1 *What is markedness?*

1.1 Aims

This book is about the concept of 'markedness' in human language. Certain structures are often avoided while others are generated; the avoided structures are called 'marked' while the generated ones are 'unmarked'. As an example, many languages have processes which turn the phonological segment /k/ into [ʔ]. In a sense, this process avoids [k] and favours [ʔ], so [k] (or some component of [k]) can be said to be more marked than (some component of) [ʔ].

The focus here is on phonology, but the general conclusions and theoretical devices extend to other linguistic domains as well. While the leading ideas proposed here are applicable to a wide range of theories of linguistic Competence, they are particularly suited to formal expression within Optimality Theory (Prince & Smolensky 1993).

Markedness has a very long history in phonological theory (Trubetzkoy 1931, 1939; Jakobson 1941; Chomsky & Halle 1968:ch. 9; Stampe 1972; Prince & Smolensky 1993; see Battistella 1990, 1996 for an overview). Even so, an important question remains: Which processes are sensitive to markedness and which ones are not? While many phenomena have been proposed as being markedness diagnostics, there has as yet been no explanation of which particular phenomena should show markedness effects. This book presents a theory of markedness that answers this question. Behind the theory are four leading ideas, summarized in (1).

(1) (a) *Competence markedness*
Markedness is part of grammatical Competence (i-language). Markedness in Competence is distinct from sometimes apparently similar Performance-related phenomena.

(b) *Preservation of the Marked*
There is grammatical pressure to preserve marked elements. If *x* is more marked than *y*, *x* can be unaffected by a process while *y* is forced to undergo it.

(c) *Markedness conflation*
Markedness distinctions can be conflated (i.e. ignored), but never reversed. If a language treats *x* as more marked than *y*, another language can treat *x* and *y* as being equally marked; however, no language will treat *y* as less marked than *x* (caveat: subject to (d) below).

(d) *Hierarchy conflict*
Markedness hierarchies can conflict: one hierarchy may favour *x* over *y* while another favours *y* over *x*.

The following chapters will argue that there are formal objects called 'markedness hierarchies', following much previous work (markedness hierarchies are enclosed by | |, and the symbol ⟩ means 'is more marked than'). For example, the hierarchy for major place of articulation (PoA) is given in (2), building on work by Lombardi (1995, 1998) and others.

(2) Major place of articulation markedness hierarchy
| dorsal ⟩ labial ⟩ coronal ⟩ glottal |

The PoA hierarchy is related to a set of output constraints in (3) (building on work by Prince 1997a,b,c,d, 1998, 1999; de Lacy 2000a, 2002a, 2004) and a set of faithfulness constraints in (4). The PoAs 'dors', 'lab', 'cor', and 'gl' are abbreviations of dorsal, labial, coronal, and glottal respectively.

(3) Output constraints for major place of articulation

(a) *{dors}	'Assign a violation for each [dorsal] feature.'
(b) *{dors,lab}	'Assign a violation for each [dorsal] and each [labial] feature.'
(c) *{dors,lab,cor}	'Assign a violation for each [dorsal], each [labial], and each [coronal] feature.'
(d) *{dors,lab,cor,gl}	'Assign a violation for each [dorsal], each [labial], each [coronal], and each [glottal] feature.'

Together, the output constraints in (3) express the PoA hierarchy. For example, dorsals are more marked than other PoAs, and *{dorsal} is violated by dorsals like [k] but not by other segments like [p], [t], and [ʔ]. Labials are never less marked than coronals, so every constraint that coronals violate is also violated by labials (i.e. *{dor,lab,cor}, *{dors,lab,cor,gl}). In this way relative markedness relations are translated into violations of the PoA output constraints.

The constraints in (3) differ from Prince & Smolensky's (1993) theory of markedness in not imposing any universally fixed rankings – the constraints can be ranked in any way. Fixed ranking is shown to have undesirable empirical consequences in §1.3.3 and §5.2.2.

(4) Faithfulness constraints for major place of articulation

IDENT {dors}	'If x is dorsal, then x' has the same PoA as x.'
IDENT {dors,lab}	'If x is dorsal or labial, then x' has the same PoA as x.'
IDENT {dors,lab,cor}	'If x is dorsal, labial or coronal, then x' has the same PoA as x.'
IDENT {dors,lab,cor,gl}	'If x is dorsal, labial, coronal, or glottal, then x' has the same PoA as x.'

• Segment x' corresponds to segment x

Because the feature dorsal is the most marked PoA, there are specific constraints that preserve it and no other. The effect is that dorsals can be exempted from undergoing a PoA-altering process like neutralization or assimilation while other PoAs undergo it. Labial is the next most marked PoA, so there is a constraint IDENT {dors,lab} that preserves it and the more marked PoA dorsal, and so on down through the PoA hierarchy.

To make one thing clear, the theory presented here is not a necessary consequence of Optimality Theory's formalism. The OT framework is not committed to any particular theory of markedness, or even to expressing markedness at all. For example, OT could in principle allow freely rankable *{dors}, *{lab}, *{cor}, and *{gl}, so disregarding PoA markedness. OT's relationship to the theory presented here is analogous to Chomsky & Halle's (1968) formalism and the theory of markedness in their chapter 9: the theory presented here does not derive from fundamental properties of the OT framework, but rather imposes restrictions on its constraint component and feature representation.

As a terminological note, confusion could arise because 'markedness constraint' is used in OT to mean a constraint that is violated if a particular structure is present in an output candidate (Prince & Smolensky 1993; McCarthy 2001b:14). Because such constraints do not necessarily express markedness relations, I hereby rename them 'output constraints' for the rest of this book.

To be clear from the outset, this book is not a history of markedness (cf. Battistella 1990, 1996; Rice to appear); its aim is to present a markedness theory. Of course, a great deal of work on markedness has preceded this book, including the Prague School's work (e.g. Trubetzkoy 1939; Jakobson 1941, 1949a,b, 1978), Greenberg (1966, 1975, 1978), Natural Phonology (Stampe 1972), Kean (1975), Cairns & Feinstein (1982), and work in Underspecification Theory (e.g. Kiparsky 1982; Archangeli 1984, 1988). The formalism owes much to work in OT: the proposals take from Prince & Smolensky (1993) the insight that markedness relations can be formally expressed by constraints, they adopt Smolensky's (1994) proposal that markedness asymmetries can be explained through constraint interaction, and they build on Prince's (1997a,b,c,d, 1998, 1999) and

my previous work (de Lacy 2002a, 2004) in using constraint forms to express markedness hierarchies (cf. Prince & Smolensky 1993). All of these theories will be discussed when they shed light on the theory; explicit comparison is provided in chapter 8.

It is also not the aim of this book to identify all the markedness hierarchies that exist; instead, a methodology for discovering them is provided. In a related vein, it is not the aim to determine whether markedness hierarchies are entirely innate or derive from functional factors (Chomsky & Lasnik 1977:435ff.; Lasnik 1990; Bermúdez-Otero & Börjars 2005 cf. articles in Hayes, Kirchner & Steriade 2004; Bybee 2001:194ff.). Instead, the purpose of this book is to present a theory that explains why the Competence mechanisms specific to phonology exhibit markedness effects.

So, what is markedness? Before giving an answer it is necessary to identify the phenomena that a theory of markedness must explain, then the mechanisms that provide the explanation must be identified. To do this, §1.2 identifies three major challenges for a theory of markedness. The solutions are given in §1.3; in effect they identify the empirical scope of the theory. The theoretical proposals are also given in brief; they are used to identify valid and invalid diagnostics for markedness in §1.4. With the necessary background in place, §1.5 provides a definition of markedness and outlines how the theory is developed in the rest of the book.

Markedness hierarchies only refer to features and their values, not to segments as a whole. However, for convenience I will occasionally refer to 'unmarked segments' instead of saying 'segments that have some particular unmarked feature value', and make statements such as '[k] is very marked' instead of 'some feature value in [k] is very marked'.

1.2 Challenges for markedness

The term 'markedness' has been used to refer to many different linguistic concepts. One of the most common uses relates to observations about asymmetries in the output, triggers, and undergoers of certain processes.[1] 'Unmarked' elements can be the sole output of processes, fail to trigger alternations, and undergo processes alone. In contrast, 'marked' elements are rarely the output, are often the only triggering elements, and are often exempt from undergoing processes.

1 In OT, the term 'process' refers to a situation involving an unfaithful input→output mapping.

For example, (2) expressed the claim that glottals and coronals are commonly accepted as having a relatively unmarked Major place of articulation while labials and dorsals are relatively more marked. The output of consonant epenthesis is always a coronal (e.g. [t n ɹ]) or glottal ([ʔ h]), and never a labial or dorsal (e.g. [p f m], [k x ŋ]) (§3.2; Lombardi 2002). Similarly, if coronal consonants force a preceding segment to assimilate, then so do labials and dorsals (i.e. if /mt/→[nt] then /mk/ must become [ŋk]). However, the reverse is not true: if dorsals trigger assimilation, then coronals do not have to do so: /mk/→[ŋk] does not imply that /mt/→[nt] (Mohanan 1993; Jun 1995; de Lacy 2002a:ch. 7). So, coronals and glottals behave differently from labials and dorsals in many different situations and languages. Therefore it seems reasonable to appeal to an overarching principle that treats glottals and coronals differently from labials and dorsals – i.e. markedness. Glottals and coronals can be said to be 'unmarked' relative to labials and dorsals.

As in a great deal of previous work, markedness relations can be expressed as a hierarchy like the one in (2). 'Dorsal' is the most marked element and 'glottal' is least marked; evidence for the hierarchy's details will be provided in the following chapters (esp. §2.2.1). Elements are marked or unmarked relative to some other element on the same hierarchy. So, dorsals are relatively marked in comparison to coronals. The terms 'marked' or 'unmarked' have often been used as if referring to an absolute state; however, no segment or prosodic structure is ever 'not marked' – it is only comparatively less marked than some other segment or structure. There are many markedness hierarchies, including ones for voice, sonority, and tone; markedness relations between prosodic structures have also been identified, but are not the focus of this book. Of course, a pressing question is 'How do we discover markedness hierarchies?' This question can only be answered after a theory of markedness effects is introduced and used to identify valid diagnostics for markedness. The theory will be discussed below and valid diagnostics identified in §1.4.

A central issue for markedness is whether it is a formal concept or merely an informal descriptive term. In other words, is it necessary to employ special Competence mechanisms to account for the phonological asymmetries observed in natural language? The proposals in this book take the view that there are i-language devices that produce markedness effects: there are formal objects that are the equivalent of markedness hierarchies and formal mechanisms that relate constraints to those hierarchies.

However, there is a great deal of disagreement and even scepticism about markedness. An alternative to having a Competence account is to allow

the phonological and syntactic components to produce virtually any output. External pressures like restrictions on language learning are then used to explain why some outputs are infrequent or non-existent (e.g. Hale & Reiss 2000; Blevins 2004). §8.3.2 argues that this approach is not adequate in accounting for all markedness phenomena.

Those who agree that markedness asymmetries require a Competence-based explanation often disagree on the details. For example, Rice (1996) proposes that velars can be less marked than coronals, while they are seen as more marked here and in Paradis & Prunet 1990, Lombardi 2002, and here. Views also differ as to which phenomena give insight into markedness (e.g. §1.4; Greenberg 1966; Rice 1999a,b, 2006).

A great deal of the scepticism about markedness and the variation in what is considered unmarked seems to be due to three apparent problems: (a) some markedness diagnostics do not work all the time, (b) *marked* elements are favoured for some phenomena, and (c) markedness distinctions can be ignored. The following three subsections expand on each of these. §1.3 provides solutions and shows how they give insight into a formal theory of markedness.

1.2.1 Some markedness diagnostics do not work all the time

A significant difficulty in providing a comprehensive account of markedness is that many processes do not treat categories in an asymmetrical way; they seem to ignore markedness hierarchies.

For example, there are no height asymmetries in vowel epenthesis. Epenthetic vowels may be high [i ɨ], mid [e ɛ ə], or low [a] (§7.2; Lombardi 2003). The only asymmetry relates to roundness: [round] vowels cannot be epenthetic (factoring out interfering processes like round harmony – §7.2.5).

There are almost no implicational universals for segmental inventories; almost every possible set of segments is attested (§3.3). The term 'inventory' is used here to refer to the output segments found in a language. For example, Hawaiian and Yellowknife Chipewyan have the voiceless stops [k p], which are highly marked in PoA, and the relatively unmarked [ʔ], but have no intermediately marked coronal [t] (Elbert & Pukui 1979; Haas 1968 respectively). In contrast, Tahitian has [p t ʔ] and no [k] (Coppenrath & Prévost 1974), Ayutla Mixtec has only [k t ʔ] in native words and no [p] (Pankratz & Pike 1967), and Māori has [k p t] but no [ʔ] (Bauer 1993).

Similarly, there are no implicational universals relating to the undergoers of assimilation; both marked and unmarked segments can assimilate while other segments do not (§4.3). For example, only coronals undergo assimilation in

Catalan; labials and dorsals do not (§4.3.1; Mascaró 1976; Wheeler 2005a). In contrast, only labials and dorsals undergo assimilation in Sri Lankan Portuguese creole while coronals do not (§4.3.2; Smith 1978; Hume & Tserdanelis 2002).

There are also no asymmetries relating to the output of segment coalescence: if two segments are fused into one the resulting segment may retain either the marked or the unmarked value of the input segments (§6.3). For example, coalescence of /bʰ/ and /t/ in Pāli yields [dːʰ] – an output that preserves the unmarked coronal PoA of the /t/ (e.g. /labʰ-ta/ → [ladːʰa] 'take' {passive perfect participle}). In contrast, Pāli's coalescence of /k/, /ʃ/, and [t] results in preservation of the more marked PoA: [kːʰ] (e.g. /a-sakː-ʃ-ti/ → [asakːʰi] {aorist-'be able'-aorist-3sg}).

The proposal that vowel epenthesis, inventories, undergoers of assimilation, and coalescence do not show markedness effects disagrees with a number of previous claims (e.g. de Haas 1988; Paradis & Prunet 1991a; Mohanan 1993; Jun 1995; Lombardi 2003). Evidence to support the view here is provided in the following chapters (see the sections cited above for specific discussion).

So, does the lack of markedness sensitivity in the phenomena just mentioned mean that there is no such thing as markedness? This question broaches a much broader issue: which phenomena give insight into markedness and which do not? Only once this latter question is answered is it possible to address the former. Consequently, this book aims to identify the phenomena that can be expected to show markedness effects, and – more importantly – to provide principles that can be used to determine whether a particular phenomenon will exhibit markedness asymmetries. To put it slightly differently, this book will provide a theory in which the phenomena listed above are insensitive to markedness distinctions. Before discussing the solution, though, there are two more challenges to markedness to consider.

1.2.2 Marked elements are favoured

Another apparent difficulty for markedness is that occasionally *less* marked elements can be eliminated while *more* marked elements remain. This situation is apparently contrary to expectations: the traditional notion behind markedness is that grammars seek to eliminate highly marked structures – i.e. markedness reduction.

The Nepalese language Yamphu provides a relevant example of preservation of marked place of articulation (see §4.2 for details). The coronal stop /t/ debuccalizes to [ʔ] in codas while the more marked /k/ and /p/ remain unchanged. The first three forms in (5a) have an underlying /t/, the second two have an underlying

geminate /tː/. The data are from Rutgers 1998. In an unrelated process, voiceless stops become voiced inter-vocalically (e.g. /hæːt-u-ŋ/ → [hæːduŋ]).

(5) Yamphu /t(ː)/-debuccalization (Rutgers 1998)

(i) /t/ → [ʔ], /tː/ → [ʔ]
- (a) [namːiʔ] 'daughter-in-law'
  cf. [namːid-æʔ] {instrumental/ergative}
- (b) [tʰeʔ-nani] 'I don't lift you(sg.)'
  cf. [tʰed-a] 'I lifted him'
- (c) [triʔ-ma] 'contrary+{infin.}'
  cf. [kap-trid-u] 'he has (unexpectedly)'
- (d) [imːeʔ-na] 'I had caused to buy you(sg.)'
  cf. [imːetː-uŋ] 'I caused to buy him'
- (e) [siʔ-ma] 'hit+{infin.}'
  cf. [sitː-a] 'hit+past', [sitː-iŋ] 'hit+exp.'

(ii) /p/→[p]
- (a) [kʰap] 'language'
- (b) [kep-ma] 'stick+{infin.}'

(iii) /k/→[k]
- (a) [æʔlik] 'bendy'
- (b) [kʰaːk-pa] 'scrape one's throat+perform act'
- (c) [hæk-ma] 'cut with a knife or sickle+{infin.}'
- (d) [aktok] 'like that'

(iv) /ʔ/ →[ʔ]
- (a) [asiʔ] 'previously' cf. [asi.ʔ-em-ba] 'before'
- (b) [jiːw-æʔ-mu] 'river-poss.-down' cf. [kaniŋ-æʔ-æ] 'we-poss.'

The challenge raised by Yamphu is why the more marked dorsals and labials are exempt from debuccalization while the less marked coronals are not.

Similar situations are found in assimilation and coalescence. Chapter 4 describes cases where coronals undergo assimilation but dorsals and labials do not. A famous case is Catalan, in which the coronal /n/ assimilates while the labial /m/ and dorsal /ŋ/ do not (§4.3). There are also many cases of segment coalescence where the most marked feature value is retained. For example, when Attic Greek vowels coalesce the resulting output vowel keeps the marked [+round] feature: /mistʰo+ɛːte/ → [mistʰɔːte] 'you may hire out', *[mistʰɛːte] (de Haas 1988; de Lacy 2002a:ch. 8).

On the other hand, all of the processes just cited also allow the most marked element to be eliminated while the least marked element is retained. For example, the highly marked dorsal /k/ is eliminated in Standard Malay codas while the less marked labial /p/ and coronal /t/ are faithfully preserved §3.2. Sri Lankan

Portuguese creole is the exact opposite to Catalan: labials [m] and dorsals [ŋ] assimilate while coronals [n] do not (§4.3.2; Smith 1978; Hume & Tserdanelis 2002). Finally, the unmarked value can emerge in coalescence, as noted for Pāli in §1.2.1 (also §6.3).

The challenge that these phenomena raise for markedness is that marked and unmarked elements seem to be treated in the same way. Both marked and unmarked features can be eliminated (through neutralization, assimilation, and so on), and both marked and unmarked features can survive in coalescence. Lack of sensitivity to markedness makes it seem that there is no use for a marked–unmarked distinction. However, the proposals in this book will argue that markedness is at the heart of phenomena like those above. Marked elements can be selectively retained because there is a principle of Preservation of the Marked: highly marked elements can be preserved while less marked ones are not. However, such preservation only applies in phenomena for which preservation of the input is possible. It is noteworthy that there is no phenomenon which only ever produces marked elements (cf. de Haas 1988); this fact follows from the theory presented here. Before discussing the proposal in more detail, one other challenge to markedness needs to be mentioned.

1.2.3 Markedness distinctions are conflated

The idea that markedness relations are universally invariant faces a challenge in apparent markedness reversals: for example, some languages treat coronals as more marked than glottals, while others treat them in the opposite way. Another challenge relates to markedness distinctions in particular hierarchies: while some languages treat labials as more marked than coronals, others seem to conflate or ignore this markedness distinction. Markedness reversal and conflation can also occur in different phenomena within the same language.

A markedness reversal can be seen in neutralization. While input glottals become coronal in Korean codas, coronals become glottal in Yamphu (§3.3). The following chapters will argue that markedness reversal is due to conflicting markedness hierarchies. There is no meta-restriction that all markedness hierarchies have to be consistent as to which segment or feature value is the least marked. So while glottals are less marked than coronals on the PoA hierarchy, coronals are less marked than glottals in other hierarchies. Languages, and even different processes in the same language, may differ as to which hierarchy they favour; the result is that there is some variation as to which segment is the 'least marked'. It is important to point out that markedness hierarchies do not conflict over everything – there are some feature values and segments that are

not least marked on *any* hierarchy; for this reason markedness effects are visible in natural language.

Markedness conflation is an entirely different issue. To give an example, the PoA markedness hierarchy has dorsals as more marked than coronals, and coronals more marked than glottals. Markedness can therefore make sense out of a situation in which dorsals and coronals become glottals, as in Kashaya codas: e.g. /mihjoq/ → [mihjoʔ] 'woodrat', /sulemat/ → [sulemaʔ] 'rope' (§5.2). However, glottals are blocked from appearing before other glottals. If the coda neutralization process is indeed one of simply reducing markedness, it is reasonable to expect two things in pre-glottal environments: (a) coronals should remain coronals, as there is nothing less marked for them to become, but (b) dorsals should become coronals, as this would achieve at least partial markedness reduction. However, in Kashaya dorsals remain dorsals in this situation: /mihjoq+ʔ/ → [mihjoqʔ], *[mihjotʔ] 'it's a woodrat'. The reason for dorsal survival is that Kashaya coda neutralization treats dorsals and coronals as being equally marked, so there is no motivation for either to change in situations where they cannot become glottals.

There are many other situations where markedness distinctions on a particular hierarchy are ignored or conflated. Examples can be found in languages where stress is sensitive to sonority, discussed in §5.3.2.

Markedness conflation seems to pose a problem because it suggests that there is no such thing as universality in markedness hierarchies. Nevertheless, this book will argue that markedness hierarchies are universal, as are the theoretical devices that refer to them. Markedness distinctions may be conflated in a principled way within the same hierarchy. Such collapse follows from the form of the theoretical devices that refer to markedness hierarchies. Crucially, it is not possible for a hierarchy to be reversed, showing that it is not possible to get rid of the concept of markedness altogether.

1.3 Solutions

Part of the solution to the apparent problems identified in §1.2 is that 'markedness' has been used to refer to several distinct concepts. One use refers to linguistic Competence, or 'i-language' mechanisms (Chomsky 1986). The other relates to Performance. This book aims to provide an account of markedness in a theory of linguistic Competence, so all asymmetrical effects that are not due to Competence mechanisms must be eliminated from the discussion. This point is developed in §1.3.1.