

Cambridge University Press

0521825784 - Phonetically Based Phonology

Edited by Bruce Hayes, Robert Kirchner, and Donca Steriade

Excerpt

[More information](#)

1 Introduction: the phonetic bases of phonological Markedness

Bruce Hayes and Donca Steriade

If phonological systems were seen as adaptations to universal performance constraints on speaking, listening and learning to speak, what would they be like?
Lindblom (1990: 102)

1 Introduction

Our starting point is a hypothesis central to contemporary phonology: that the markedness laws characterising the typology of sound systems play a role, as grammatical constraints, in the linguistic competence of individual speakers. From this assumption, a basic question follows: How are grammars structured, if markedness laws actively function within them as elements of linguistic competence? We find the answer offered by Optimality Theory (Prince and Smolensky 1993) worth investigating: the grammatical counterparts of markedness laws are ranked and violable constraints and the latter form ‘the very substance from which grammars are built: a set of highly general constraints which, through ranking, interact to produce the elaborate particularity of individual languages’ (Prince and Smolensky 1993: 217). With qualifications, this view is adopted by many of the contributions to this volume.

The focus of our book is on a different, complementary question: Where do markedness laws come from? Why are sound systems governed by these laws and not by some conceivable others? What is the source of the individual’s knowledge of markedness-based constraints? The hypothesis shared by many writers in this volume is that phonological constraints can be rooted in *phonetic knowledge* (Kingston and Diehl 1994), the speakers’ partial understanding of the physical conditions under which speech is produced and perceived. The source of markedness constraints as components of grammar is this knowledge. The effect phonetic knowledge has on the typology of the world’s sound systems stems from the fact that certain basic conditions governing speech perception and production are necessarily shared by all languages, experienced by all speakers, and implicitly known by all. This shared knowledge leads learners to postulate independently similar constraints. The activity of similar constraints is

Cambridge University Press

0521825784 - Phonetically Based Phonology

Edited by Bruce Hayes, Robert Kirchner, and Donca Steriade

Excerpt

[More information](#)2 *The phonetic bases of phonological Markedness*

a source of systematic similarities among grammars and generates a structured phonological typology.

In this introduction, we explain why it is useful to explore the hypothesis that knowledge of markedness derives from phonetic knowledge: how one's view of markedness changes under this hypothesis and what empirical results come from this change of perspective. We outline first how research on phonetically based markedness can be beneficially explored in the framework of Optimality Theory (section 2); and how the OT search for the right constraint set can be speeded up on the view that markedness is phonetically based (sections 3 and 4). We then discuss a specific example of a phonetically based Markedness constraint that illustrates several options in mapping the facts of phonetic difficulty to the elements of grammar (section 5). In the remaining sections, we relate the general discussion of markedness to the specific contents of the book, noting that despite differences of analytical strategy or general theoretical outlook, the diverse phenomena analysed by our contributors can be viewed in a unified fashion.

2 **Phonetically based Markedness and Optimality Theory**

The idea that phonological Markedness has phonetic roots has particular antecedents in *The sound pattern of English* (Chomsky and Halle 1968), in the theory of Natural Phonology (Stampe 1973), and in the more recent work on Grounded Phonology by Archangeli and Pulleyblank (1994). Optimality Theory makes it worth returning to these issues, since it provides tools with which the questions can be addressed in novel ways. OT takes on a difficulty that held back earlier approaches to naturalness: the *what* is phonetically difficult is not the same as the *how to fix it*. In a rule-based framework, one must provide the theory with multiple fixes, all of which address the same phonetic difficulty. OT separates the problem (embodied in the Markedness constraints) from the solution; the latter is the general procedure at the core of OT, namely creation of a large candidate set by GEN, with the choice from among them determined by the relative ranking of the Markedness constraints with respect to Faithfulness and each other. As a result, OT allows the phonetic principles that drive the system to be expressed directly (Myers 1997): a constraint can embody a particular form of phonetic difficulty, with the issue of how and whether the difficulty is avoided relegated to other parts of the grammar. For a clear case of this sort, see the discussion of postnasal voicing in Pater (1999) and Kager (1999).

The separation of Markedness and Faithfulness also provides a cogent response to an ancient canard: If phonetic optimality is important, why don't sound systems contain nothing but the Jakobsonian optimal [ba]? The answer

Cambridge University Press

0521825784 - Phonetically Based Phonology

Edited by Bruce Hayes, Robert Kirchner, and Donca Steriade

Excerpt

[More information](#)

is that not all the constraints can be satisfied at once. Faithfulness and Markedness constraints conflict; and moreover, there are conflicts between different types of Markedness constraints (notably, those grounded in production vs those grounded in perception). There is no reason to expect the resolution of these conflicts to be uniform across languages. The postnasal voicing example just mentioned is a plausible case of multiple resolutions of the same difficulty.

The more direct argument for OT is that phonetically based constraints discussed here are frequently both active and violated, yielding Emergence of the Unmarked effects (McCarthy and Prince 1994) which require explicit ranking. Kirchner's, Kaun's, and Crosswhite's chapters provide extensive evidence of this type, as does a voicing example discussed below.

3 Markedness

The term *markedness* is ambiguous. It can be used in a strictly typological sense, to identify structures that are infrequently attested or systematically missing, as in *Active use of [-ATR] is marked* (Archangeli and Pulleyblank 1994: 165 and passim). The term can also refer to an element of a formal linguistic theory, as in OT, where the term *markedness* characterises a constraint type, often disambiguated by capitalisation: Markedness constraints penalise particular structures in surface forms, whereas Faithfulness constraints evaluate dimensions of similarity between specified pairs of lexically related structures, such as the underlying and surface representations.

The definition of markedness in OT is also sometimes related to the hypothesis that Markedness constraints are universal and innate. This claim is logically independent of the central tenets of OT about constraint interaction.¹ Accordingly we are free to assume that a constraint need not be universal or innate to qualify as a Markedness constraint; rather, we use the term in the purely technical sense of a constraint whose violations are evaluated solely on surface forms. We use the term *markedness law* to denote patterns found in typological data, which Markedness constraints are often meant to explain. We may add that the correspondence conditions themselves are formulated with the intention of deriving key aspects of phonological typology.²

The terms thus clarified, we turn now to the options available to phonologists who study markedness in either of these two senses.

4 Inductive and deductive approaches to the study of Markedness

Lindblom (1990: 46)³ observes that the study of distinctive features can proceed in two ways: inductively and deductively. The inductive approach in the

Cambridge University Press

0521825784 - Phonetically Based Phonology

Edited by Bruce Hayes, Robert Kirchner, and Donca Steriade

Excerpt

[More information](#)4 *The phonetic bases of phonological Markedness*

study of features is to introduce a new feature whenever the descriptive need arises. The deductive approach, for example Stevens' Quantal Theory (1989) or Lindblom's Dispersion Theory (1986), proceeds not from a question of description ('What are the features used in language?') but from a principled expectation: 'What features should we expect to find given certain assumptions about the conditions [under which] speech sounds are likely to develop?' (Lindblom and Engstrand 1989: 107). The deductive approach can thus hope to provide not only an empirically verifiable feature theory, in the form of principles from which feature sets derive, but may also yield answers to further questions, such as 'Why are the mental representations of speech sounds feature-based (and likewise segment-, syllable-, foot-based)?' These questions simply do not arise under approaches that take for granted the existence of such units and merely aim to discover in the data a basis for their classification.

The distinction between inductive and deductive approaches applies equally to research on markedness. Most attempts to discover markedness principles in phonology have proceeded, until recently, in inductive fashion: phonologists accumulate factual observations about languages and, in due course, a cluster of such observations coheres into a law. The law may be absolute ('There are no initial or final systems in which all obstruent combinations are heterogeneous with regard to voicing'; Greenberg 1978: 252), or implicational ('The presence of syllabic [h] implies the presence of syllabic fricatives'; Bell 1978: 183), or only a trend ('If a nasal vowel system is smaller than the corresponding basic vowel system, it is most often a mid vowel that is missing from the nasal vowels'; Crothers 1978: 136). But in most cases the laws originate as generalisations over known languages, not as principles explaining why these laws should be expected to hold. A set of such laws, when they survive peer review, forms a proposed theory of markedness.

The markedness questions asked in earlier typological work seem to have been those for which evidence happened to be available. We cannot exclude the possibility that a priori principles have guided the search for typological generalisations, as reported in the classic work of Trubetzkoy 1938, Jakobson 1941, Hockett 1955, and Greenberg 1978, but these guiding principles were not spelled out and cannot be reconstructed. One may ask, for instance, why the search for clustering universals (Greenberg 1978) proceeds by asking some questions (*Is there an implicational relation between initial [ln] and initial [lt]?*) but not others (*Is there an implicational relation between initial [ht] and initial [th]?*).

There is an issue of research strategy here. The number of conceivable typological observations is so vast that our results will be haphazard if we examine the data in arbitrary order. Without a general conception of what makes a possible markedness principle, there is no more reason to look into the markedness

Cambridge University Press

0521825784 - Phonetically Based Phonology

Edited by Bruce Hayes, Robert Kirchner, and Donca Steriade

Excerpt

[More information](#)

patterns of, say, initial retroflex apicals (a useful subject, as it turns out; see section 6.1) than into those of prenasal high tones (a topic whose interest remains unproven). The researcher has to take a stab in the dark. In light of this, it seems a sensible research strategy to hypothesise general principles concerning why the constraints are as they are, and let these principles determine a structured search for markedness patterns. We also see below that pursuing the deductive strategy can yield a completely different picture of markedness in several empirical domains.

The work reported in this volume proceeds deductively – as advocated by Lindblom (1990) and Ohala (1983, and much later work) – by asking at the outset variants of the following question: Are there general properties distinguishing marked from unmarked phonological structures, and, if so, what are they? Earlier work in phonetics⁴ and phonology⁵ suggests that a connection can be found between constraints governing the production and perception of speech and markedness patterns. Certain processes (cluster simplification, place assimilation, lenition, vowel reduction, tonal neutralisation) appear to be triggered by demands of articulatory simplification, while the specific contexts targeted by simplification (e.g. the direction of place assimilation, the segment types it tends to target) are frequently attributable to perceptual factors.

Deductive research on phonological markedness starts from the assumption that markedness laws obtain across languages not because they reflect structural properties of the language faculty, irreducible to non-linguistic factors, but rather because they stem from speakers' shared knowledge of the factors that affect speech communication by impeding articulation, perception, or lexical access. Consider the case discussed below, that of the cross-linguistic dispreference for voiced geminates. The deductive strategy starts from the assumption that this dispreference cannot reflect an innate constraint that specifically and arbitrarily bans [b: d: g:], but must be based on knowledge accessible to individual speakers of the factors that might interfere with the production and perception of voicing. This knowledge and its connection to the grammar have then to be spelled out.

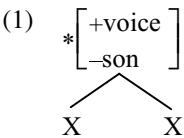
Is the deductive strategy reductionist? Clearly so, but in specific respects. The research presented here bears only on the possibility of systematically deducing the contents of phonological constraints from knowledge of grammar-external factors. This is not the same as deducing the grammar itself: on the contrary, structural properties of the grammar may well filter phonetic knowledge and limit the ways it is mapped onto grammatical statements, as suggested by Gordon (chapter 9) and summarised below (section 5.7). Further, none of the contributions addresses systematically the nature of phonological representations or deduces *their* properties from extra-grammatical factors or discusses whether such reduction is feasible (Gafos 1999). The same

goes for the nature of constraint interaction. On the issue of external grounding for all of these components, see Pierrehumbert’s overview (2000), and the discussion of representations and constraint interactions by Flemming (2001).

5 Markedness from phonetics: a constraint and its phonetic basis

We now examine a specific example of the deductive strategy. This section introduces a markedness scale and points out its sources in the aerodynamics of speech.

In the phonological analysis of a number of languages, a constraint is needed that penalises voiced obstruent geminates; (1) is a first approximation.



Variants of (1) are active in Ancient Greek (Lupas 1972), Ossetic (Abaev 1964), Nubian (Bell 1971), Lebanese Neo-Syrian (Ohala 1983), Tamil (Rajaram 1972), Yakut (Krueger 1962), Limbu (van Driem 1987), Seleyarese and Buginese (Podesva 2000), and Japanese (Ito and Mester 1995). No language known to us bans just the *voiceless* geminates.⁶ The constraint in (1) thus has a typological counterpart, the implicational law in (2):

- (2) The presence of a voiced obstruent geminate in a given language implies, in any context, that of the corresponding voiceless geminate.⁷

If a Markedness constraint like (1) reflects, directly or not, an implicational law like (2), then we must consider the possibility that the constraint is universal, in the sense of being potentially active in any grammar. In the next section we explore the hypothesis that some version of (1) is universal in the sense of being inferable from generally available phonetic knowledge.

5.1 From phonetics to grammar

As indicated earlier, we assume that constraints may be universal without being innate (cf. Lindblom 1990; Donegan 1993; Boersma 1998; Hayes 1999). We view Universal Grammar (UG) primarily as a set of abstract analytical predispositions that allow learners to induce grammars from the raw facts of speech, and not as a – dauntingly large – collection of a priori constraints. The project then is to understand how constraints like (1) are induced from evidence

Cambridge University Press

0521825784 - Phonetically Based Phonology

Edited by Bruce Hayes, Robert Kirchner, and Donca Steriade

Excerpt

[More information](#)

about the conditions under which voicing is perceived and produced and what form they take if they are so induced. It is useful here to make the four-way distinction shown below:

- (3) a. Facts of phonetic difficulty
- b. Speakers' implicit knowledge of the facts in (a)
- c. Grammatical constraints induced from the knowledge in (b)
- d. Sound patterns reflecting the activity of the constraints in (c)

Facts about phonetic difficulty (3a) and sound patterns (3d) are, in principle, accessible; they are obtainable from experiment, vocal tract modelling, and descriptive phonological work. But the precise contents of (3b) and (3c) have to be guessed at. We see no alternative to drawing these distinctions and making some inferences.

With Prince and Smolensky (1993), we assume that constraint organisation, (3c), reflects transparently the structure of markedness scales, (3b).⁸ We also assume that the correspondence between the facts of phonetic difficulty (3a) and the markedness scales (3b) is necessarily indirect: the crucial question is how indirect.

The markedness scales phonologists have mainly relied on so far do not, in their current formulations, explicitly relate to scales of articulatory or perceptual difficulty. Examples are: (a) the nucleus goodness scale in Prince and Smolensky 1993; (b) a place optimality scale like (*{Labial, Dorsal}* < *Coronal* < *Pharyngeal*), where < denotes 'worse than'; Lombardi (in press); and (c) syllabic markedness scales like *CVCC*, *CCVC* < *CVC* < *CV*. This may reflect the fact that there is no connection between Markedness constraints and phonetic scales or that the exact ways in which phonetic scales map onto phonological markedness has no consequences for the functioning of the phonology. However, the research reported in this book as well as in earlier work indicates that there is often evidence for a much closer connection.

In the next subsections we summarise the articulatory difficulties involved in sustaining vocal cord vibration in different obstruents and consider ways in which speakers can encode knowledge of these difficulties in markedness scales. Our point will be that among several types of mapping (3a) onto (3b)–(3c), a more direct one yields more predictive and more successful models of grammar.

5.2 Aerodynamics of voicing

Phonetic studies (Ohala and Riordan 1979; Westbury 1979; Westbury and Keating 1986) have located the rationale for the markedness law in (2) in the aerodynamics of voicing production:

Cambridge University Press

0521825784 - Phonetically Based Phonology

Edited by Bruce Hayes, Robert Kirchner, and Donca Steriade

Excerpt

[More information](#)8 *The phonetic bases of phonological Markedness*

- (4) a. Voicing requires airflow across the glottis.
- b. In obstruents, the supraglottal airflow is not freely vented to the outside world.

For these reasons, active oral tract expansion (for example, by tongue root advancement or larynx lowering) is necessary to maintain airflow in an obstruent. These manoeuvres cannot be continued indefinitely or controlled tightly. It is therefore more difficult to sustain production of voicing in long obstruents. The difficulty is directly witnessed in languages like Ossetic, whose speakers attempt to maintain a voicing distinction in long obstruents but nonetheless lose 'part or all of the voiced quality' (Abaev 1964: 9) in [b: d: g:]. No comparable difficulty exists in sustaining voicelessness in [p: t: k:] or voicing in long sonorants, while the problem of maintaining voicing in singleton stops is necessarily one of shorter duration. So far the discussion motivates a simple voicing difficulty scale of the form $D_i: < D_i$ where $D_i:$ is a geminate voiced obstruent, and D_i is the corresponding singleton.

Consider now a second factor that influences phonetic difficulty in obstruents, namely place of articulation. As Ohala and Riordan (1979) observe, the size of the cavity behind the oral constriction affects the aerodynamics of voicing. The time interval from the onset of stop closure to the point where passive devoicing will set in varies with the site of the oral constriction: in one experiment, voicing was observed to continue in [b] for 82 ms, but for only 63 and 52 ms respectively in [d] and [g]. This is because the larger cavity behind the lips offers more compliant tissue, which allows the cavity to continue for a longer time to expand passively in response to airflow. A consequence of this is the known asymmetry (Gamkrelidze 1978) between voicing markedness in singleton bilabials as against alveolars and velars: [g] implies [d] which implies [b].⁹ This asymmetry holds, according to Ohala (1983), among voiced geminates as well: a geminate [b:]'s duration will certainly exceed 82 ms, and thus some active expansion of the oral tract must be taking place, just as for [d:] and [g:]. But a difference in ease of voicing maintenance persists among the voiced geminates, because there are more options for expansion available in front than in back articulations.

5.3 *From aerodynamics to markedness to constraints*

There are then at least two sources of articulatory (and indirectly perceptual) difficulty in maintaining voicing: the duration of oral closure and the size of the cavity behind the oral constriction. Phonologically, these are completely different, yet at the level of phonetic difficulty, they are essentially the same thing: in both [g] (a singleton with a small cavity behind the constriction) and [b:] (a geminate with a large cavity) there is difficulty in maintaining voicing

Cambridge University Press

0521825784 - Phonetically Based Phonology

Edited by Bruce Hayes, Robert Kirchner, and Donca Steriade

Excerpt

[More information](#)

past the point where passive devoicing normally sets in. Thus at the phonetic level we can posit a single scale of difficulty that includes both singletons and geminates.

(5) * $[+voice]$: $\{g: < d: < b: < g < d < b\}$

The scales we formulate henceforth distinguish a shared target property – $[+voice]$ in (5) – and the set of contexts in which this property is realised with greater or lesser difficulty: (5) states that the $[+voice]$ feature is hardest to realise in $[g:]$, next hardest in $[d:]$, and so on, and easiest to realise in $[b]$.

The scale in (5) identifies $[b:]$, the best voiced geminate, as harder to voice than short $[g]$, the worst singleton. The difference between a singleton and a geminate consonant is typically much more than the 30 ms that separate the onset of passive devoicing in $[b]$ vs $[g]$ (Lehiste 1970; Smith 1992). Thus the difficulty involved in sustaining voicing should be far more extreme for any geminate obstruent than it would be for any voiced singleton: (5) reflects this point.

If knowledge about the difficulty of sustaining voicing in obstruents resembles the scale in (5), then its grammatical counterpart cannot be a single constraint; nor can the constraints against voiced geminates remain unrelated to those against voicing in singletons. This is because the voicing difficulty in $[g: d: b:]$ is of the same type – if not of the same magnitude – as that involved in $[g d b]$. We need a constraint set that reflects the whole scale in (5), not just its upper region. The more general point is that knowledge of markedness, when viewed as phonetic knowledge, generates constraint families and rankings whose structure reflects a broader map of phonetic difficulty, as the learner understands it, rather than isolated points and relations on this map.

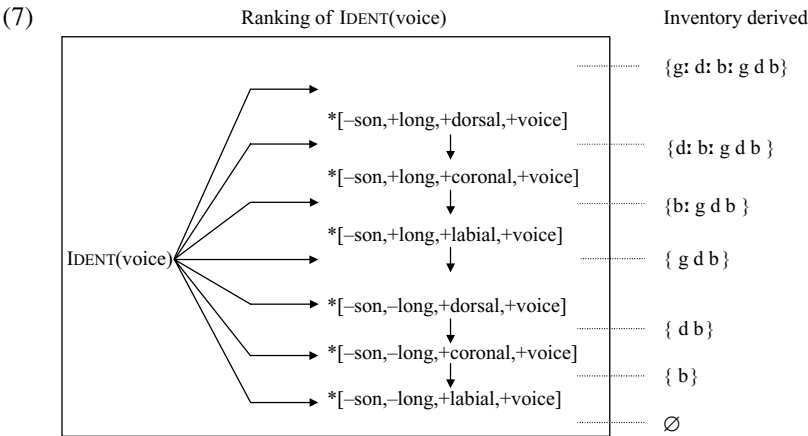
As a specific proposal to this end, consider the set of Markedness constraints in (6). These constraints are assumed to be ranked a priori, according to the phonetic difficulty of the segments that they ban (but see fn. 8 above on the issue of fixed rankings).

- (6) a. * $[-son, +long, +dorsal, +voice]$ ‘no voiced long dorsal obstruents’ \gg
 b. * $[-son, +long, +coronal, +voice]$ ‘no voiced long coronal obstruents’ \gg
 c. * $[-son, +long, +labial, +voice]$ ‘no voiced long labial obstruents’ \gg
 d. * $[-son, -long, +dorsal, +voice]$ ‘no voiced short dorsal obstruents’ \gg
 e. * $[-son, -long, +coronal, +voice]$ ‘no voiced short coronal obstruents’ \gg
 f. * $[-son, -long, +labial, +voice]$ ‘no voiced short labial obstruents’

If the rankings in (6) are fixed, then the relative ranking of this constraint family with respect to the Faithfulness constraint IDENT(voice) determines the

10 *The phonetic bases of phonological Markedness*

inventory of voiced obstruents, as shown in (7):



An interesting aspect of the constraint set in (6) is that it uses very fine categories, each embodying information about both place and length. Phonologists characteristically judge that constraints are based on rather broader categories. One thus could imagine a more modular characterisation of voicing markedness, as in (8):

- (8) a. *[−son, +dorsal, +voice] ‘no voiced dorsal obstruents’ >>
 *[−son, +coronal, +voice] ‘no voiced coronal obstruents’ >>
 *[−son, +labial, +voice] ‘no voiced labial obstruents’
 b. *[−son, +long, +voice] ‘no long voiced obstruents’ >>
 *[−son, −long, +voice] ‘no short voiced obstruents’

The constraints in (8) are simpler than those of (6), and involve separate chains of a priori rankings for the dimensions of place and length. As a result, this constraint set is silent on how closure duration and cavity size interact – that is, on the [b:] vs [g] comparison – and thus makes rather different predictions. Notably, we find that in ranking IDENT(voice) amid the chains of (8) (interleaving the chains freely), we cannot derive the inventories for two of the crucial cutoff points in (5): {b: g d b} (forbidding *[d:] and harder) and {d: b: g d b} (forbidding just *[g:]).¹⁰

5.4 *From scales to sound patterns: some language data*

The special possibilities implied by (6) (i.e., the constraint set that embodies a unitary scale of voicing difficulty) are confirmed by examples from real