, we discuss a number of issues relating to the formal structure of primitives. Section 1.2.1 examines the question of whether primitives should be viewed as privative or equipollent, and also considers how, in some approaches, an enriched phonological structure has led to a reduction of the number of primitives. Section 1.2.2 discusses the status of multi-valued (or scalar) features and looks at some of the ways in which phonological approaches have formalised apparently scalar properties such as vowel height, stress, and sonority. Finally, Section 1.2.3 briefly considers the question of whether segments themselves must also be regarded as primitives.

1.2.1 Privativity vs equipollence

One of the main topics in the discussion on the formal structure of features revolves around Trubetzkoy's (1939) distinction between privative and equipollent contrasts. Formally speaking, the issue concerns the question of whether (all) features are attribute-value pairs, or whether features are attributes only, so that the presence of a feature can contrast only with its absence (see e.g. Kim 2002).

A general advantage of privative approaches is that they provide a straightforward way to assess the relative complexity of segments in a language. For example, in a language with a voicing contrast in stops, the voiced stop series will have one primitive more (e.g. the monovalent feature [voice]) than the corresponding voiceless series. This appropriately reflects the fact that the voiced series will behave as marked; the voicing contrast may be neutralized in final position, for example. A monovalent feature [voice] also provides an explanation for the fact that, in segmental inventories, the presence of voiced stops generally implies that of voiceless ones (but not vice-versa). In binary-valued approaches, such cases of relative markedness require additional theoretical machinery, such as underspecification and redundancy rules.

However, a methodological challenge is that it is difficult to compare a privative and an equipollent approach as long as we can also change other aspects of the model. A typical argument in favour of equipollence involves the phonological status of nasality: the absence of nasality is arguably never phonologically active, i.e. it does not spread, no other features refer to it, etc. Therefore, a model which assigns equal status to [+nasal] and [−nasal] is less restrictive than a one which assumes a (monovalent) feature [nasal] only. However, this argumentation holds true only if there is a principled reason against a monovalent feature [oral]. Indeed, some oppositions which appear to be binary phonetically are sometimes translated into two distinct privative features. For example, laryngeal contrasts involving glottal width are sometimes formalised in terms of two monovalent features, viz. [voice] and [spread glottis], sometimes in the same language (an approach that is referred to as 'laryngeal realism'; see among others Iverson and Salmons 1995; Honeybone 2005). Note that this kind of approach leads to the theoretical possibility of a four-way contrast (because a segment in a language could be hypothesized to have neither feature, or even both).

Privativity therefore forces the analyst to assume a model which contains as small a set of primitives as is possible. We can see this in the development of Element Theory (Kaye, Lowenstamm and Vergnaud 1985; Harris and Lindsey 1995; Backley 2011), an offshoot of the framework of Dependency Phonology that is associated mainly with work in Government Phonology but is in principle independent from this approach. An explicit goal of Element Theory is to reduce the number of elements (which are formally equivalent to privative features, in that they consist of an attribute only). The textbook account of Element Theory in Backley (2011) assumes a total of six elements: the 'resonance' elements |A|, |I|, |U|, which are familiar from earlier dependency-based approaches (e.g. Anderson and Jones 1974; Anderson and Ewen 1987), and the 'manner elements' |ʔ|, |H|, |L|, which have their origin in the theory of 'charm' (Kaye et al. 1985). Various attempts have been made to reduce this 'standard' set, by replacing one or more primitives by an enriched phonological structure (see e.g. Jensen 1994, Pöchtrager 2006, 2013).

Much the same development can be observed in feature-based approaches that were modelled on SPE. For example, the introduction of syllable structure into generative phonology in the 1970s and 1980s obviated the need for the feature [±syllabic]. Another dimension in which the majority of scholars have accepted the replacement of a feature by structure is segmental length, which is nowadays usually seen as involving the attachment of a segment to more than one skeletal slot, root node, or mora, depending on one's representational assumptions. (In mora theory, the difference between a short and a long consonant is usually represented in terms of association to zero or to one mora; this means that, for consonants at least, we can interpret the mora as a kind of monovalent feature.) The main argument for a structural approach to length is that long segments behave as bisegmental. For example, in some languages long vowels can occur in open syllables only. In other languages, vowels undergo (compensatory) lengthening when a coda consonant is deleted. In an analysis in which length is represented by a feature, there is no logical relation between the length of a vowel and its distribution in open and closed syllables. On the other hand, there are also languages in which tense vowels occur in open syllables only. In some of these (for instance, Eastern Javanese), tenseness also shows harmonic behaviour, which suggests that it functions as an (autosegmental) feature.

The idea that features can be replaced by structure has been extended by some authors into other domains as well. For instance, in the model of Aperture Theory (Steriade 1993ab, 1994) there is no need for the feature [continuant], because the slots to which features are associated are specified in terms of their aperture properties. Golston and van der Hulst (2000) argue that 'stricture is structure', while, as noted, Pöchtrager (2006, 2013) reanalyses lowness in terms of length (in the sense of structural complexity). Finally, Botma and van Oostendorp (2012) observe that the relation between tenseness and syllable structure in Germanic languages suggests an analysis in terms of the Trubetzkoyan 'Silbenschnitt';[1] if some aspects of syllable structure are underlying (and distinctive), tenseness may be seen as a predictable phonetic interpretation of syllable structure, rather than as a feature. None of these proposals can be regarded as mainstream, but all of them have attracted at least some attention.

Finally, notice that in the SPE system the features [syllabic] and [long] had a problematic status in any case, since neither appears to have clearly definable articulatory correlates (see also the discussion in Section 1.3 below).

Another motivation to reduce the set of primitives concerns the relation between the number of primitives and the size of segment inventories. The size of an inventory grows exponentially with the size of the set of primitives. One primitive (a feature or an element) x gives a maximal size of 2 segments ([+x] and [−x], or [x] and ∅); two primitives give a size of four segments ([+x, +y], [+x, −y], [−x, +y], [−x, −y]), three primitives eight, four sixteen, and so on. With only ten primitives we reach an inventory size of 1024. Very few theories have been successful at curtailing the number of primitives in such a way that they fit better with the size of segment inventories. The largest known inventory sizes are found in Khoisan languages, some of which

[1] *Silbenschnitt* 'syllable cut' divides the post-nuclear transition into *smooth* and *abrupt* cuts. This opposition is then employed to explain the well-established correlation in many Germanic languages between a vowel's length and tenseness and the following consonant's strength and fortisity. Long tense vowels followed by short lenis consonants are said to implement a smooth cut, while short lax vowels followed by a (long) fortis consonant are said to implement an abrubt cut. The silbenschnitt proposal thus seeks to explain these surface-correlations as the manifestation of a single underlying dichotomy between these two types of post-nuclear transition.

have around 120 segments (of which the majority are clicks). Such an inventory could in principle be described with as few as 7 mono- or bivalent primitives,[2] in Element Theory with an additional headedness distinction for each segment 5 to 6 primitives (giving 112 and 256 possible segments) would suffice. Similarly, a bivalent system in which underspecification is construed as contrastive would need at most 5 features (243 possibilities). In reality, however, most theories employ a number vastly exceeding what is necessary on a purely combinatorial basis, and arguments have been made either way as to whether that is desirable or not. For further discussion of these issues see for instance Reiss (2012), Breit (2013), and Bafile (2019).

The question of why segment inventories are in general restricted has not, to our knowledge, received a satisfactory answer. (It may be due to a combination of factors). It is unclear why there are not more languages with large consonant inventories (and more languages with clicks, which can be easily combined with other contrasts). The fact that Khoisan languages have large inventories has been attributed to their being spoken in the area where *homo sapiens* originated (e.g. Henn et al. 2011; Fleming 2017); but as far as we are aware, there is no real explanation for why very large inventories are so rarely found outside of Southern Africa. From a functional perspective, inventories that are even larger than those in Khoisan (consisting of, say, two or three hundred sounds) might be undesirable from the viewpoint of contrast, but this could be mitigated by the possibility of having very short words.

Summarising, the advantage of an equipollent approach seems to lie mainly in that it allows a straightforward formalisation of the notion of markedness. While in a binary approach the two ends of a pole are formally equally complex, an equipollent approach treats one of the ends as relatively more complex than the other. The difference is therefore mostly one of notation; but such notational issues are important (Ganesalingham 2013). We have also seen that feature reduction is an important driving force behind privative approaches; many approaches which assume a restricted set of primitives assign a correspondingly greater role to phonological structure in accounting for phonological generalizations.

1.2.2 Multivalued scalar features

As we already noted, features have also been treated as being multivalued. For example, Ladefoged (1971) identifies two contrastive degrees of nasalization in Palantla Chinantec (subsequently confirmed by Merrifield and Edmondson 1999), while Ladefoged and Maddieson (1996) suggest that the two types of nasals found in Acehnese can be distinguished in terms of the degree of nasal airflow. Such contrasts can be handled by a scalar feature [nasal] (with the values [0 nasal], [1 nasal], [2 nasal]), although this is not the only possible analysis. Ladefoged himself speculates that the different degrees of nasalization in Palantla Chinantec may result from different underlying representations (Merrifield and Edmondson make a similar suggestion), while the Acehnese contrast has been analysed in terms of underspecification (Grijzenhout 2001) and also in terms of a distinction between obstruent and sonorant nasals (Durvasula 2009).

Another phonetic dimension which has been interpreted in scalar terms is vowel height. Ladefoged (1975) assumes a multivalued scalar feature [high] to represent a four-way height

[2] The two are equivalent in this respect, since in a monovalent privative system, for each primitive there is the option of inclusion and exclusion, while in a bivalent equipollent system each primitive can be annotated with either of two values, i.e. in both cases there are 2^n possible combinations.

contrast like /æ ɛ e i/, such that /æ/ is [1 high], /ɛ/ is [2 high], and so on (see also Williamson 1977). An approach with binary [high] and [low] cannot represent such a system, since one of the four possible combinations (i.e. [+high, +low]) is impossible. A different type of scalar approach to vowel height is found in Particle Theory (Schane 1984), where the place features |i|, |u|, and |a| (as used in Dependency Phonology) can occur more than once in the specification of a segment. Thus, on the assumption that |i| is a property of front vowels and |a| of low vowels, a four-way height contrast among front vowels (e.g. /i e ɛ æ/) is represented as |i|, |ia|, |iaa|, |iaaa|, respectively.

The use of multivalued scalar features has not been widely accepted — at least, not as a way of encoding underlying contrasts. A general reason for this is that such features are unconstrained; there does not appear to be a principled upper limit on the number of ways that a particular phonetic dimension can be categorized. Aside from this, we suspect that many analysts who assume a forward feeding model of the kind set out in SPE (in which categorical phonological representations serve as the input to gradient phonetic representations) prefer to view multivalued features as part of the phonetic implementation, rather than of phonology 'proper'. For this reason, analyses of vowel height usually involve additional features (e.g. [ATR]) or additional formal mechanisms. For example, the 'standard' model of Dependency Phonology (Anderson and Ewen 1987) represents the contrast between high-mid /e/ and low-mid /ɛ/ in terms of headedness; the two vowels have the same primitives, viz. |i| and |a|, but they differ in which of these primitives is the 'head' of the structure. In /e/, headedness of |i| implies that here 'frontness' is more prominent, while in /ɛ/, headedness of |a| implies that 'lowness' is more prominent. For a similar approach, see e.g. van der Hulst (1988) and Backley (2011).

The formally most rigorous interpretation of headedness is that it is an asymmetric relation between two sister nodes, such that one is the head and the other the dependent. This is the conception of headedness in Radical CV Phonology (van der Hulst 2020). Other approaches apply the notion of headedness less rigorously. For example, the model of Dependency Phonology in Anderson and Ewen assumes different kinds of dependency relations (including a relation of 'mutual dependency'), while in many versions of Element Theory headedness is viewed as an intrinsic property of elements, rather than a structural relation between them.[3] For example, Backley (2011) assumes that elements which occur on their own can also be headed (in his analysis of English vowels, this distinguishes short monophthongs, which are headed, from reduced vowels, which are unheaded). This interpretation of headedness is conceptually quite similar to a multi-valued approach, in that both provide a way to represent the degree of prominence of a particular phonetic dimension.

We already noted that in the period following SPE, scalar features have by and large been restricted to two domains: to distinguish between different levels of stress, and to account for the sonority hierarchy. With respect to stress, Chomsky and Halle (1968) assumed that at any cycle of the phonological derivation in English, one vowel would receive a value [1 stress] for primary stress while other vowels could receive a larger value (for secondary stress, tertiary stress, etc.) In a new cycle, the rules would pick out one [1 stress] to stay at the highest level, while all other stresses would be demoted by one. Later work made the intuitively more

[3] Though note that for instance Breit (2013, 2017, in press) argues explicitly for a structural conception of headedness in Element Theory, a view also taken up by Ulfsbjorninn (2021).

appealing assumption that primary stress has the highest value. This is the main insight behind grid-based models of stress, in which stress is represented by columns of grid marks. A popular alternative to represent stress is in terms of a hierarchical tree structure of some sort, in which the primarily stressed syllable is the head (of the head of the head) of the overall structure. This option makes it possible to represent a potentially infinite number of stress levels without recourse to natural numbers. In this respect, we see a clear parallel between the introduction of headedness in subsegmental structure and in prosodic structure; both allow the representation of relative prominence without the need to assume multi-valued features.

As regards sonority, there is a long tradition, dating back to Jespersen (1904) to represent sonority classes in terms of a grid-like structure or in terms of sonority values (see e.g. Clements 1990; Zec 1994). Such values have also been invoked to account for sonority distance effects. For example, although the clusters in /kwa/, /kla/, and /kna/ all have the appropriate sonority profile for onsets, a language like (Modern) English only tolerates sequences in which a voiceless stop is followed by an approximant. The ungrammaticality of /kn/ onsets has been attributed to the fact that voiceless stops and nasals are too close to each other on the sonority scale. Sonority is therefore a scalar property, standardly treated in scalar interpretations of certain aspects of phonological structure.

While we do not know of any approach to phonological primitives that uses scalar features *tout court*, we have seen that there are approaches which use scalar features in certain specific domains, or employ formal mechanisms that are conceptually similar to them. This is also the case for some of the contributions to this volume; we find different instantiations of head-dependency relations in BARONI (this volume) and SCHWARTZ (this volume), while PUGGAARD-RODE, BOTMA AND GRIJZENHOUT (this volume) provide a scalar (or, in their terms, 'quantised') interpretation of subsegmental timing positions.

1.2.3 Segments as primitives?

While there is currently general consensus that the smallest units in the phonological organization are subsegmental 'primes' (be they features, elements, or gestures), the status of segments is a matter of some debate: do they have an independent status in the theory or are they merely configurations of features? The main reason for recognising subsegmental primitives was, of course, that many phonological generalizations that were traditionally made in terms of segments can be more appropriately stated in terms of features. Consider in this respect also the problem of crazy rules.

In the influential model of Autosegmental Phonology (Goldsmith 1976), the strict one-to-one relationship between features and segments as assumed in SPE was abandoned. As a result of this, the role of segments became less prominent. The possibility for a feature to be simultaneously linked to more than one segment (e.g. a place feature in a homorganic cluster) meant that segments no longer functioned as primitive units. (One might say that this brought phonological representations a small step closer to 'phonetic reality'.) At the same time, something like the traditional segment can still be derived from an autosegmental model as an ×-slot (or a root node) with all features that are associated to it. It is therefore still possible in principle to extract an IPA notation directly from an autosegmental structure – one would just have to ignore the floating features.

The role of segments is downplayed even more in some current models, to the extent that segments lack a formal status. For example, in the framework of Articulatory Phonology the grammar regulates the timing of articulatory gestures (see e.g. Browman and Goldstein 1986, 1989; Pouplier 2020; see also SORENSEN AND GAFOS, this volume). Coordination of different gestures may create the illusion of a segment (as when a labial gesture, a velic lowering gesture, and a voicing gesture together yield [m]), but the segments are themselves epiphenomenal.

Segments also lack a formal status in approaches where phonological acquisition involves the activation of both features and feature co-occurrence constraints (see VAN ‘T VEER, this volume). In such an approach, the incremental expansion of feature combinations in the grammar will at some point give rise to segment-sized units, but, as VAN ‘T VEER notes, these are epiphenomenal; the phonological contrasts in the emerging grammar are expressed by features, as are the phonological processes and phonotactic constraints.

Finally, some recent approaches explicitly reject segments, at least as cognitive units. This is the case in some versions of Exemplar Theory (e.g. Bybee 2001) and in the work of Silverman (e.g. Silverman 2017). It should be noted, though, that these approaches are in general highly sceptical of phonological structure, including the types of primitives that we are discussing in this book. For example, Silverman (2017: 283) maintains that “there is no reason to assume that language users subdivide the words they learn into distinct sound components unless there is evidence from alternation to do so.” He notes that such sound components may be useful for the purpose of linguistic analysis, but that such analyses are not an adequate reflection of the phonological knowledge of language users.

While many of the traditional functions of segments (or phonemes) have been taken over by subsegmental units, there seem to be at least some cases for which it is unclear whether a purely non-segmental account is sufficient. For example, sonority (to the extent that this is considered a grammatical property; see e.g. Ohala 1990 and Harris 2006 for a different view) appears to be special in that it is one of the few domains in which segments, rather than features, appear to act as primitives. There does not seem to be a generally accepted way to represent the sonority hierarchy in terms of features. (Note in this regard that the sonority scale in Parker (2011) contains no fewer than 17 levels.) The most common approach, at least for some of these levels, is to place whole segments on the scale, so that e.g. [r] takes up a different position on the scale than [l]. We can also count vowel height as an aspect of the sonority hierarchy. Such holistic approaches to segments therefore seem to still be the major challenge to any idea that segments have an internal structure.

1.3 The content of phonological primitives

While we have been principally concerned with the question of how phonological primitives are employed by phonological computations in the previous section, another important and long-standing question that has more recently attracted renewed attention concerns the very nature of the primitives themselves. This question can be usefully broken down into two parts. First, we may ask what it is that the phonological primitives encode mentally: is each primitive closely tied in some way to a specific phonetic cue, are the cues that they correlate with purely articulatory, purely acoustic, or a mixture of both, and do their eventual phonetic identities play any role in their phonological distribution and behaviour? Second, what is the origin of phonological primitives: are they innately pre-specified in some way, or do they emerge from the speaker/hearer’s input during acquisition?

If one imagines for a moment a representative sample of the input that phonology students across the world's linguistics departments receive — particularly in introductory courses, but in most cases equally in those at a more advanced level — then it might well appear as though the primitives that phonology operates on are almost invariably equipollent, temporally unorganised, atomic mental representations of the actions of the active articulators during speech production, such as [±high] for vertical tongue body elevation or [±coronal] for movement of the tongue blade. The input sample may contain some features which reference acoustic properties, e.g. Jakobson, Fant and Halle's (1952) [±grave] or the |L| of Element Theory (cf. e.g. Backley 2011), and some which do not appear to refer to any direct physical substrate, such as the [±FB] and [±WB] of boundaries (cf. Chomsky and Halle 1968:66f). At best, however, such occurrences may be barely distinguishable from the noise, leaving us with the impression that any debate over the nature of the phonological primitives is all but settled. Conversely, both the recent history of the field, as well as the current literature on the topic give quite a different impression. There is now a great plurality of positions in the field, and ongoing work has both revealed new questions and made the finding of answers to long-standing questions more pressing than ever. This latter impression is also one that we see readily confirmed by a sample taken not from the world's phonology classrooms but from the contributions to this volume, which are of course tied together by a shared interest specifically in the subsegmental aspects of phonology.

1.3.1 Representationalist concerns: Early developments

While the move toward cognitivism has by virtue of its subscription to indirect representationalism centre-staged the question what a given mental representation is a representation of, as long as they construe phonology as in some sense different from articulatory and auditory phonetics, even those pre- and post-cognitivists who may subscribe to direct realism have to answer the question of what object in the real world it is that the phonological primitives correlate with.

Indeed, the question is a long-standing one. Even if not phrased in terms of the now common computationalist conception of inter-modular representational mappings, the question was readily apparent in the structuralists' concern for the delineation of the distinctive properties of phonemes. To wit Trubetzkoy's (1939) discussion of the functional valence of the encoded oppositions, which explicitly acknowledges the possibility of a mismatch across levels of analysis, and crucially presupposes that what we understand to be the referents of the distinctive features plays a significant role in characterising their phonological identity.

In terms of modern feature theories, Jakobson, Fant and Halle's (1952) Preliminaries most prominently represents a turn toward explicitly addressing the question of how the distinctive features that form the primitives of phonology correlate to the external(-ised) speech signal. Jakobson, Fant and Halle strongly argue for a primacy of acoustic factors over the articulatory, based on the observation that distinct articulatory configurations may lead to the same perceptual outcomes, so that a system built on primarily articulatory distinctions may end up being richer than what is perceptually recoverable. While this concern is reflected in their extensive discussion of both acoustic and articulatory aspects of the 12-feature system they propose, they end up with a system in which at least some distinctive features seem to remain best characterised in articulatory terms — something perhaps foreshadowed in their early remark that the "systematic exploration of the [perceptual and aural] levels belongs to the future and is an urgent duty" (p. 12). Jakobson and Halle's (1956) Fundamentals of Language continues this line of inquiry, reaffirming on the one hand the acoustics-directed path of their earlier work, but also explicitly addressing the importance of the mentalistic, representational aspect of the question and potential challenges to it.

While the acoustic-leaning conception of phonological primitives carries over to Halle's (1959) Sound Pattern of Russian, we see a significant deviation and turn toward an articulatory definition with the advent of SPE (Chomsky and Halle 1968). Though it is important to note that while the nomenclature, discussion, and illustration of the mapping of distinctive features to their phonetics in SPE is heavily skewed toward articulatory properties, Chomsky and Halle explicitly acknowledge that this approach does not necessarily imply the representational irrelevance of perceptual-acoustic factors. Rather, they propose that phonological features can be interpreted both as "a set of instructions to the articulatory system, or as a refined perceptual representation" (p. 65). The key to this apparent ambivalence is the generativist conception that despite what may or may not be transmitted in the physical speech signal, a listener will actively try to match and recover what they perceive to the underlying structure, a view that is in its essence congruent with recent experimental evidence that hearers engage in active articulatory motor simulation during perception (Berent et al. 2020). In doing so, they argue, the hypothesis space a listener must entertain is sufficiently constrained to facilitate the recovery of articulatory events not directly represented in the signal (p. 294). In defining and exploring their set of distinctive features, it may seem sufficient then to discuss only the articulatory aspects, and Chomsky and Halle (1968) limit themselves to this "not because we regard [acoustical and perceptual] aspects as either less interesting or less important, but rather because such discussions would make this section [...] much too long." (p. 299).

Given the great influence of SPE on generative phonology and the ensuing success in furnishing explanatory hypotheses about phonological behaviour, it is perhaps no surprise that in many cases the explanatory adequacy of representational correlates of a particular set of features was seen as less significant than its explanatory power on a higher level. After all, for many a phonologist pre- and post-SPE, the principal phonological function of the primitives is classificatory, and the thesis of representational recoverability may appear to obviate the need for anything more than a descriptively adequate set of primitives regardless of its label or precise external correlate — in other words, if a set of features seems to sufficiently explain natural class behaviour, one may assume this sufficiently justifies the set of features itself. We stipulate that this is precisely the effect we observe in our imagined sample of phonology students' classroom input above.

1.3.2 Articulatory vs acoustic correlates: Current developments

If the situation described above were all there is to explain with respect to the nature of phonological primitives and their phonology-external correlates, we might be tempted to conclude that any deeper questions about these primitives are relatively unimportant to phonological theory itself. In fact, however, a stimulating diversity of different positions has over time (re)emerged from this, which variously accord different levels of importance to the correlates of phonological primitives. In many cases, these developments have also re-emphasised that what external entity we believe the primitives represent internally will matter for phonology more widely, as we may end up with different conclusions about many aspects of the phonological features, such as their granularity, interrelation, valency, and so on.

At one end of the spectrum, Articulatory Phonology (Browman and Goldstein 1986, 1988, 1989; et alios) has taken on board many issues brought to light over the years regarding the phonetic adequacy of an SPE-type mapping from sequential categorical feature complexes to phonetically rich, temporally ordered representation and sought to much more radically embrace a close link between phonology and articulatory phonetics, conceptualising the phonological primitives as direct mappings of the articulatory gestures involved in speech planning, otherwise usually relegated to a much lower level of cognition such as a phonetic component that is much better characterised as a transducer than an input system in Fodor's sense (cf. e.g. Lenneberg 1967:75-120). This goes so far as to include the phonological encoding of the relative temporal expression of such gestures with respect to one another, as opposed to the standard assumption of temporal simplicity under close juncture (i.e. within segments) shared by most other phonological frameworks. As SORENSEN AND GAFOS (this volume) point out, the gestural direct-mapping conception of phonological primitives forces us to treat phonology not purely as an abstract symbol manipulation device but as a dynamical system, i.e. a system where state is time-dependent in addition to its 'static' (i.e. atemporal) context. As their contribution shows, this assumption which is at the core about the primitives more than the phonological system fundamentally shapes both what we assume about the wider organisational principles of phonological structure and the phonological computations that apply to them.

An interesting converse trajectory can be observed in the development of an alternative approach to phonological primitives rooted in Dependency Phonology (Anderson and Ewen 1987) and Government Phonology (Kaye, Lowenstamm and Vergnaud 1990; Charette 1991; Harris 1994), now mainly represented by Element Theory (Kaye et al. 1985; Kaye and Harris 1990; Harris and Lindsey 1993, 1995; Backley 2011). Early in the development of Dependency Phonology, Anderson and Jones (1974) argued that significant weight should be given to the observation that phonology appears to often treat its representations at a less finely granular level than what would be suggested by a gestural or even an SPE-style distinctive feature analysis. Although not extensively addressing the correlation issue, they propose that the primitives of phonology may be more akin to prototypical versions of certain vowels or consonants, e.g. |a, i, u|, which either autonomously or as combined components phonologically represent the primitive units of segments. Similar arguments were also set forth by Schane (1984) and Kaye et al. (1985), both of whom sought to find a set of primitives which they argued more accurately characterised the level of granularity that really matters at the level of phonological representation as opposed to the more fine-grained phonetic details that didn't appear to be of such phonological import. While there is some ambiguity on the precise nature of these prime's phonology-external correlates in earlier work, at least Kaye et al. (1985) drew a very definitive link to the articulation-centric system established with SPE by developing on the one hand an Element Calculus which derived a set of binary distinctive features from the less granular, autonomous elements such as |A, I, U| (again autonomously representative of prototypical vowels such as /a, i, u/), and on the other as part of that calculus drew a distinction between two types of functions in the phonetic exponence of such features, namely that elements in a structural head impart what is mostly reflective of the perceptually salient vocalic properties, while the operators throughout supply more of the articulatory base. While Kaye (1989:42-49) quite strongly rejects the hypothesis that phonology is articulation-based, the quite radical and fundamental departure from the articulatory link that is commonly associated with Element Theory today came to fruition principally through the proposals of Lindsey and

Harris (1990) and Harris and Lindsey (1993, 1995), who explicitly argued for a purely acoustic conception of the elements' phonetic exponence much in line with Source-Filter Theory (Fant 1960; Stevens 1998) and Traunmüller's (1994, 1998) Modulation Theory. Under this view, the phonological primitives represent acoustic modulations upon a schwa-like carrier signal, such as the suppression of amplitude (|ʔ|) or periodicity (|h|). Both the phonology-focused reductionism and the acoustic primacy that underlie this view continue to be fruitful avenues of investigation with clear consequences for the phonological system. For instance, Backley (2011, 2017) proposes that the acoustic identity of the primes embodies antagonistic phonological relations in between the primes while BARONI (this volume) follows a long-standing issue in this theory where some elements appear to be more commonly affiliated with consonantal properties and others with vocalic properties by proposing that even within the group of the three prototypically vocalic elements |A, I, U| some elements might be more affine to the encoding of vocalic contrast than others. In terms of the coarse granularity of elements and the way these combine into melodic representations, PRINCE (this volume) shows how the resultant measure of combinatorial complexity of segments (cf. e.g. Kaye and Harris 1990) can be drawn on to explain patterns of acquisition and loss in French children and aphasics a la Jakobson (1941).

While Articulatory Phonology and Element Theory as characterised above represent relatively puristic positions on the articulation vs acoustics debate, as we have already seen of course the mainstream of current phonological theorising operates on the articulation-leaning but nonetheless relatively intermediate position rooted in SPE. Taking as an example Hayes' (2009) textbook, and the explanations of the distinctive features supplied there, we find that of the 31 distinctive features employed, 3 (syllabic, stress, main) are given neither a clear articulatory nor acoustic correlate, 4 (consonantal, approximant, delayed release, tense) are given a purely acoustic correlate, a further 4 are given both an articulatory and an acoustic correlate (sonorant, long, strident, voice), and the remaining 20 (continuant, trill, tap, front, back, high, low, round, ATR, nasal, labial, coronal, dorsal, anterior, distributed, lateral, labiodental, spread glottis, constricted glottis, implosive) are given a purely articulatory correlate.[4] While clearly skewed toward articulatory primacy, there appears to be a clear place for an acoustic base in this conception of the phonological primitives. An important question for such a view, though as far as we know it has unfortunately not been addressed to any great extent by its exponents, is how one might explain which features of articulation and which features of acoustics ought to receive a mental representation, and of course especially in those cases where both types of correlates appear to be valid, whether they contribute equally and whether there is a degree of language-specificity in which factor is accorded a more prominent status.

There are of course also theoretical positions to which such correlative metrics do not apply in the exact same manner as in the views discussed above, because their wider psychological conception of phonology does not attribute the same status to whatever may be considered their phonological primitives. For instance, in Exemplar Theory, phonological forms are constituted by clouds of exemplars, which themselves store specific experienced instances of the form and

[4] Although some of these, e.g. [high, low, front, back, nasal] could quite easily also be given an acoustic correlate, these features are not discussed in acoustic terms in Hayes (2009), which we have liberally interpreted here as signalling a principally articulatory conception of their representational nature.

are quite rich in information, including rich acoustic and articulatory information as well as information relating to context, semantics, extra-linguistic context, etc. What is perhaps most functionally reminiscent of the role taken by phonological primitives in the structuralist and generativist approaches does not necessarily constitute a mental entity as such but rather are generalisations over essentially continuous data as represented by the exemplars in a cloud and its wider network. Of course, there is also a wide variation of views regarding abstraction and discretisation within exemplar-based models (cf. e.g. Bybee and McClelland 2005, Ernestus 2014, Pierrhumbert 2016), but at a fundamental level the primitives of exemplar-based theories are the exemplars themselves, which by their own nature cannot be fairly tried in the court of a symbol-based dichotomy between acoustic and phonological correlates. A more apt question for exemplar-based views may be to ask whether a separatistic treatment of acoustic versus articulatory generalisations over exemplars makes similar explanatory promises as what is likely hoped for by subsegmental phonologists subscribing to an exclusionary view of either type of correlate. The architecture of Functional Phonology (Boersma 1998) is also interesting in this regard, proposing that phonology has separate perceptual and articulatory representational levels. Consequently, rather than the hybrid position seen in many feature systems (recall e.g. Hayes 2009), both acoustic and articulatory correlates take primacy, but they do so at their respective levels of the grammar, with some mediation through interaction between the perceptual output of the articulatory production grammar and the perceptual recognition grammar. The main question raised by this proposal is then not so much about the type of correlates involved in phonology, but rather jabs back at the Chomskyan proposal that it is not necessary (and perhaps not even desirable) to accord equal linguistic status to both the perceptual and articulatory aspects of speech given that a hearer is in the business of reconstructing an abstract stored representation given the cues fed into the system's input.

1.3.3 Primitives and substance

Now that we have looked in some depth at the phonology-external correlates of phonological primitives, let us turn to a question that has (at least in some parts of theoretical phonology) been a major theme in recent times. Namely, given that phonological primitives tend to correlate more or less closely with specific external phonetic cues and/or events (whether articulatory, acoustic, perceptual, or a mixture), does this external identity of the primitives play any role in phonology? Simplistically, advocates of substance freeness would say that the answer is no, while advocates of a more phonetically grounded persuasion would answer in the affirmative. The specific answer we give to this question has potential ramifications for many aspects of phonology, including the acquisition of phonology, the mapping between phonology and phonetics, the mental representation of phonological objects, and phonological computation. As with many of the questions we have discussed so far, there is a whole range of positions that have been taken on the issue at each of these levels. While we will turn more extensively to the acquisition question in Section 1.4, and focus our discussion here on the role of substance at the other levels.

At the level of computation, the principal question is whether phonology is sensitive to the substantive properties of the primitives (i.e. what its phonetic correlates are and what they 'mean' or 'imply'). Take for instance the SPE-style vocalic features [±high] and [±low] in their articulatory conception. Introductory phonology students are typically taught that mid vowels in systems with three or more height distinctions are represented with the features [−high, −low]. That is, compared to high and low vowels, they are 'neutral' in terms of the height of the tongue body and the aperture of the jaw. When students (typically some time later in their training) ask why there are no [+high, +low] vowels. A common answer that seems to be invariably satisfactory is that the articulatory instructions that are represented by these features are not compatible: since [+high] instructs the lower jaw to adduce and the tongue body to elevate, while [+low] instructs the lower jaw to abduce and the tongue body to depress, such vowels are not possible. Note that this does not actually address the actual question however.

While it may be all true that these two sets of opposing instructions could not be implemented simultaneously and therefore there can't be any vowels that arise from a combination of these articulatory events, it doesn't actually explain why there shouldn't be any vowels that are *phonologically* specified [+high, +low]. Sure, we might expect this to be a rare occurrence given no vowel is substantively produced in this way, but we can perfectly well conceive of situations that will phonologically derive a representation containing [+high, +low]: suppose for instance a height harmony rule that appears to simply spread the feature [+high] leftward. We might want to try and capture this with a rule such as V → [+high] / ___ C_0 [+syllabic, +high]. If such a rule is applied to a low vowel (specified [+low]), this should result in a segment with the feature combination [+high, +low], the 'impossible vowel'. In no language with height harmony do we find evidence of ineffability in such circumstances however. Rather one of three possibilities obtains: i) a low vowel remains low, ii) a low vowel raises to mid, or iii) a low vowel raises to high.

There are two common solutions to account for the first case, which are surprisingly interesting. If we say that rule application is blocked because (universally) [+high, +low] cannot be combined we explain the absence by appealing to the phonetic substance within phonology — phonology does not combine features whose correlates are incompatible. On the other hand, if we say that this is simply because the rule does not apply to low vowels, i.e. we modify the input to [+syllabic, −low] → [+high], the absence of the rule not including [−low] for the input specification must be explained. We could again appeal to the substance and say phonology not only doesn't allow rules to apply if the correlates of the output's features are incompatible, but it doesn't even allow rules that could lead to such an output. An alternative explanation appeals to externalisation: a speaker/hearer would not posit this rule because it would lead to ineffability if they had no other rule that fixed these incompatible feature specifications before they are phonetically implemented, but since there is no evidence for a rule that turns low into non-low vowels, the only plausible solution for the learner is to posit that the rule does not apply to low vowels. The substance matters not at the level of phonology, but at the level of phonetic implementation.

The second and third situation, where low vowels raise to mid or high can similarly be given a substantive and a non-substantive explanation. If phonology does not allow the feature complex [+high, +low] to be formed due to the substantive conflict, then any rule with V as an input in a system with [+low] vowels must prevent [+high, +low] as a possible outcome and so only V → [+high, −low] or V → [−high, −low] are possible phonological rules. Alternatively, we may let phonology derive [+high, +low] and then have one of the rules [+high, +low] → [+high, −low] or [+high, +low] → [−high, −low] as 'repairs'. If we assume these repairs to be somehow innately hardcoded (and the learner's choice boils down to which they order before the other) this again appeals to the substance within phonology, but we need not do this. As before, we can alternatively appeal to externalisation: if [+high, +low] is unpronounceable but words which the rule applies to are not ineffable, the learner is motivated to look for another rule that matches the data, and can motivate this by the observed alternation. The only solution to the phonetic restrictions that remains open is to encode the result in the rule itself, or to posit an additional 'repair' rule, but the substance doesn't matter to phonology at all. (In fact, we could also just let [+high, +low] pass through, and quite plausibly assume that if phonetics is faced with the task of opposing instructions it will take the middle ground, so [+high, +low][5] may not even problematic to externalisation).

What this perhaps overly laborious example illustrates well is that it might seem quite natural and explanatorily fruitful to exploit the phonetic substance in the phonology, but in virtually all cases this could also be relegated to a later level and it is extremely difficult from phonological behaviour alone to clearly tell whether substance matters at the level of computation or whether its consequences come to bare only post-phonologically.[6] Consequentially, arguments for one or the other position is usually sought elsewhere. The substantive approach may appear to be more explanatorily satisfying to the phonologist, since it gives an explanation from within the phonology for many common phenomena while at the same time providing biases against those which are rarer. On the other hand, the substance free approach forces us to find explanations for such matters elsewhere, but offers a much simpler conception of the phonological system itself, and if some phenomenon can already be explained extra-phonologically, why should we even try to explain it again through the phonology? Phonological computation without appeal to phonetic is more readily compatible with modular conceptions that separate phonology and phonetics into different components. A clear consequence of domain-specificity and informational encapsulation is that modules do not directly share their vocabulary and inner workings, which precludes phonetically substantive analyses within a separate phonological module, and it is not surprising that modularist phonologists typically assume phonology to be substance free at this level. If we do admit reference to phonetic substance within the phonology we do of course face another question: is phonology potentially fully sensitive to all the phonetic correlates of the primes, or are there restrictions on the types of properties and contexts within which they can be referred to?

Somewhat independently of the question whether phonological computation may be sensitive to phonetic substance or not is the question whether substance can inform the selection of the primitives to be stored in underlying forms and to be referenced in phonological processes. If substance is admitted, this seems like a fairly straight-forward task where we can start out with what-you-see(/hear/do)-is-what-you-get and then whittle away everything that becomes predictable or conflicting as we acquire more processes (cf. e.g. the approach suggested by Hale and Reiss 2000, 2008). Crucially, in all this a learner as well as a mature hearer faced with the task of matching their input against some hypothesis space for acquisition or lexical access can (ignoring the issue of noise) rely on the fact that if a vowel is low then its [+low], etc. However, if no reference to the phonetic substance is allowed in establishing the mental

[5] This is of course exactly what is assumed by acoustic approaches such as Element Theory, where mid vowels contain high |I| or high |U| in combination with low |A|.

[6] While challenging, the two positions are in principle testable. In our example here, what we would need is to find speakers with a vowel system that does not make use of [+low] and has high harmony plausibly implemented with a V → [+high] rule, then let them learn an artificial grammar that requires them to posit vowels with underlying [+low] but plausibly lets them transfer the height harmony rule without giving any evidence for the outcome on low vowels. If phonology is perfectly content with [+high, +low] *and* phonetics can implement it, they should always come up with mid vowels.

representations of phonological objects, the assignment of a primitive to a phonological form has to be based on phonological evidence alone. One type of evidence may be constituted by contrast, if /n/ and /t/ are contrastive then their representation has to be different in at least one aspect, if /n, t, d/ are all contrastive at least two primitives must be used to represent them (but there are already many different ways in which this could be done with the two features), if we notice that /n/ regressively nasalises vowels but /d, t/ do not then /n/ must contain at least one primitive that neither /t/ nor /d/ contains (viz. [nasal]), if we notice that both /t, d/ undergo flapping but no other stop does they must contain some common feature, etc. Similarly, if there is no evidence in the language that nasals show shared behaviour with voiced stops to the exclusion of voiceless ones, we would have no reason to posit that /n/ ought to be specified with a primitive contained in other voiced stops but not voiceless ones (viz. [voice]).

Finally, at the level of the interface, we may ask whether substance in some way restricts the mapping relations between primitives and their phonetic exponents. For instance, one may take the position that both at the levels of computation and the representation of phonological the substance is not relevant, but that nonetheless specific phonological features must have specific mapping relationships to phonetic cues. The most restrictive option is a restriction to one-to-one mapping, e.g. not both features [F_1] and [F_2] may map to the raising of the tongue body, and some feature [F_n] may not map to more than one phonetic cue. This is clearly incompatible with some theories, especially those which seek to starkly reduce the set of phonological primitives. For instance, Element Theory |U| encodes both cues for rounding and for backness, while |L| is commonly assumed to encode both voicing (periodicity) and nasality (Ploch 1999; Botma 2004; Nasukawa 2012; Breit 2017). This view is compatible with a slightly loosened mapping requirement allowing one-to-many relations from phonological primitives to phonetic cues, which appears to be a view widely entertained. Inversely, we could of course imagine a restriction to many-to-one relations from phonological primitives to phonetic cues, a view that (as far as we know) is not represented among feature theories. Finally, many-to-many relations may be permitted, e.g. [F_1] may map to phonetic cues A, B while [F_2] maps to phonetic cues B, C. Scheer (2020) argues that this is essentially what should be assumed if we take seriously a commitment to list-based translation at the phonology–phonetics interface, i.e. the mapping between the two is entirely arbitrary. An additional factor that comes into play, and can be seen mainly in dependency-based approaches such as Element Theory is context-dependency: may the phonetic interpretation of some phonological primitive be dependent upon its phonological environment? |L| in Element Theory again serves as a useful illustration. Standardly, this is assumed to encode voicing in head position (viz. |$\underline{\text{L}}$|) in onsets, low tone in head position in nuclei, and nasality as a dependent (cf. e.g. Backley 2011: 159). Of course, such questions pose important questions about what may and may not serve as context, and we would expect not only that locality conditions obtain, but that they give rise to certain observable phenomena, similar to what is known from morphosyntax (cf. e.g. the *ABA generalisation in morphology, Bobaljik and Sauerland 2018) — we would not, for example, expect that a feature [F] in a non-initial onset maps to voicing if the onset of the word-initial syllable is also specified for [F] but to nasality otherwise, although this type of contextual mapping could be employed in accounting for phenomena such as Lyman's Law in Japanese (Lyman 1894). The major phonological effect of what restrictions (if any) obtain on the mapping relations between phonological primitives and phonetic cues is distributional and combinatorial. If we assume for instance one-to-one mappings more phonological specification is necessary while simultaneously the same feature cannot be employed to different phonetically incompatible ends in different environments. If anything goes, then the same feature [F] may be employed for completely different phonetic purposes in different positions, predicting that there can be phonological interaction (due to the shared featural specification) without any tangent phonetic cue toward that shared identity.

In practice, many combinations of these positions across the various levels where substance may show some influence on phonology are taken. Grounded Phonology (Archangeli and Pulleyblank 1994) very radically advocates for a view where substance plays a significant role at every level based on arguments from phonological behaviour, phonological representation, and the interface. In Articulatory Phonology (Browman and Goldstein 1986, 1988, 1989; Pouplier 2020; SORENSEN AND GAFOS, this volume) substance naturally plays a significant role due to the very conception of phonology as a relatively direct representation and computation over articulatory gestures. Government Phonology (and its theory of primitives Element Theory) traditionally assume what has been termed the *Phonological Epistemological Principle* (Kaye 1989, 2011), which claims that both phonological computation and representation are based on observable phonological behaviour alone, thus denying the role of substance at these levels, while simultaneously assuming a relatively strict mapping relation (recall Kaye et al.'s (1984) Element Calculus; but also the mappings suggested in acoustics-based Element Theory such as Harris and Lindsey (1993, 1995)). Hale and Reiss (2008) version of substance-free phonology appears to go in a similar direction, assuming innate phonological features with presumably some type of pre-established (or at least biased) correlate so that substance may play a role in bootstrapping the initial setup of phonological representations but is otherwise completely irrelevant to phonological representation and computation. Building on this view it has also been proposed that some commonly assumed-to-be phonological phenomena can be relegated to an enriched version of the phonetic component, as in Reiss and Volenec's (2017) Cognitive Phonetics. This contrasts with versions of substance-free phonology that do not assume innate features or correlations in this way, for instance IOSAD (2017, this volume) combines the substance-free view with featural emergentism, such that the phonological features themselves are acquired only on the basis of the phonological distributions and behaviours that give rise to them, with relative consequentially freedom in their association with phonetic cues across the interface. HALL (this volume) is also notable in this regard, arguing for a middle-ground position between the radically grounded (e.g. Archangeli and Pulleyblank 1994; SORENSEN AND GAFOS, this volume) and the radically substance-free (e.g. Hale and Reiss 2000, 2008; Reiss 2021; IOSAD 2017, this volume). We will take up this issue again in Section 1.4, which is concerned with the origin of phonological primitives.

That there is no real agreement on such issues in sight is illustrated well by recent developments in Element Theory appear to manifest an ongoing split between those seeking to admit a more significant role of substance and those seeking to further eradicate it. For instance, Backley's (2011, 2017) theory of multiple headedness includes pair-wise combinatorial restrictions based on what Backley terms *element antagonism*, a relationship that obtains between two elements that encode acoustically opposed cues such as the aperiodic noise of |H| and the periodic murmur of |L|. BARONI's (this volume) work on the place elements |A, I, U| goes in a similar direction, advocating a closer tie between elements' (phonetic) identity and phonological behaviour than what has been traditionally assumed in the framework. Conversely, Cyran's (2014, 2017, in press) *laryngeal relativism*, which proposes that some voicing systems may phonologically be treated more like aspiration systems and vice-versa, clearly moves away from the narrow mapping assumptions that have long been made about the fortis and lenis primitives (|H| and |L|, respectively) in Element Theory (cf. e.g. Harris 1994:133-138) and have also been widely argued for outside that framework under the banner of *laryngeal realism* (cf. Honeybone 2005, Iverson and Salmons 2008). Most radically perhaps, Scheer (2014, 2020) argues that if modular commitments to domain-specificity are taken seriously, even at the mapping across interfaces the substance eventually correlated to some primitive ought to be treated as irrelevant — the appearance of such then may be rather epiphenomenal in that, for instance, if we had a system where the primitive [spread glottis] became to be associated consistently with phonetic voicing for some reason, and [voice] as phonetic aspiration, on the phonological level at least in the next generation we would simply relabel these and no real 'phonologically chaotic' effect persists. A very interesting question arising from this is, if we assume full substance-freeness across all phonologically relevant levels but do not adopt an emergentist view of phonological primitives themselves, why the set of primitives should be restrained in the way that many theories suggest. In other words, if this is how phonology is best characterised, then how come that language after language appears to have the same (or at least a remarkably similar) set of primitives, whether they be 6 elements, 23 distinctive features, or something else?

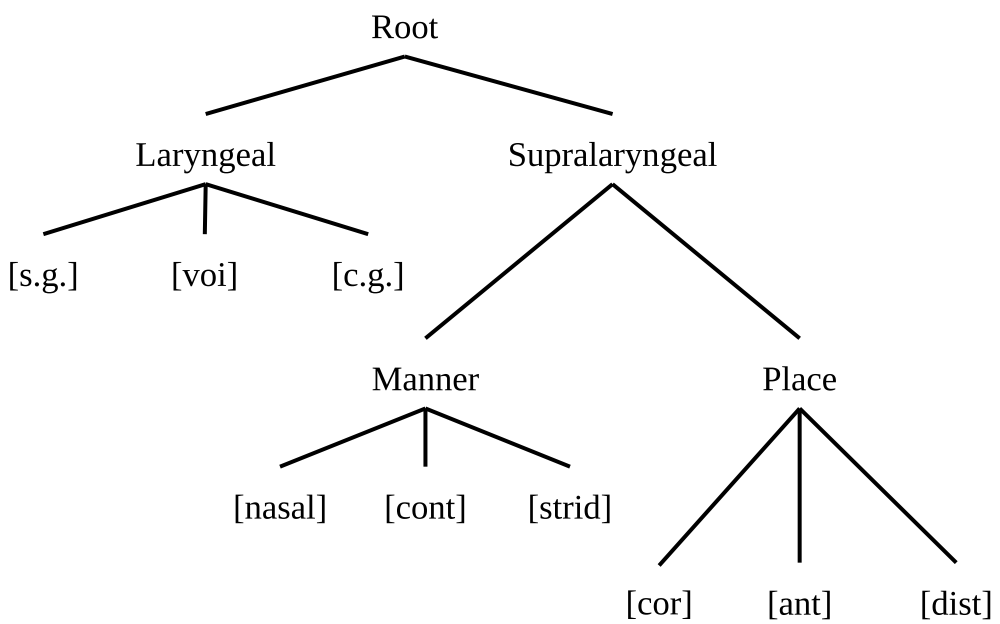
1.3.4 Subsegmental and subatomic structure

In Section 1.2.1 we have already discussed proposals which express or replace a function or property otherwise encoded by a phonological primitive with phonological structure, and obviously the question of whether some external contrast is actually correlated phonologically to a difference in feature specification or in structure is a highly interesting one. There is also quite clearly a strong theory-internal relation between a reduced number of features and enriched phonological structure, and theories might actively seek to displace properties onto structure in order to reduce the number of primitives. For instance, the replacement of |A| with structure has been somewhat of a mission-statement for GP2.0 (Pöchtrager 2006). At the extreme end, one may even contemplate replacing *all* of the primitives with 'pure structure' as suggested by Cavirani and van Oostendorp (2020). Since we have already discussed the benefits of a structural conception of features such as [(±)long] and [(±)syllabic] at some length earlier on, we will not take up this issue again and instead here turn our attention to other aspects of structure below the segment, such as the question whether phonological primitives are organised below the segment and whether they are really primitives in the sense that they are not constituted of any smaller parts themselves.

In Feature Geometry (Clements 1985, Clements and Hume 1995), there is the idea that feature organisation (structural relations between features) follows from the articulatory definition of the features involved. As Ewen (1995: 581) points out, in Feature Geometry "the interpretation of the dependency relations represents an attempt to formalize the constraints on human articulators." The main effect of such organisation of features below the segment is referential in nature. Whereas in non-structured feature theories phonological processes have to reference individual features, geometric arrangements allow reference to pre-established groups of features that may behave together. For instance, it is well-known that the place features ([labial], [coronal], [dorsal], etc.) often show class behaviour in assimilation processes such a progressive nasal place assimilation. Non-geometric theories in the SPE tradition have conventionally adopted notational convenience devices such as [α place] to denote such processes. However, under the hood [α place] is nothing more than a shorthand for [α labial, β coronal, γ dorsal, ...] — nothing in the phonological theory of features itself differentiates a class of features that commonly show shared behaviour (such as the place features) from an arbitrary one that doesn't (such as for example [α labial, β voice, γ high]). A segment-internal hierarchical (geometric) organisation into nodes of features formally accords phonological status to such classes of features, e.g. [labial], [coronal], and [dorsal] (possibly together with features such as [high], [low], [front], and [back]) may be grouped under a 'place' node, and features such as [anterior] and [distributed] themselves might be grouped under the [coronal] feature, etc. Such a feature geometry recognises that there are substantive relations between the features, usually based on shared attributes of their correlates (see Section 1.3 below), and the class behaviour exhibited by these specific subsets of the features. Crucially, in doing so, geometric approaches make predictions about which of the many possible subsets of the set of features are expected to show class behaviours and which may not. The fruitfulness of this endeavour is reflected in the fact that the geometric idea has not been confined to articulation-based feature theories in the SPE tradition, but has for instance also been adopted by many working in Element Theory (e.g. Harris 1994).

An interesting observation to be made about traditional feature geometries is that they are more information-rich than they strictly need to be. For example, Clements (1985: 248) illustrates a partial feature geometry by organising the nine features [spread glottis, constricted glottis, voiced, nasal, continuant, strident, coronal, anterior, distributed] into the hierarchical structure shown in (1):

(1)



While the feature geometry in (1) presents only a subset of the feature system assumed by Clements (1985), it will serve us well in illustrating the point of information duplication in the system. Suppose we were to replace each of the substantive labels in the terminal-dominating nodes of (1) with the symbols A, B, C instead. That is, the Laryngeal node dominates A, B, C, as does the Manner node, as does the Place node. If we were to do this, knowing what articulatory dimension the dominating nodes refer to, we could easily reconstruct the substantive information about the actual articulatory correlate described by each terminal, despite only using three distinctive features in place of the nine presented in (1). In familiar phonological terms, what A, B, and C refer to is predictable from its environment and thus need not be encoded in the theory.

Another interesting aspect of the A, B, C analogy is that it shows the capability of geometric feature arrangements to not only encode dominance relationships, but if we abstract away from the context-free interpretation of the features, also has the potential to encode orthogonal relations between features across the different branches of the tree. That is, we might not only expect some class behaviour between the features grouped under any one node (e.g. Manner), but also between all the A or B features in the tree, whether they be dominated by the Laryngeal, Manner, or Place node. This potential for 'orthogonal relations' is exploited explanatorily in Radical CV Phonology to account for specificational redundancy, i.e. relations in SPE-type feature systems where a specific value v_1 of a feature F_1 implies a specific (identical or different) value v_2 for a different feature F_2, i.e. $[v_1 F_1] \rightarrow [v_2 F_2]$. For instance, in very many languages if a vowel is [+back] it will also be [+round]. Van der Hulst (1994) argues that no matter what the phonology-external advantages of such a system may be (roundness has been argued to enhance the lowered F2 cue of backness), in some cases the external correlates and cues associated with the redundant specification are redundant, while at the same time SPE-type feature theories cannot capture any relation between these features in the phonology where they are effected.

While of course one solution to the redundant specification problem is unification, as in the |U| of Element Theory and related frameworks, which encodes both backness and roundness, what van der Hulst (1994) proposes is that these effects are due orthogonally shared feature specifications within a hierarchically organised segment. Like Feature Geometry, Radical CV Phonology assumes a geometric arrangement of segmental features under different nodes called subgestures (e.g. a primary and a secondary location subgestural node, the former of which encodes place, the latter harmonic vocalic properties such as height and backness). Only two primitives, C and V, are assumed, which may enter into different dependency relations, and each terminal of the segment can be specified by either just C, just V, or a dependency-arrangement between both primitives. The crucial effect of allowing each terminal to be specified by either primitive (or their dependent combinations) is that the orthogonal relations entertained above will obtain and redundancy relations can be explained as harmony processes between nodes. For instance, a location gesture where both the primary and secondary subgesture are C-headed is harmonic, but one where the former is V-headed and the latter C-headed is disharmonic and thus dispreferred, or 'marked'.

There are two interesting insights that can be gleaned from the Radical CV position described above:

First, it shows that to at least some degree, the encoding of orthogonal relations between 'primitives' has to be bought at the cost of substance. In a non-substance free feature theory, because the primitives are inherently tied to their correlates, they are fixed in their gestural-organisational affiliation: if [low] literally represents an instruction to the tongue and jaw to lower for greater aperture and [voice] literally represents an instruction to the larynx to effect adduction of the vocal folds, then it would seem preposterous to propose that either [low] or [voice] may be specified both as part of the laryngeal and the place node — no matter what we do, aperture of the oral cavity is not a property of laryngeal specification, and vocal fold adduction is not a property of place specification.

Second, while the smallest melodic building blocks of Radical CV theory are indeed just the two primitives C and V, much of phonology appears to nonetheless seem to be sensitive more to the larger units constituted by dependent configurations of the two features in context than the individual Cs and Vs themselves. For example, in the analysis of vowel harmony the appropriate level of description in Radical CV phonology appears to be the entire subgestural node with its particular C-V configuration (cf. e.g. van der Hulst 2012, 2018), which can in fact be relabelled into a set of elements very much akin to those found in Element Theory (cf. van der Hulst, in press), so much so that often analyses can be transferred nearly one to one between the frameworks. Van der Hulst (1994) also points out that phonology may be sensitive to the individual C and V primitives in some dependency configurations (he calls these complex) but oblivious to the compositional nature in others (he calls these simple). This very much brings back the question raised earlier by Kaye et al.'s (1985) Element Calculus: may what we have here referred to as the primitives actually be internally complex units even if they appear to be treated as atoms at the level of phonological computation?

As with most aspects, the range of answers varies widely. In Articulatory Phonology and 'standard' Element Theory the primitives are indivisible monoliths that stay constant throughout the phonological computation. In the SPE conception, the various levels of featural specification present a continuum mainly expressed through valuation, i.e. the primitives themselves are monolithic and stay constant throughout while the values computed over them shift successively from more abstract phonological computations involving underspecification and the markedness values *u* and *m* to segments fully specified in all binary features to more fine grained scalar phonetic representations that can be fed into the articulatory process. Classical Element Calculus assumes that the elements are internally complex fully specified SPE-style feature matrices (including internal annotations for *hot* and *cold* features, essentially analogues of SPE's *m* and *u*, respectively). However, at the level of phonological computation, they appear as autonomous atoms and phonology has no access to their internal parts — the complex internal make-up only becomes relevant during the actual element calculus as part of the spell-out procedure of the final phonological representations. Radical CV phonology, as we have seen above, takes somewhat of an intermediate position: simple dependency configurations such as {C}, {V}, {C_V} and {V_C} present to the phonology as atoms, while the internal make up of complex dependency configurations such as {C;V} and {V;C} is transparent to the phonology. What we are not aware of is any theory where all the primitives are internally complex *and* the internal make up is consistently transparent and relevant to phonological computation. A possible reason for this might be that such a theory might appear definitionally inconsistent, as one might reason that if the subatomic units are consistently accessible and phonologically meaningful, then we cannot possibly define the primitives *of phonology* as anything but the units that internally make up these primitives.[7] However, there is a possible conception of such a transparently internally complex yet phonologically atomic primitives that would appear to come fairly close, namely where one where phonological

[7] The problem of whether there can be a meaningful way in which we can speak about a *phonological primitive* in such a theory is remarkably reminiscent of the philosophical problem of religious language, which asks whether given some supposed divine attribute of God we can actually meaningfully talk about God's attributes (see Weed 2021 for an overview).

processes may be fully sensitive to the subatomic make-up of the primitive but they can do nothing that has any influence whatever that has influence (directly or indirectly) on the subatomic components at any stage of the computation, crucially including the spell-out process. Such a theory could perhaps be described as a formalisation of a maximally substance-abusing theory in the sense of Hale and Reiss (2008).

Finally, we may ask what happens if we take the ideas of structure-maximisation (vis-a-vis primitive reduction) and combine it with subatomic complexity. An example of such an endeavour is Cavirani and van Oostendorp's (2020) proposal that the vocalic elements |A, I, U| of Element Theory may internally be entirely void of any type of fundamental particle. Instead, they propose that the three resonance elements can all be reified as a set of different structures, consisting of treelets built out of empty multisets with a set of combinatorial restrictions (such as balancing of branches) which allow essentially four different types of treelet[8] encoding schwa, |A|, |I|, and |U|. The structures may be embedded within each other as long as no branching node has a depth difference > 1 across its children. By stipulating that what is phonetically interpreted as the equivalent of 'standard' Element Theory's |A| is the most complex of their base structures, they can in addition to all the subsets of the three elements themselves predict that |A| can exceptionally show two levels of embedding while the other two elements cannot. While interesting in that they can link this to the established observation (also underlying some of Pöchtrager's gripe with |A|) that |A| seems to show slightly different phonological behaviour to the other two resonance elements, the approach raises a fundamental ontological question about this type of structural solution. The structures and restrictions which they propose are fairly specifically designed to get the result they seek, to a degree we might say that both the structures and their linking to phonetic interpretations is based on stipulation alone — there are theoretically infinitely many different combinations of structure-building processes and tailored restrictions which would reify the exact same outcome. By way of a very silly demonstration of this, suppose that we take as a starting point the exact system Cavirani and van Oostendorp propose, but instead of the balance requirement we implement a pruning mechanism before phonetic interpretation which prunes all the leaves of the tree until the maximal possible structure remains that contains no depth difference greater than 1 across any two sister nodes. There are then an infinite number of structures we could build and all link to the exact same interpretational outcomes as before, and we could arbitrarily start by allowing first 1, and then successively n+1 levels of embedding to get an infinite number of different generating mechanisms all with the same interpretive result. What then, we should ask, is the difference to phonology (or more generally our conception of the mental representation of the primitive units of some external reality), whether we use a symbol |A| and stipulate that |A| may recur twice but |U, I| may not, or we come up with some arbitrary mechanism that does exactly this except without labels? If discernible by anything but appeals to conditions of economy etc. it seems most likely that, ceteris paribus, the answer would be found not in the predicted behaviour of the 'primitives' (or 'primitive structures') themselves, which can be manipulated relatively arbitrary through stipulation, but through deep insight into the

[8] These are two structures with depth 1: a unary projection {{}} and a balanced binary projection {{}, {}}; and two structures with depth 2: an unbalanced 2-level binary projection {{{}}, {}} and a balanced 2-level binary projection {{{}}, {{}}}. The 2-level binary structures can in fact be understood as embedding of the two 1-level structures within themselves.

representational, computational, and translational mechanism that are available to the relevant components of the mind.

1.4 The origin of phonological primitives

One might think that in a volume about phonological primitives, considerable attention would be paid to the developmental origins of such primitives. However, even the one contribution that deals with the acquisition of features only briefly mentions the issue, without taking a strong position either in favour of, or against, the innatist presumption.

This situation is indicative for the field as a whole, where questions of innateness are not usually tackled head on — unless in order to demonstrate the supposed fallacy of the innatist position. Most arguments, however, are made on the basis of theoretical considerations, not empirical results from studies on acquisition per se.

We may conclude at least two things from this situation: first, the phonological enterprise can fare quite well without knowing the answer to the question of innateness of primitives, and second, the question must be very difficult to answer, otherwise, we might presume, it would by now have been settled.

The latter conclusion, however, is phrased slightly misleadingly. Not only is the question difficult to answer, much worse it is very difficult to phrase — at least in a way that would lead to a testable hypothesis. Before we elaborate on the intricacies, however, it might be helpful to provide a brief background sketch on the stages of early phonological L1 development.

1.4.1 Proto-phonological development

A fair amount is known about the stages of linguistic development that children go through before we can unequivocally say that they are phonologically active agents - let us call such stages “proto-phonological”. Necessarily, such knowledge is heavily skewed towards perception. Although some studies of babbling have been done (De Boysson-Bardies et al. 1980; De Boysson-Bardies & Vihman 1991), most studies on pre-speech children have focused on perception. This is simply because children start speaking much later than they start hearing: first words typically occur around or shortly after the first birthday (Visser-Bochane et al. 2020), whereas perception has been shown to be *en route* to language specificity, perhaps even prenatally (Mehler et al. 1988; Rasmus et al. 2000; see also Gervain 2018 for an overview). Whether this simply comes down to a higher cognitive task load for production (for one thing, it must presuppose highly intricate motor control), or whether some other reason is at the core of this perception-production lag is something we cannot address here.

Seminal studies such as Werker and Tees (1984) and Kuhl et al. (1992) have mapped out the stages of early speech perception along the following lines: from birth, humans are able to accurately distinguish between all possible speech sounds. Exposure to a surrounding language, however, leads to a reorganisation of the perceptual map: at around 6-8 months of age, children no longer detect differences between vowels that do not contrast in the language that surrounds them, and with respect to vowel discrimination, perform equally poor as adult speakers of that language (Kuhl et al. 1992). Only a few months later, at around 10-12 months, a similar development occurs in the consonantal realm (Werker and Tees 1984).[9]

Not only do infants redraw their perceptual map to better suit the phonetic categories of their language, they proceed further to warp the perceptual space, so that the sensitivity to acoustic distances in absolute sense (for example, the F1-F2 plane in Euclidean space) is decreased near the centres of phonetic categories, yet enhanced in border regions. We know this as the Perceptual Magnet Effect (Kuhl, 1991), and together with the aforementioned redrawing of the lines (or, better said, erasing of irrelevant lines), it presumably serves two functions: by decreasing the sensitivity to minute differences in areas of relative certitude, the listener's perceptual system becomes more economical, and importantly, more robust. At the same time, strengthening the abilities to detect differences in regions where such differences may be expected to lead to categorical leaps (that is, from one potential phonemic category to the next), robustness is further enhanced.

It would seem, then, that around the first birthday, infants possess a relatively advanced, adult-like speech sound perception system. Unfortunately, from here on things become considerably more complicated. Such a perceptual map, with native-like boundaries and category-specific space-warping, is a far cry from what is traditionally considered "phonology". For one thing, while such knowledge of categories may be *prerequisite* to compute a system of actual phonological contrasts, it is by no means sufficient.

1.4.2 The origins question and the approaches to primitives

The difficulty of phrasing the innateness question becomes very apparent at this point. To take one naive example, such a perceptual system might seem difficult to rhyme with emergent, substance-free primitives: the substance is *right there*, why not use it. But this would be to make a caricature of substance free phonology: there is no necessity at all to deny that the perceptual map is used to construct emergent features, not even to deny that such features should be substantive or not. The real question is whether the map is used to provide substantive and language-specific mappings for a set of innate features, or whether to guide the process of parsing the input acoustically so that relevant generalisations can be detected. In other words, an empirical result that at first seems relevant to the matter at hand turns out to be compatible with two very different conceptions of phonology. In effect, the problem of how to pose the question of innateness of primitives boils down to a simple counter question: what are your primitives?

The question of substance was taken up in Section 1.3.3 above, but it is worth remembering here that the question of innateness cannot be properly asked without making absolutely clear what it is that is supposed to be innate or not — substance? No substance? Some substance? If primitives do contain reference to phonetic substance, does that mean acoustic information is encoded? Or articulatory instructions?

The contributions to the current volume, no matter how different their conceptions of phonological primitives might be, do all agree on one important insight which is so

[9] Recent studies have thrown doubt on whether this same timeline can be assumed for all speech sound differences. For one thing, a directionality effect has been shown to exist, where changes in the direction of a more peripheral vowel were detected, but not changes in the opposite direction (Polka & Bohn, 2011)

fundamental to most of current phonology that it usually goes unmentioned: there *are* primitives of phonology, and what is more, they are subsegmental in nature. This is by no means a logical necessity; not only was phonology practiced long before the advent of the distinctive feature (for most of the first half of the twentieth century, the primitive of phonology was the phoneme; see also the discussion in Section 1.2.3 above). More modern approaches such as the Exemplar Theory, spearheaded among others by Bybee (2001) (see also the contributions in Vihman and Keren-Portnoy 2013) challenge the notion of the subsegmental primitive (or at least its existential unavoidability) from a more *gestalt*-driven perspective. Here the acquisition of phonology consists of computing generalisations over the developing lexicon, which retains as much phonetic detail as it possibly can. This position is not represented in the current volume, and to our knowledge there has been no successful attempt to simulate the complete acquisition of a full phonological system using the principles of Exemplar Theory.[10] In the rest of this section, we will address the origins question in relation to some of the contributions to this volume. We will not address each one individually, as some are quite similar with regards to the acquisitionist perspective pursued in this section.

The proposals put forth in the eleven chapters presented here generally fall into three families of primitives: SORENSEN AND GAFOS (this volume) propose a set of *articulatory gestures* as being the primitives of phonological computation and representation (an idea that is virulently opposed by VAUX AND MILLER (this volume)). On the other end of the physical spectrum, BARONI (this volume) and PRINCE (this volume) work in Element Theory, where primitives are conceptualised as acoustic targets, and where the means by which the speaker reaches these targets seems unimportant. The remaining contributions work within one or another version of distinctive feature theory.

For both the gestural and the element-theoretic approaches, the matter of substance seems to be settled beforehand: a substance-free Element Theory is almost a contradiction in terms, as is a substance-free Gestural Phonology. Both approaches boast that substance — be it articulatory or acoustic — is built right into the very core of their respective primitives. With respect to the developmental origins of primitives, note that this does not automatically mean that the primitive cannot be emergent; we shall return to this matter further below.

Within distinctive feature theory, however, the landscape is more varied. In fact, we might see degrees of substance represented in the same proposal (consider, for example, the difference in concreteness between place and laryngeal features on the one hand, and major class features on the other). Similarly, while most of feature theory is mostly articulation-leaning, some features (e.g. [±sonorant]) may be principally acoustics-oriented.

With that in mind, let us review some of the perspectives offered in this volume, always asking the question: what does this say about the developmental origins of these primitives?

1.4.3 The origins of gestures and elements

In the gestural approach put forward by SORENSEN AND GAFOS (this volume) the mapping between the phonetic expression of primitives and their availability for phonological

[10] To be sure, even under an Exemplarist view, something in the child's disposition must bias them to retain all that phonetic detail, and to extract the meaningful generalisations over it. It would seem, then, that just as with all the other approaches, the question of innateness does not magically disappear.

computation and representation is most strict. In fact, they go so far as to say that theirs is the only framework in which "... the unit of phonological form is identical to the unit of speech production". This close connection opens the door to a comparatively early acquisition of phonological primitives; as soon as the child produces the sounds they underlie, the primitive may be said to be acquired, regardless of higher-level phonological activity. It appears exceedingly difficult, however, to determine whether such gestural primitives are learned or not. In models presented in Smith et al. (2021) and Smith and O'Hara (2021) it would seem that the initial state includes precise knowledge of articulatory targets — even if the learner is allowed to update these.

Presumably, perception in Articulatory Phonology proceeds by way of the listener reconstructing the gestures by which the incoming speech signal most likely was produced That is, the learning model in Smith and O'Hara (2021) affords the artificial "learner" with direct access to the precise measure of the degree of constriction produced by the "teacher". With regards to the acquisition of phonology, this then leads to a contradictory situation: as we saw above, a native language-like perceptual map is acquired before any meaningful speaking commences. Why then, must production lag behind perception? If the perceptual map is acquired so early, and if perception proceeds by the listener reconstructing the articulation of the speaker, it would seem that this information should be immediately accessible to the phonological system. One potential response is that the child's early perceptual map is qualitatively different from the later map, in that the former does not proceed by means of reconstructing articulation, whereas the second does. We could say that the early map is enriched with articulatory reconstructions later in development, but that hardly explains how the early system comes into being, and why it is necessary to make the latter. Some studies (e.g. Imada et al. 2006) have pointed to a possible neurological correlation between perception and production in pre-verbal infancy, but all in all the picture remains far from complete. Alternatively, perhaps we must conclude that Articulatory Phonology assumes a separate perception grammar.

Element Theory, represented here by BARONI (this volume) and PRINCE (this volume), makes strong claims of universality. The small, monovalent set of primitives is set to reduce generative capacity of the phonological grammar, but this aim towards numerical minimalism comes at the price of having to assume that the mapping between primitives and their substance (here: an acoustic target) is relatively loose, and language specific. Acquisition, therefore, must consist of either finding the correct mapping between the innate primitives and the language-specific instantiation, or alternatively, of finding the right primitives to begin with. In other words, if we go with Backley's (2011) six elements (or any other set for that matter), must we assume these are innate because they are universal? Or is the combination of species-specific learning bias and the input/uptake of such a design that precise set of stimuli an unavoidable outcome? Other than emergent features, which are often proposed hand-in-hand with substance-free phonology (see below), emergent elements need not arise solely motivated by phonological activity, but by virtue of their built-in substantive component, knowledge of the lexicon may be a substantial source of information for the learner. Interestingly, despite their firmly generativist background, if elements are shown to be emergent in such a manner, the question arises how far we are, conceptually, from the generalisations over the lexicon proposed in exemplar theory.

While Articulatory Phonology and Element Theory are very different in many ways, they do share the similarity that their primitives have an intimate connection to the phonetic reality. A very strict and immediate one in the case of Articulatory Phonology, and a more idealised and variable one in Element Theory, but substance is baked-in (on at least some level) nonetheless. As we have seen, however, this does not automatically lead to testable versions of the origins question. In the case of gestures, questions of perception remain, and in the case of elements, it remains an open question (albeit an intriguing one) whether a supposedly universal set of elements must be assumed to be innate, or whether they shall arise inevitably from generalisations over the lexicon and phonological interactions.

1.4.4 The origins of features

Having considered the two non-feature-based approaches presented in the current volume, let us now turn to those who do assume some kind of Distinctive Feature Theory. Whereas gestures and elements both appear to inherently come with strong ties to phonetic substance, the matter is much more diverse when it comes to features. The contributions in this book range from a strong substance-free perspective (IOSAD, this volume) to HALL's (this volume) proposal to curb the amount of substance, without giving up on the idea of substance at all. Most of the other feature-based contributions are not primarily concerned with the question of substance, but given the intertwined nature of the substance question and the origins question, it can be fruitful to attempt to ascertain their position in this landscape nonetheless. VAUX AND MILLER (this volume) for example, do acknowledge the relevance of "articulatory grounding", even if they warn against what they appear to view as an excess of articulatory grounding in Articulatory Phonology. LAHIRI AND KOTZOR's (this volume) Featurally Underspecified Lexicon (FUL) does assume some phonetic grounding, but in their approach the relationship between phonology and phonetic substance is sometimes crucially defined in negative terms (hence, the term *underspecified*). It should be noted that the FUL-model has been successfully applied in studies of language acquisition (see, for example, van der Feest 2007 and Altvater-Mackensen 2010), but here too, the theory remains in principle compatible with both emergent and innate features.

IOSAD (this volume) is an exponent of the position that phonological primitives (here, features) are both *emergent* and *substance free*. The argument goes that such assumptions are necessary to account for phonological processes that are not phonetically (substantively) natural. If we define membership of a natural class by means of participating in shared phonological behaviour, and if it can be shown that such behaviour is shared by a set of phonetically unrelated segments, *and* if natural classes are defined by a shared phonological primitive, it must follow that such primitives are blind to the phonetic substance that the class members bring to the table. In addition, barring divine insight, infants have no way of knowing which classes they will encounter, and hence which features they must posit. In other words, features must be *emergent* and cannot be innate.

While at first this seems to be a logical line of reasoning, there is a real problem with phrasing the question "where do features come from?". The innatist position goes hand in hand with the assumption of a universal set of features, but if features are really devoid of phonetic substance, how can we be sure there is *not* such a universal set? Any theory of phonology, except those that adhere to the strictest mapping conventions, must allow for languages to differ in whether and how they employ features to denote classes (this is a hallmark of the Modified

Contrasitivist Hierarchy, see Dresher 2009; and HALL, this volume), even if the features are the same from one language to the next. If languages are fully free to employ features as they please, since those features possess no phonetic substance whatsoever, there is no way of testing whether those features are truly emergent or not. The more abstract the primitive, the more difficult it is to track its origins. Again, we run into the virtual impossibility of the question where features come from.

HALL (this volume) shares with IOSAD (this volume) its emphasis on class behaviour as fundamental to defining primitives. His "contrastivist hypothesis" predicts that only those features that denote contrasts in the language can be available for phonological computation, and vice versa. Features are substantive to some degree, and not necessarily emergent.

This shared emphasis on phonological computation as the foundation of primitives is quite interesting with regards to our origins-question: regardless of whether features are substantive or not, any empirical evidence for their having been acquired can only be obtained as soon as the child becomes an active user of phonology. This means that an approach followed by, for example, Levelt (1994) cannot be considered valid, as she regards features as "acquired" if and when they appear in the child's output, regardless of whether there is any evidence for *activity* of that feature whatsoever. To put it differently: may we assume that children in the single-word-stage have lexical representations in terms of features?[11]

An admittedly rather formalistic way out of this conundrum is proposed by VAN 'T VEER (this volume). In his proposal there is no way for any feature to be present in the child's output without it also being the subject of phonological computation, specifically by means of Feature Co-occurrence Constraints. In this proposal, the acquisition of a primitive at the representational level immediately and automatically triggers the acquisition of that same primitive at the computational level. The problem here is that phonological computation is understood in such broad terms that it encompasses much more than what is traditionally understood to constitute evidence of phonological computation, such as alternations.

Within the framework of Distinctive Feature Theory, the Onset Prominence (OP) framework (SCHWARTZ, this volume) proposes that of the three traditional degrees of freedom for consonants (place, laryngeal, manner) the former two are simple in that their substance relates to one plane only (place and timing, respectively) whereas manner distinctions relate to both timing and degree of constriction simultaneously. Hence, the proposal goes, manner should not have the same representational status as place and laryngeal, but rather be replaced by a downward extension of the prosodic hierarchy, consisting of the levels of *Closure*, *Noise*, *Vowel Onset* and *Vowel Target* from top to bottom. Each of these levels is binary branching, such that the leftmost daughter hosts the next manner level down, and the rightmost daughter presents a potential docking place for place and/or laryngeal features. Typological differences in for example the degree to which consonant-vowel co-articulation is permitted can be accounted for by differing the level at which vocalic features dock.

[11] An interesting intermediate solution has been proposed by Fikkert & Levelt 2008, who argue that while the child has access to features from a very early age, they may not necessarily be *subsegmental*. Rather, features apply first at word level, with the vowel being dominant, whereas onsets and codas gain the ability to carry their own feature specifications later in development. This approach is close in spirit to the Featurally Underspecified Lexicon model proposed by LAHIRI & KOTZOR (this volume).

This presents an interesting perspective on an observation that has been making the rounds among acquisition-oriented phonologists for a while now. As proposed by Rose (2000), Inkelas and Rose (2007), but also in the context of atypically developing children (Chiat, 1989; Marshall et al., 2002; Marshall and Chiat, 2003), children appear willing to readily sacrifice segmental faithfulness in order to safeguard the expression of prosodic strength relations. If, as the thinking in the OP framework goes, what we formerly thought of as “manner” features really is a matter of prosody instead, we should predict that children prioritise manner over place and laryngeal in honing their phonological accuracy. On the surface, this appears to be a problematic prediction, as it is well known that children have a strong preference for stop-vowel-continuant word shapes, making for (possibly near) homophony of words like Dutch *poes* (target: /pus/; child form: [puf]) and *soep* (target: /sup/; child form: [puf]) for example.[12] However, until thoroughly tested, such objections are premature.[13]

1.4.5 Where we are: the irrelevance of innateness

We began this subsection with the observation that as a whole, phonologists have no answer to the question of the developmental origins of primitives. While at first this seems a sorry state of affairs, it makes it all the more remarkable that the pursuit of phonological descriptions and theories does not appear to be hindered by this gap. Does this mean, then, that we can put the question aside as both unanswerable and irrelevant?

Perhaps the value of the origins question does not lie in whatever its answer might ultimately turn out to be, but precisely in the difficulty it represents when trying to formulate it in a concrete and testable way. In this section, we have tried to show how different assumptions about primitives may lead to different questions, focusing mostly on the relationship between the proposed primitive and its phonetic correlate: the question of substance. This is motivated in part because it sheds light on the question what it is, that is supposed to be either emergent or innate, and furthermore, what type of evidence is available to the learner. It would seem that the more substantive the primitive, the more the learner may take cues from the lexicon, whereas in the case of fully substance-free primitives, the learner must rely on phonological behaviour to identify the relevant classes.

1.5 Summary

In this chapter we have attempted to ask, in very many different ways, the question what a phonological primitive is, and sought to find out something about this by looking at the various positions that have been set out about phonological primitives, regardless of the commitments of any particular framework or theoretical perspective on phonology. This has been quite insightful in showing that, while of course interdependent, many of these issues can nonetheless be quite fruitfully discussed in isolation and we can find theoretical positions that answer quite differently to different subquestions about the nature of phonological primitives. One might be disheartened at the picture of a field so severely fragmented on as fundamental a notion as the melodic building blocks of the very thing (phonology) that is studied by its community.

[12] Both examples from Fikkert & Levelt (2008: examples 9a, 9b). Note that these examples also include PoA deviations from the target, but that is beside the point for the current discussion.

[13] For one thing, such superficial objections are a far cry from taking (the development of) prosodic structure seriously. Again, see Fikkert & Levelt (2008) for ideas on how the tandem acquisition of prosodic structure and segmental phonology provides valuable insights.

However, what we hope to have shown here instead is that there is quite clearly some common ground, both in that for instance virtually all phonologists subscribe to a notion of an atomic discrete representational unit at the subsegmental level, the properties of which should have some explanatory power regarding phonological phenomena, and in that there seems to be a common interest in addressing these at times quite difficult and interdependent questions. Overall, if phonologists disagree about many of these aspects but care about them and show interest in investigating them, there is a good chance that this will in the long run lead to progress. Therefore, without recapitulating in any way the many questions (and corresponding answers) we have discussed throughout the chapter, we hope that the situation we have illustrated at large here will make clear the importance of continued work on the primitives, and perhaps also require us to find new ways to test some of the claims we have encountered. The various chapters that form the remainder of this book should serve well set against this background, as they very much illustrate many of the issues at stake, offer some potential avenues to address them, and at times leave us with yet more questions to ask.

References

Anderson, J.M. & Ewen, C.J. (1987) *Principles of dependency phonology*. Cambridge: Cambridge University Press.

Anderson, J.M. & Jones, C. (1974) Three theses concerning phonological representations. *Journal of Linguistics*, 10:1-26.

Altvater-Mackensen, N. (2010) *Do manners matter? Asymmetries in the acquisition of manner of articulation features*. Doctoral dissertation, Radboud Universiteit Nijmegen.

Archangeli, D. & Pulleyblank, D. (1994) *Grounded Phonology*. Cambridge, MA: MIT Press.

Backley, P. (2011) *An introduction to Element Theory*. Edinburgh: Edinburgh University Press.

Backley, P. (2017) Headedness in Element Theory: The case for multiple heads. *Glossa: a journal of general linguistics*, 2(1):89. https://doi.org/10.5334/gjgl.463

Bafile, L. (2019) The question of overgeneration in Element Theory. In: L. Franco & P. Lorusso (eds.), *Linguistic variation: structure and interpretation*, Berlin: De Gruyter Mouton, pp. 19-28. https://doi.org/10.1515/9781501505201-003

Berent, I., Platt, M., Theodore, R., Balaban, E., Fried, P.K. & Pascual-Leone, A. (2020) Speech perception triggers articulatory action: evidence from mechanical stimulation. *Frontiers in Communication*, 5:34. https://doi.org/10.3389/fcomm.2020.00034

Bobaljik, J. & Sauerland, U. (2018) ABA and the combinatorics of morphological features. *Glossa: a journal of general linguistics*, 3(1):15. https://doi.org/10.5334/gjgl.345

Boersma, P. (1998) *Functional Phonology: Formalizing the interactions between articulatory and perceptual drives.* LOT Publications.

Botma, B. (2004) *Phonological Aspects of Nasality: An Element-based Dependency Approach.* LOT Publications.

Botma, B. & van Oostendorp, M. (2012) A propos of the Dutch vowel system 21 years on, 22 years on. In: B. Botma & R. Noske (eds.), *Phonological Explorations: Empirical, Theoretical and Diachronic Issues*. Berlin: De Gruyter, pp. 135-154.

Breit, F. (2013) *Formal aspects of Element Theory*. MRes dissertation, University College London.

Breit, F. (2017) Melodic heads, saliency, and strength in voicing and nasality. *Glossa: a journal of general linguistics*, 2(1):85. https://doi.org/10.5334/gjgl.462

Breit, F. (in press) Sets of Elements. In: F. Breit, Y. Yoshida & C. Youngberg (eds.), *Elements, Government and Licensing: Developments in Phonology*. London: UCL Press.

Browman, C.P. & Goldstein, L. (1986) Towards an articulatory phonology. *Phonology*, 3(1):219-252.

Browman, C.P. & Goldstein, L. (1988) Some Notes on Syllable Structure in Articulatory Phonology. *Phonetica*, 45:140-155.

Browman, C.P. & Goldstein, K. (1989) Articulatory gestures as phonological units. *Phonology*, 6:201-251.

Bybee, J. (2001) *Phonology and Language Use*. Cambridge: Cambridge University Press.

Bybee, J. & McClelland, J.L. (2005) Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. The Linguistic Review, 22(2-4):381-410. https://doi.org/10.1515/tlir.2005.22.2-4.381

Cavirani, E. & van Oostendorp, M. (2020) A theory of the theory of vowels. In: K. Nasukawa (ed.), *Morpheme-internal Recursion in Phonology*. Berlin: Mouton de Gruyter, pp. 37-56.

Chiat, S. (1989) The relation between prosodic structure, syllabification and segmental realization: Evidence from a child with fricative stopping. *Clinical Linguistics & Phonetics*, 3(3):223-242.

Chomsky, N. & Halle, M. (1968) *The Sound Pattern of English.* Cambridge, MA: MIT Press.

Clements, G.N. (1985) The Geometry of Phonological Features. *Phonology Yearbook*, 2:225-252.

Clements, G.N. (1990) The role of the sonority cycle in core syllabification. In: J. Kingston & M. Beckman (eds.), *Papers in Laboratory Phonology I: Between the grammar and physics of speech.* Cambridge: Cambridge University Press, pp. 282-333.

Clements, G.N. & Hume, E. (1995) The internal organization of speech sounds. In: J. Goldsmith (ed.), *Handbook of Phonological Theory*. Oxford: Blackwell, pp. 245-306.

Cyran, E. (2014) *Between phonology and phonetics: Polish voicing.* Berlin: De Gruyter Mouton.

Cyran, E. (2017) ‘Voice’ languages with no [voice]? Some consequences of Laryngeal Relativism. *Acta Linguistica Academica*, 64(4):477-511. https://doi.org/10.1556/2062.2017.64.4.1

Cyran, E. (in press) Production-bias and substance-free representation or laryngeal distinctions. In: F. Breit, Y. Yoshida & C. Youngberg (eds.), *Elements, Government and Licensing: Developments in Phonology*. London: UCL Press.

De Boysson-Bardies, B., Halle, P., Sagart, L. & Durand, C. (1989) A crosslinguistic investigation of vowel formants in babbling. *Journal of Child Language*, 16(1):1-17. https://doi.org/10.1017/S0305000900013404

De Boysson-Bardies, B. & Vihman, M.M. (1991) Adaptation to language: Evidence from babbling and first words in four languages. *Language*, 67(2):297-319. https://doi.org/10.2307/415108

de Saussure, F. (1916) *Cours de linguistique générale*. Publié par Charles Bally et Albert Sechehaye ; avec la collaboration de Albert Riedlinger. Éd. critique préparée par Tullio de Mauro. Paris: Payot.

Dresher, B.E. (2009) *The contrastive hierarchy in phonology*. Cambridge: Cambridge University Press.

Dresher, B.E. (2021) Contrast in Phonology 1867-1967: History and Development. In: H.G. van der Hulst & E. Dresher (eds.), *The Oxford Handbook of the History of Phonology*. Oxford: Oxford University Press, pp. 231-252.

Durvasula, K. (2009) *Understanding nasality*. Doctoral dissertation, University of Delaware.

Ernestus, M. (2014) Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, 142:27-41. https://doi.org/10.1016/j.lingua.2012.12.006

Ewen, C.J. (1995) Dependency Relations in Phonology. In: J.A. Goldsmith (ed.), *The Handbook of Phonological Theory*. Oxford: Blackwell, pp. 570-585.

Fant, G. (1960) *Acoustic Theory of Speech Production, with calculations based on X-ray studies of Russian articulations*. The Hague: Mouton.

Fikkert, P., & Levelt, C. (2008) How does Place fall into Place? In: B. Avery, E. Dresher & K. Rice (eds.), *Contrast in phonology*. Berlin: De Gruyter Mouton, pp. 231-270.

Fleming, L. (2017) Phoneme inventory size and the transition from monoplanar to dually patterned speech. *Journal of Language Evolution*, 2(1):52–66,

Ganesalingham, M. (2013) *The Language of Mathematics: A Linguistic and Philosophical Investigation*. Berlin: Springer.

Gervain, J. (2018) The role of prenatal experience in language development. *Current Opinion in Behavioural Science*, 21:62-67. https://doi.org/10.1016/j.cobeha.2018.02.004

Goldsmith, J.A. (1976) *Autosegmental Phonology*. Doctoral dissertation, Massachusetts Institute of Technology.

Golston, C. & van der Hulst, H.G. (2000) Stricture is Structure. In: B.J.H. Hermans & M. van Oostendorp (eds.) *The derivational residue in Phonology*. Amsterdam: John Benjamins, pp. 152-175.

Grijzenhout, J. (2001) Devoicing and voicing assimilation in German, English, and Dutch: a case of constraint interaction. In: K. Dziubalska-Kołaczyk (ed.), *Constraints and Preferences*. Berlin: De Gruyter Mouton, pp. 201-228.

Hale, M. & Reiss, C. (2000) 'Substance abuse' and 'dysfunctionalism': Current tends in phonology. *Linguistic Inquiry*, 31(1):157-169. https://doi.org/10.1162/002438900554334

Hale, M. & Reiss, C. (2008) *The Phonological Enterprise*. Oxford: Oxford University Press.

Hall, D.C. & Mielke, J. (2011) Distinctive features. *Oxford Bibliographies in Linguistics*. https://www.oxfordbibliographies.com/view/document/obo-9780199772810/obo-9780199772810-0012.xml

Halle, M. (1959) *The Sound Pattern of Russian.* The Hague: Mouton & Co.

Harris, J. (1994) *English Sound Structure*. Oxford: Blackwell.

Harris, J. (2006) The phonology of being understood: Further arguments against sonority. *Lingua*, 116(10):1483-1494. https://doi.org/10.1016/j.lingua.2005.07.009

Harris, J. & Lindsey, G. (1993) There is no level of phonetic representation. *UCL Working Papers in Linguistics*, 5:354-373.

Harris, J. & Lindsey, G. (1995) Segmental decomposition and the signal. In: W.U. Dressler, M. Prinzhorn & J.R. Rennison (eds.), *Phonologica 1992: proceedings of the 7th International Phonology Meeting*. Turin: Rosenberg & Sellier, pp. 97-106.

Hayes, B. (2009) *Introductory Phonology*. Oxford: Wiley-Blackwell.

Henn, B.M., Gignoux, C.R., Jobin, M., Granka, J.M., Macpherson, J.M., Kidd, J.M., Rodríguez-Botigué, L., Ramachandran, S., Hon, L., Brisbin, A., Lin, A.A., Underhill, P.A., Comas, D., Kidd, K.K., Norman, P.J., Parham, P., Bustamante, C.D., Mountain, J.L. & Feldman, M.W. (2011) Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proceedings of the National Academy of Sciences*, 108(13):5154-5162. https://doi.org/10.1073/pnas.1017511108

Honeybone, P. (2005) Diachronic evidence in segmental phonology: the case of obstruent laryngeal specifications. In: M. van Oostendorp & J. van de Weijr (eds.), *The internal organization of phonological segments.* Berlin: De Gruyter Mouton, pp. 319-354.

Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A. & Kuhl, P.K. (2006) Infant speech perception activates Broca's area: a developmental magnetoencephalography study. *Neuroreport*, 17(10):957–962. https://doi.org/10.1097/01.wnr.0000223387.51704.89

Inkelas, S. & Rose, Y. (2007) Positional neutralization: A case study from child language. *Language*, 83(4):707-736. https://doi.org/10.1353/lan.2008.0000

Iosad, P. (2017) *A substance-free framework for phonology: An analysis of the Breton dialect of Botha*. Edinburgh: Edinburgh University Press.

Iverson, G.K. & Salmons, J.C. (1995) Aspiration and laryngeal representation in Germanic. *Phonology*, 12(3):369–396.

Iverson, G.K. & Salmons, J.C. (2008) Germanic aspiration: Phonetic enhancement and language contact. *Sprachwissenschaft*, 33(3):257-278.

Jakobson, R. (1941) *Kindersprache, Aphasie und allgemeine Lautgesetze*. Uppsala: Almqvist & Wiksell.

Jakobson, R. (1949) On the identification of phonemic entities. *Travaux du Cercle Linguistique de Copenhague*, 5:205–213.

Jakobson, R., Fant, G. & Halle, M. (1952) *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates.* Cambridge, MA: MIT Press.

Jakobson, R. & Halle, M. (1956) *Fundamentals of Language*. The Hague: Mouton & Co.

Jensen, S. (1994) Is ʔ an element? Towards a non-segmental phonology. *SOAS Working Papers in Linguistics and Phonetics*, 4:71-78.

Jesperson, O. (1904) *Lehrbuch der Phonetik.* Leipzig: B.G. Teubner

Kaye, J. (1989) *Phonology: A cognitive view*. Hillsdale, NJ: Lawrence Erlbaum.

Kaye, J. (2011) "GP, I'll have to put your flat feet on the ground". In: H. Broekhuis, N. Corver, R. Huybregts, U. Kleinhenz & J. Koster (eds.), *Organizing Grammar: Linguistic Studies in Honor of Henk van Riemsdiijk*. Berlin: De Gruyter Mouton, pp. 283-288. https://doi.org/10.1515/9783110892994.283

Kaye, J. & Harris, J. (1990) Segmental complexity and phonological government. *Phonology*, 7(1):255-300.

Kaye, J., Lowenstamm, J. & Vergnaud, J.-R. (1985) The internal structure of phonological elements: a theory of charm and government. *Phonology Yearbook*, 2:305-328. https://www.jstor.org/stable/4419961

Kaye, J., Lowenstamm, J. & Vergnaud, J.-R. (1990) Constituent structure and government in phonology. *Phonology*, 7(2):193-231. https://www.jstor.org/stable/4420016

Kim, Y. (2002) Phonological Features: Privative or Equipollent? AB dissertation, Harvard University.

Kuhl, P.K. (1991) Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50(2):93-107.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., & Lindblom, B. (1992) Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044):606-608.

Ladefoged, P. (1971) *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.

Ladefoged, P. (1975) *A Course in Phonetics.* New York: Harcourt Brace Jovanovich.

Ladefoged, P. & Maddieson, I. (1996) *The sounds of the world's languages.* Oxford: Blackwell.

Lenneberg, E.H. (1967) *Biological Foundations of Language*. New York, NY: John Wiley & Sons.

Levelt, C. (1994) *On the acquisition of place.* Doctoral dissertation, Rijksuniversiteit Leiden.

Lindsey, G. & Harris, J. (1990) Phonetic interpretation in Generative Grammar. *UCL Working Papers in Linguistics*, 2:355-369.

Lyman, B.S. (1894) Change from surd to sonant in Japanese compounds. *Oriental Studies of the Oriental Club of Philadelphia*, 17pp.

Marshall, C. & Chiat, S. (2003) A foot domain account of prosodically-conditioned substitutions. *Clinical Linguistics & Phonetics*, 17(8):645-657.

Marshall, C., Ebbels, S., Harris, J., & van der Lely, H.K. (2002) Investigating the impact of prosodic complexity on the speech of children with Specific Language Impairment. *UCL Working Papers in Linguistics*, 14:43-68.

Mehler, J., Jusczyk, P., Lambertz, G., Halsted, N., Bertonici, J & Amiel-Tison, C. (1988) A precursor of language acquisition in young infants. *Cognition*, 29(2):143-178. https://doi.org/10.1016/0010-0277(88)90035-2

Merrifield, W.R. & Edmondson, J.A. (1999) Palantla Chinantec. Phonetic Experiments on Nasalization, Stress, and Tone. *International Journal of American Linguistics*, 65:303-323.

Nasukawa, K. (2012) *A Unified Approach to Nasality and Voicing.* Berlin: De Gruyter Mouton. https://doi.org/10.1515/9783110910490

Ohala, J.J. (1990) Respiratory activity in speech. In: W.J. Hardcastle & A. Marchal (eds.), *Speech production and Speech Modelling.* Dodrecht: Kluwer, pp. 25-53.

Parker, S. (2011) Sonority. In: M. van Oostendorp, C.J. Ewen, E. Hume & K. Rice (eds.), *The Blackwell Companion to Phonology* (Vol. 2). Oxford: Wiley-Blackwell, pp. 1160-1184. https://doi.org/10.1002/9781444335262.wbctp0049

Pierrehumbert, J.B. (2016) Phonological Representation: Beyond Abstract versus Episodic. *Annual Review of Linguistics*, 2:33-52. https://doi.org/10.1146/annurev-linguistics-030514-125050

Ploch, S. (1999) *Nasals on my mind.* Doctoral dissertation, SOAS, University of London.

Pöchtrager, M. (2006) *The Structure of Length*. Doctoral dissertation, University of Vienna.

Pöchtrager, M. (2013) *On A.* Paper presented at the workshop on melodic representation. University College London, 12 March 2013.

Polka, L. & Bohn, O.S. (2011) Natural Referent Vowel (NRV) framework: An emerging view of early phonetic development. *Journal of Phonetics*, 39(4):467-478.

Pouplier, M. (2020) Articulatory Phonology. *Oxford Research Encyclopedia of Linguistics*. https://doi.org/10.1093/acrefore/9780199384655.013.745

Rasmus, F., Hauser, M.D., Miller, C., Morris, D. & Mehler, J. (2000) Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science*, 288(5464):349-351. https://doi.org/10.1126/science.288.5464.349

Reiss, C. (2012) Towards a bottom-up approach to phonological typology. In: A.M. Di Scuillo (ed.), *Towards a Biolinguistic Understanding of Grammar: Essays on Interfaces.* Amsterdam: John Benjamins.

Reiss, C. (2021) Substance-Free Phonology. In: S.J. Hannahs & A. Bosch (eds.), *The Routledge Handbook of Phonological Theory*. London: Routledge, pp. 145-179.

Reiss, C. & Volenec, V. (2017) Cognitive Phonetics: The Transduction of Distinctive Features at the Phonology-Phonetics Interface. *Biologinguistics*, 11(SI):251-294.

Rose, Y. (2000) *Headedness and prosodic licensing in the L1 acquisition of phonology*. Doctoral dissertation, McGill University.

Schane, S.A. (1984) The Fundamentals of Particle Phonology. *Phonology Yearbook*, 1:129-155.

Scheer, T. (2014) Spell-Out, Post-Phonological. In: E. Cyran & J. Szpyra-Kozlowska (eds.), *Crossing Phonetics-Phonology Lines.* Newcastle-upon-Tyne: Cambridge Scholars, pp. 255-275.

Scheer, T. (2020) On the lexical character of intermodular communication. *Radical: A Journal of Phonology*, 1:183-239.

Silverman, D. (2017) *A Critical Introduction to Phonology: Functional and Usage-Based Perspectives*, 2nd Edition. London: Bloomsbury Academic.

Smith, C., O'Hara, C., Rosen, E. & Smolensky, P. (2021) Emergent Gestural Scores in a Recurrent Neural Network Model of Vowel Harmony. In: *Proceedings of the Society for Computation in Linguistics*, Volume 4, pp. 61-70.

Smith, C. & O'Hara, C. (2021) Learnability of Derivationally Opaque Processes in the Gestural Harmony Model. In: *Proceedings of the Society for Computation in Linguistics*, Volume 4, pp. 396-400.

Steriade, D. (1993a) Segments, Contours and Clusters. In: A. Crochetiere et al. (eds), *Proceedings of the XVth International Congress of Linguistics*. Quebec: Université Laval, pp. 71-82.

Steriade, D. (1993b) Closure, Release and Nasal Contours. In: M.K. Huffinan, & R.A. Krkow (eds.), *Nasals, Nasalization, and the Velum* (Phonetics and Phonology 5). San Diego, CA: Academic Press, pp. 401-470.

Steriade, D. (1994) Complex Onsets as Single Segments: The Mazateco Pattern. In: J. Cole, & C. Kisseberth (eds.), *Perspectives in Phonology: Proceedings of the University of Illinois Phonology Conference*. Stanford, CA: CSLI, pp. 203-291.

Stevens, K.N. (1998) *Acoustic Phonetics.* Cambridge, MA: MIT Press.

Sweet, H. (1877) *A handbook of phonetics*. Oxford: Clarendon Press.

Traunmüller, H. (1994) Conventional, biological, and environmental factors in speech communication: a modulation theory. *Phonetica*, 51: 170-183.

Traunmüller, H. (1998) Modulation and demodulation in production, perception, and imitation of speech and bodily gestures. *Proceedings of FONETIK 98*. Stockholm University, pp. 40-43.

Trubetzkoy, N. S. (1939) *Grundzüge der Phonologie*. Göttingen: Vandenhoek & Ruprecht.

Ulfsbjorninn, S. (2021) Labiovelars and the labial-velar hypothesis: phonological headedness in Bare Element Geometry. *Glossa: a journal of general linguistics*, 6(1): 111. https://doi.org/10.16995/glossa.5718

van der Feest, S.V.H. (2007) *Building a Phonological Lexicon. The acquisition of the Dutch voicing contrast in perception and production.* Doctoral dissertation, Radboud University Nijmegen.

van der Hulst, H.G. (1988) The geometry of vocalic features. In: H.G. van der Hulst & N. Smith (eds.), *Features, segmental structure and harmony processes* (Vol. 2). Dordrecht: Foris, pp. 77-125.

van der Hulst, H.G. (1994) An introduction to Radical CV Phonology. In: S. Shore & M. Vilkuna (eds.), *SKY 1994: Yearbook of the linguistic association of Finland*. Helsinki, pp. 23-56.

van der Hulst, H.G. (2012) A framework for vowel harmony. In: B. Botma & R. Noske (eds.), *Phonological Explorations: Empirical, Theoretical and Diachronic Issues*. Berlin: Mouton de Gruyter, pp. 155-190.

van der Hulst, H.G. (2018) *Asymmetries in vowel harmony: a representational account.* Oxford: Oxford University Press.

van der Hulst, H.G. (2020) *Principles of Radical CV Phonology. A theory of segmental and syllabic structure.* Edinburgh: Edinburgh University Press.

van der Hulst (in press) A Guide to Radical CV Phonology, with special reference to tongue root and tongue body harmony. In: L. Voeltzel (ed.) *Perspectives on Element Theory*. Berlin: Mouton de Gruyter.

Vihman, M.M. & Keren-Portnoy, T. (eds.) (2013) *The emergence of phonology: Whole-word approaches and cross-linguistic evidence.* Cambridge: Cambridge University Press.

Visser-Bochane, M.I., Reijneveld, S.A., Krijnen, W.P., van der Schans, C.P. & Luinge, M.R. (2020) Identifying Milestones in Language Development for Young Children Ages 1 to 6 Years. *Academic Pediatrics*, 20(3):421-429. https://doi.org/10.1016/j.acap.2019.07.003

Weed, J.H. (2021) Religious Language. In: J. Fiester & B. Dowden (eds.) *The Internet Encyclopedia of Philosophy*. https://iep.utm.edu/rel-lang/

Werker, J.F. & Tees, R.C. (1984) Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1): 49-63.

Williamson, K. (1977) Multivalued features for consonants. *Language*, 53:843-871.

Zec, D. (1994) *Sonority Constraints on Prosodic Structure*. London: Routledge. https://doi.org/10.4324/9780429437168