CHAPTER 2
GRAMMATICAL AND PHONOLOGICAL MARKEDNESS

0. Introduction

This chapter is devoted to a review of the notion ‘markedness’ in morpho-syntax (‘grammatical markedness’) and phonology (‘phonological markedness’). According to the main claim of this dissertation, a connection exists between these two kinds of markedness in the sense that marked grammatical categories sponsor marked phonological material to an extent greater or equal than in the unmarked grammatical category.

I start by discussing grammatical markedness in §1, where the most important criteria for grammatical markedness are presented, with emphasis on frequency of occurrence. §2 looks into the expression of marked versus unmarked grammatical categories and refutes relative phonological size as a reliable property of the respective output forms. In §3 the most important criteria for phonological markedness are laid out. §4 tests the frequency of occurrence – phonological markedness correlation expressed by the Marked in the Marked (MIM) generalization on open versus closed class items and roots versus affixes. §5 is devoted to a functional grounding account of MIM effects introduced in §4. §6 is a brief recapitulation of the issues discussed in the chapter.
1. Criteria for grammatical markedness

1.1 Structural coding

An important criterion for distinguishing the marked from the unmarked member of an opposition involving grammatical (morpho-syntactic) categories is structural coding, defined in (15) below following Croft (2003):

(15)  Structural coding

The marked value of a grammatical category will be expressed by at least as many morphemes as the unmarked value of that category. (Croft 2003:92)

For example, it may often be the case that the unmarked value of a category has a null grammatical morpheme as the specific marker, so the output for the category is equivalent to the bare root. The marked category will have either a null affix mark, or a non-null one, which attaches to the root. Either way, there are at least as many phonologically overt morphemes in the marked value of the category as in the unmarked one.

When applied to the category of number, structural coding allows for the following situations (16):
(16) (a) zero coding for the unmarked value (Singular) and overt coding for the marked value (Plural), as in English:

dog-Øs vs. dog-spl.

(b) zero coding for both values, Singular and Plural, as in Minor Malabri (Croft 2003:89):

?ɛɛw ‘child/children’

(c) overt coding for both the singular and plural, as in Zulu (Croft 2003:88):

umu-ntu ‘Singular-person’
aba-ntu ‘Plural-person’

A fourth, conceivable case is represented by languages that overtly code the Singular but have a zero Plural. The markedness criterion of structural coding predicts that this type is non-existent⁷. The predictions made on the basis of structural coding are illustrated in (17):

(17) Table 2 Overt versus null inflection in Number

<table>
<thead>
<tr>
<th>Null singular morpheme</th>
<th>Overt singular morpheme</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overt plural morpheme</strong></td>
<td>English</td>
</tr>
<tr>
<td><strong>Null plural morpheme</strong></td>
<td>Minor Malabri</td>
</tr>
<tr>
<td></td>
<td>Mandarin Chinese</td>
</tr>
</tbody>
</table>

⁷ Such cases would only be predicted to occur if independent factors prevented the realization of the Plural in some language.
An important insight into the nature of markedness is the statement of grammatical markedness properties as implicational universals rather than absolute properties. For the pattern illustrated in (17) the implicational universal is stated in (18):

(18) If a language uses an overt inflection in the singular, then it also uses an overt inflection in the plural. (Croft 2003:89)

In a two-way (Singular - Plural) system, the implicational universal for the category of Number (15) is conventionally stated as Plural > Singular. This formalism, understood as ‘Plural is more marked than Singular’, is also meant to express the implication in (18).

Starting with the seminal work of Greenberg (1966ab) and the advent of the implicational perspective in grammatical typology, grammatical markedness has been expressed in terms of universal feature hierarchies, where features are understood as the abstract expression of morpho-syntactic categories. Grammatical markedness hierarchies have been proposed for a large variety of categories, including cross-categorial relations. Examples of hierarchies are given in (19) below, following Croft (2003). In (19) members or values of the categories are arranged in the decreasing order of their markedness:
Examples of morphological markedness hierarchies

Number (noun, pronoun, adjective, verb): trial (paucal) > dual > plural > singular

Gender (noun, adjective): neuter > feminine > masculine

Case (noun, adjective): oblique > nominative

Person (verb): 2\textsuperscript{nd} > 1\textsuperscript{st} > 3\textsuperscript{rd} (or 1\textsuperscript{st}, 2\textsuperscript{nd} > 3\textsuperscript{rd})

Tense (verb): future > preterit (past) > present

Aspect (verb, verb phrase): perfective > imperfective

Mood (verb, sentence): hypothetical (subjunctive) > indicative

Voice (verb): passive > active

Inflectional categories: gender > number

Likewise, in the area of phonology hierarchies have been proposed for distinctive features (20):\footnote{For work on phonological feature hierarchies, see Jakobson and Halle (1956), Dresher (2003) and the references therein, and Croft (2003).}

Examples of phonological feature hierarchies

Nasality (vowels): [+nasal] > [-nasal]

Voice (vowels, sonorants): [-voice] > [+voice]

Voice (obstruents): [+voice] > [-voice]

The main question remains what the very nature of grammatical markedness is. While structural coding remains an important diagnostic of markedness, a fuller understanding of the concept is impossible without reference to frequency of occurrence, as we shall see in §1.2.

1.2. Frequency

1.2.1 The failure of the iconicity approach to grammatical markedness

Survey work on grammatical linguistic typology (Croft 1990, 2003) has attempted to establish a possible connection between markedness and iconicity, the latter understood as a way in which “the structure of language [...] reflects the structure of experience” (Croft 2003:102). While a detailed discussion of the iconicity criterion for grammatical markedness would take us too far afield, suffice it to say
that less marked elements are generally more iconic than more marked ones. The idea, appealing at first sight, has considerable drawbacks.

According to the iconicity approach, Plural forms that are overtly marked by inflection can be said to reflect conceptual complexity in that they have more phonological material than the corresponding Singular forms. However, if we examine a three-way (Singular, Plural, Dual) system, we can see that the iconicity approach is confronted with a problem. The key observation is that the Plural can be realized by reduplication, but the Dual is always marked by affixation (cross-linguistically, reduplicated duals seem to be unattested). Within a theory of markedness based on iconicity, this behavior is hard, if not impossible, to accommodate, because we would expect the less marked member of the category (Plural) to be more iconic and reflect reality more faithfully than the more marked one (Dual). Yet it is the Plural that encodes duality from a phonological point of view (reduplicated forms contain a stem and a copy thereof, in other words, a pair of entities), and the Dual just reflects some general conceptual complexity (stem plus affix).

A similar difficulty is encountered when one assesses the iconicity of other grammatical categories. In a study of reduplication in Australian languages, Fabricius (1998) notes a number of mismatches between iconicity and markedness. For example, reduplication is unexpectedly used to express the ‘stative truth’ aspect in Kuku Yalangi, or the ‘attenuative’ aspect in Yankunytjatjara, and no iconicity account seems available for such cases.
As a matter of fact, as Haspelmath (2003) points out, both iconicity and (global) markedness are vague, polysemous notions\(^9\). With respect to iconicity, he convincingly argues that it should be replaced by a concept of economy (perhaps incorporating distinctiveness and parsability). As regards (global) markedness, Haspelmath proceeds to a deconstruction of the cover term ‘markedness’ and concludes that what has traditionally been regarded as markedness oppositions in the Greenbergian tradition are in fact frequency asymmetries motivated by a principle of economy.

1.2.2 Markedness as a consequence of economy

The hierarchies that express grammatical markedness can therefore be viewed as the manifestations of an economy principle that translate into differences in the frequency of use of the respective forms, an idea already present in the work of Greenberg (1966a). While it might be agreed that the grammatical markedness illustrated by the hierarchies in (19) is reflected in structural coding, there is no such consensus for the phonological hierarchy in (20). Some phonologists subscribe to underspecification or monovalent features, but this is not without controversy.

Nevertheless, the question arises whether markedness hierarchies represent primitives of linguistic description or derived objects.

---

\(^9\) Haspelmath (2003) distinguishes no less than twelve acceptations of ‘iconicity’ and ‘markedness’, respectively.
Most contemporary researchers in morpho-syntactic typology favor the latter answer, especially given the complexity and heterogeneous nature of the concept of markedness. In a review, Haspelmath (2003) argues that there are as many as twelve possible acceptations of the term ‘markedness’ that nevertheless share important characteristics and he attempts to identify the unifying factor, which he believes to lie in language use (or linguistic performance\textsuperscript{10}). Along similar lines, Hawkins (2004) shows that grammatical hierarchies are derived notions representing cross-linguistic generalizations that emerge from the interaction between grammar and performance.

Performance is a concept that can be traced back to the early days of linguistic structuralism (‘parole’ in the sense of de Saussure 1916/1972, as opposed to ‘langue’) to generative linguistics (‘performance’ (Chomsky 1965), ‘E-language’ (Chomsky 1986), as opposed to ‘competence’ and ‘I-language’, respectively). Although the labels used in linguistic parlance may differ, the term ‘performance’ is used essentially to designate the actual, physical and observable manifestations of the abstract linguistic system that speakers/hearers are endowed with (‘competence’). Performance covers aspects of language use such as production and perception or constraints imposed by the cognitive system (processing, memory load or pragmatic/informational factors).

Although since the advent of generative linguistics the focus of investigation has mainly been on the side of competence, performance issues

\textsuperscript{10} A brief discussion of the concept of performance is given below.
cannot be overlooked. A striking example of the problems performance factors pose for the understanding of how language works is provided by center embedding phenomena. Center embedding arises due to multiple recursion of linguistic structure and is well known in syntax (Chomsky 1957; Chomsky and Miller 1963; Miller and Chomsky 1963, Miller and Isard 1964 etc.)¹¹.

An example of center embedding sentence is represented by the English sentence # The rat the cat the dog chased ate died.¹² This sentence is perfectly grammatical in the sense that no rule of English syntax (i.e. competence) is violated. Yet most speakers consider the sentence hard, if not impossible to process altogether, most probably due to performance factors such as memory span limitations that lead to processing difficulty.

Center embedding phenomena show that performance factors, traditionally regarded as marginal in classical generative linguistics, need to be paid due attention, as they may cast doubt on the unbounded applicability of grammatical rules. Ultimately, it may be the case that language use (performance) can lead to shaping grammars in the direction of performance preferences, as advocated by Hawkins (1994, 2003, 2004). This issue will be dealt with in §5, in relation to the account offered for the connection between grammatical and phonological markedness.

¹¹ For a recent general discussion of center embedding and numerous references, see Uehara (2003).
¹² The symbol # shows that the sentence is hard to process.
Functional factors like minimization of effort in perception and production have been shown to form the underpinnings of other rules, such as heavy (complex) NP shift (CNPS), a process whereby a complex noun phrase direct object undergoes extraposition over a prepositional phrase and moves to the right (Ross 1967, Postal 1974, Culicover 1976, 1984 etc.).

A performance factor that has been shown to systematically correlate with grammatical or inflectional markedness is the frequency of occurrence of forms as measured in linguistic corpora. In recent years, frequency has often been considered the most important criterion for grammatical markedness.

The correlation between morphologically-marked forms and their frequency of occurrence has been noted since the early days of linguistic typology. As pointed out by Greenberg (1966a:32), the frequency of occurrence of words inflected for Number, as measured in the Sanskrit, Latin, Russian and French corpora considered for analysis, decreases from the (unmarked) Singular to the (marked) Dual. The sense in which frequency of occurrence correlates with grammatical markedness is that the higher a category ranks on a markedness hierarchy, the lower its frequency of occurrence:

(21) Table 3  Frequencies of occurrence of Number category values

<table>
<thead>
<tr>
<th>Language</th>
<th>Size of Sample</th>
<th>Singular</th>
<th>Plural</th>
<th>Dual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sanskrit</td>
<td>93,277</td>
<td>70.3</td>
<td>25.1</td>
<td>4.6</td>
</tr>
<tr>
<td>Latin (Terence)</td>
<td>8,342</td>
<td>85.2</td>
<td>14.8</td>
<td>N/A</td>
</tr>
<tr>
<td>Russian</td>
<td>8,194</td>
<td>77.7</td>
<td>22.3</td>
<td>N/A</td>
</tr>
<tr>
<td>French</td>
<td>1,000</td>
<td>74.3</td>
<td>25.7</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Greenberg measured the frequency of occurrence of other categories (Case, Person, Number and Voice in verbs etc.) and obtained similar correlations with grammatical markedness, as exemplified in (22) - (25) below:

(22) Table 4  Frequencies of occurrence of Case category values

<table>
<thead>
<tr>
<th>Language</th>
<th>Size of Sample</th>
<th>Direct Case</th>
<th>Oblique Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sanskrit</td>
<td>93,277</td>
<td>72.5</td>
<td>27.5</td>
</tr>
<tr>
<td>Latin</td>
<td>8,342</td>
<td>68.7</td>
<td>31.3</td>
</tr>
<tr>
<td>Russian</td>
<td>6,194</td>
<td>65.2</td>
<td>34.8</td>
</tr>
</tbody>
</table>

(23) Table 5  Frequencies of occurrence of Person category values

<table>
<thead>
<tr>
<th>Language</th>
<th>Size of Sample</th>
<th>3\textsuperscript{rd} person</th>
<th>1\textsuperscript{st} person</th>
<th>2\textsuperscript{nd} person</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sanskrit</td>
<td>93,277</td>
<td>54.1</td>
<td>11.3</td>
<td>34.6</td>
</tr>
<tr>
<td>Latin</td>
<td>8,342</td>
<td>45.3</td>
<td>29.3</td>
<td>25.4</td>
</tr>
<tr>
<td>Russian</td>
<td>6,194</td>
<td>50.4</td>
<td>31.9</td>
<td>17.7</td>
</tr>
</tbody>
</table>

(24) Table 6  Frequencies of occurrence of Voice category values

<table>
<thead>
<tr>
<th>Language</th>
<th>Active (%)</th>
<th>Passive (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latin</td>
<td>90.2</td>
<td>9.8</td>
</tr>
<tr>
<td>Sanskrit</td>
<td>73.1</td>
<td>26.9</td>
</tr>
</tbody>
</table>

(25) Table 7  Frequencies of occurrence of Tense category values

<table>
<thead>
<tr>
<th>Language</th>
<th>Present (%)</th>
<th>Past (%)</th>
<th>Future (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sanskrit</td>
<td>53.6</td>
<td>46.3</td>
<td>0.1</td>
</tr>
<tr>
<td>Latin</td>
<td>62.1</td>
<td>26.6</td>
<td>11.3</td>
</tr>
</tbody>
</table>

The sample sizes in (22) and (23) refer to the total number of words in the stretch of language considered for analysis. Frequencies of occurrence of individual members of grammatical categories are calculated as percentage values.
for words inflected for the relevant grammatical category. For example, for the
category of Number, in a language with a three-way Singular-Plural-Dual system,
the frequency of occurrence of the Singular is represented by the percentage ratio
of words inflected for the Singular over the total number of words inflected for
Number, and similarly for the Plural and Dual. For extinct languages like Latin
and Sanskrit, since no spoken language data are available, frequency was
estimated based on written corpora. For other languages, frequency values were
measured using transcripts of conversations, speeches and fictional prose of
various genres. Ideally, the frequency of occurrence of grammatical categories
should be computed on corpora of spoken language that include transcripts from
various stylistic registers (for the methodology of spoken corpus design, see

An important issue that requires clarification is the variation of frequency
values across individual speakers and registers. Such variations do exist (see Biber
1993 for the importance of considering such variation in corpus studies), so it is
important to determine to what extent frequency values measured on a particular
stretch of language represent reliable parameters on the basis of which
grammatical markedness can be assessed.

While there may be certain degrees of variation with respect to the
frequency of individual elements across styles, individuals or language use
situations, the frequency data for a given linguistic system can be said to hold as
relatively fixed. To see in what sense one can talk about the ‘fixedness’ of
frequency values of grammatical categories, consider the behavior of negative versus affirmative constructions in English. Givón (1993) measured the frequency of occurrence of English negative clauses in several type of style varieties. At one end of the style continuum represented by fictional varieties of language, he found that negative clauses represent about 12% of all clauses. At the other end of the continuum (academic varieties), Givón found only 5% of clauses to be negative. The frequency of occurrence of negatives in other stylistic varieties, as measured on language corpora, ranges between 5 and 12%. If we examine these results, we note that despite the higher frequency of (grammatically marked) negatives in the fictional domain as compared to the one in the academic register, (grammatically unmarked) affirmative clauses remain overwhelmingly more frequent as compared to their negative counterparts, irrespective of the language variety considered. In general, although frequency values can vary across speakers, registers or situations, the ratio of marked over unmarked categories is skewed in favor of unmarked ones. This ‘statistic invariant’ behavior renders negative constructions a marked morpho-syntactic status as compared to their affirmative counterparts. It is in this sense that the ‘fixedness’ of frequency parameters in a linguistic system is understood in this dissertation.

The data in (22) - (25) show that the frequency of occurrence of a grammatical category as measured in linguistic corpora follows the same tendency as its ranking on the relevant grammatical markedness hierarchy (see (19) above). Similar results were obtained on Czech, German and Russian by Kučera and
Monroe (1968). In the case study on Number in Romanian (Chapter 5) I computed the frequency of occurrence of number-inflected nominals in a sample of fictional prose. The results are comparable to the ones obtained by Greenberg (1966a) for other languages.

Frequency of usage is a facet of language use that has emerged in recent years as one of the most robust criteria for grammatical markedness. Research on markedness effects in morpho-syntax has lead to the idea that the very term ‘markedness’ can be to a large extent replaced by ‘frequency’ (Werner 1989; Fenk-Oczlon 1991, 2001; Haspelmath 2005). An essentially similar approach to phonological markedness is taken by Gurevich (2001), Hume (2004) and Blevins (2004), who question the status of markedness as a primitive notion and emphasize the role of language use and predictability in shaping grammatical or phonological hierarchies.

Advocates of the frequentionist view of markedness maintain that the frequency of use of linguistic forms reflects the cognitive costs involved in their processing. Forms with high frequency of use presuppose more familiarity and predictability on part of the speaker/hearer and involve a minimization of the processing load on the human processor.

At the same time, frequency of occurrence represents a corollary of economy in language use. The idea of ‘economy’ has long been shown to play a

---

13 Frequency of occurrence has been used to elucidate cases of semantic markedness that were considered ambiguous according to other criteria (Hatzivassiloglou and McKeown 1995).
14 See §5.1 for a discussion of the issue of effort minimization and economy.
role in shaping morphological patterns, the general principle being that linguistic expressions should be kept to a minimum whenever possible, thus contributing to a minimization of the processing burden placed on the speaker-hearer. Economy in language use is encapsulated in the Principle of Least Effort (see Zipf 1949); more recently, further evidence for similar claims in phonetics was brought by Ladefoged (1982) and Lindblom (1990). In syntax, the Minimalist Program (Chomsky 1995) acknowledges economy as one of the fundamental principles of grammar.

As shown as early as by Zipf (1949), the connection between economy and frequency can be informally stated as a statistically significant tendency for expressions that contain less material (or are, in general, structurally less complex) to occur more frequently than forms that are more complex or are ‘bulkier’ in point of expression.

The empirical adequacy of the economy-frequency approach has been tested on a number of phenomena that otherwise constituted notable exceptions to predictions of grammatical hierarchies that do not take language use into account, as we shall see in the following section.

1.2.3 Markedness reversals as a consequence of frequency effects

An important area of success of the frequentionist stand is represented by its ability of account for apparently problematic instances of so-called ‘markedness reversals’ (Mayerthaler 1981, Tiersma 1982, Croft 1990). The phenomenon in
question is seen in situations where categories that are assumed to be grammatically unmarked show an unusual behavior and have overt coding, while the marked counterparts have zero coding.

For example, in Welsh the Singular of some nouns is characterized by an overt suffix, but the corresponding Plural has null marking, as in $plu-en_{Sg}$ ‘feather’ versus $plu-\emptyset_{Pl}$ ‘feathers’. Similarly, we sometimes unexpectedly see overt coding in the third person of the Imperative and, in contrast, null marking in the second person of the same mood, as in Latin $lau\={a}da\=-to_{3rd}$ (‘let him/her praise!’) versus $lau\={a}da\=-\emptyset_{2nd}$ (‘praise!’). In a theory where grammatical categories represent entities arranged in a universally fixed order on a grammatical or inflectional hierarchy such phenomena seem to be hard to accommodate. However, if we consider the functional, frequency-motivated underpinnings of such reversals, the effects no longer appear exceptional\(^\text{15}\).

It is worth noting that most often these cases have a semantic or pragmatic basis; for example in nouns denoting entities that naturally occur in groups or pairs, it is the Plural that is (locally) unmarked, as the category which occurs with higher relative frequency. Similarly, in the Imperative, the second person qualifies as the unmarked/default category, since it encodes the addressee of an utterance, and accordingly it occurs more frequently than the third person. The null marking of the more frequent forms is thus in accordance with economy in language use.

\(^{15}\) It should be mentioned that markedness reversal phenomena can be dealt with in the constraint-based framework provided by Optimality Theory. For an illustration, see the discussion of the Hebrew root-affix asymmetry in Hebrew in §4.2
Along similar lines, Tiersma (1982), who discusses a considerable number of markedness reversals (in Spanish, English, German, Dutch etc.), formulates the following principle:

(26)  When the referent of a noun naturally occurs in pairs or groups, and/or when it is generally referred to collectively, such a noun is unmarked in the plural. (Tiersma 1982:835)

Instances of (apparently exceptional) markedness reversals can be easily predicted in a theory of markedness that relies on economy and frequency of use. If a principle of economy is assumed to play a relevant role in language use (‘performance’), it is possible for a category that is more frequent in point of use (like the Plural of nouns designating entities that occur in pairs or groups) to be apt to receive less expression than the less frequent category.

In this dissertation I assume a similar mechanism to be responsible for the correlation between inflectional markedness and phonological behavior. In (27) below a statement of the pivotal generalization made in this dissertation is given:

(27)  If $g_1$ and $g_2$ are members of the grammatical category $G$ such that $g_2$ ranks higher than $g_1$ on the grammatical markedness hierarchy ($g_2 > g_1$), outputs inflected for $g_2$ can sponsor marked phonological structure to an extent that is equal or greater than $g_1$ (the unmarked term), as a reflection of the economy principle in language use.
The generalization in (27) will be fleshed out in Chapter 3 §4. Informally, if a term has a high frequency of occurrence (due to conceptual, cognitive and pragmatic factors), economy dictates that language users will prefer to avoid marked phonological structure in that term. In grammatically marked items, marked phonological structure is allowed, as its presence is relatively less costly in point of language economy (occurrence in fewer items).

Before discussing the above claim in more detail, let us consider possible correlations between grammatical markedness and the phonological behavior of the respective forms. Economy suggests that such a correlation may be established with phonological bulk or size.

2. Towards a correlation between grammatical markedness and phonology: size

In keeping with the principle of economy discussed in the previous section, one may expect that marked grammatical categories (which have a relatively low frequency of occurrence) are likely to receive less phonological expression than their more frequent, less marked counterparts.

To adapt an everyday example due to Haspelmath (2003), consider phone numbers in the Los Angeles area, which has three area codes: 213, 310 and 323, and where a complete phone number is of the form 213-XXX-XXXX, 310-XXX-XXXX or 323-XXX-XXXX. If a full number is necessary for successful communication in the greater area, which is diverse, people living in the same neighborhood can leave out the area code, which is both predictable and highly
frequent, thus making economy possible. Moreover, people living in the same 
building can afford to omit both the area code and the first three digits of the 
phone number (the exchange), for similar reasons. Therefore the more frequent 
(and predictable) a form is, the more likely it is to receive less expression and to be 
of less size.

Research on economy and morphological coding of grammatical categories 
has shown that Haspelmath’s ‘phone number model’ can be illustrated by the 
behavior of grammatical categories in natural languages.

Considerations of economy have lead Haiman (1983) to state a 
generalization on the size of Case forms that is in line with the status of Case 
values on the grammatical markedness hierarchy:

(28) In no language will the phonological expression of a direct case 
[nominative/accusative] be bulkier than that of the corresponding indirect 
case [dative/locative/instrumental etc.] (Haiman 1983:792)

A straightforward example for Haiman’s generalization comes from 
Turkish, a language with a rich Case system. In Turkish, the Nominative has zero 
inflectional marking, the Accusative is characterized by the suffix -i and the 
Locative, by the affix -de (Lewis 2000), as shown below for the relevant Case 
hierarchy:
Haiman states that his generalization is confirmed in a variety of languages (Walpíri, Greenlandic Eskimo, Kalkatungu, Hungarian and Indo-European (French, German, Russian)). Although the claim is credible in light of the economy principle that demands that frequent, unmarked forms, be less ‘bulky’ than infrequent, marked ones, the demonstration lacks the support of a language sampling technique that would make statistically significant the result according to which marked categories are bulkier than unmarked ones. Also, the generalization covers only the grammatical category of Case; it is not clear whether it could carry onto other morpho-syntactic categories.

The relation between grammatical markedness and phonological expression instantiated by size is also discussed in a recent study by Brunner (2003). Brunner undertook to test the prediction made by Dryer (p.c., cited in Croft 2003) according to which the less marked value of a grammatical category should be no longer phonologically than the more marked category (‘length’ being estimated as segment or syllable count in inflected outputs). Brunner tested Dryer’s prediction for the category of Number on a sample consisting of 42 languages, compiled following the sampling method in Dryer (1989, 1992).
For the category of Number, the prototypical situation in a three-way (Singular, Plural, Dual) system is the one in a language like Yimas (Foley 1991), with null marking in the Singular, and non-null marking in the Plural and Dual, with more phonological material in the Dual:

(30) \[ \text{Dual} > \text{Plural} > \text{Singular} \]

\[
\begin{align*}
\text{yura-Ø} & \quad \text{‘dog’} \\
\text{yura-y} & \quad \text{‘dogs’} \\
\text{yura-ntrm} & \quad \text{‘two dogs’}
\end{align*}
\]

With respect to segment count, the analysis found that in nouns the Singular was longer than the Plural only in 8% of the languages, and the Singular and the Plural never exceeded the Dual. Comparable results were obtained for Number in other lexical categories (verbs, adjectives, pronouns). When phonological size was expressed as syllable count, there were no exceptions, and the marked category was always equal to or longer than the unmarked one.

The size correlation between grammatical markedness and phonological form is also attested for grammatical categories other than Case and Number (Tense, Mood, Gender etc.)\(^{16}\). However, the quantitative approach to the correlation between grammatical markedness and phonological properties is faced with a number of problems.

\(^{16}\) See Haspelmath (2003) and Newmeyer (1998) for examples.
First, as already noted, studies need to be based on reliable language samples, a desideratum that is not always met (see Dryer 1989, 1992 for the importance of language sampling in linguistic typology). Second, and more importantly, the correlations represent statistical tendencies rather than categorical results, and although they are significant in a number of cases, exceptions do exist. The theory should be able to account for such exceptions, but explanations do not readily suggest themselves (markedness reversal solutions are not always available). For example, Limbu (Tibeto-Burman, van Driem 1987) distinguishes between a zero-marked Singular and a suffix-marked Plural and Dual. In Limbu, the Plural suffix consists of three segments, while the Dual suffix has only two. Barring other phonological adjustments in number affixation that would delete material from the stem, the overall segment count of the Plural is higher than both the Singular and Dual segment counts:

(31) Number affixes in Limbu

   Singular: -Ø
   Plural: -ha?
   Dual: -si

Similarly, English verbs take the -s suffix in the 3rd person of the Present Indicative, while the 1st and 2nd persons have zero marking, while in point of
grammatical markedness the 3rd person is the least marked member of the category of Person.

Third, analyses based on phonological size are bound to be confronted with a confounding factor that is hard to tease out. This factor is represented by morphological complexity expressed by the property of structural coding already discussed in §1.1. As we have seen, marked members of grammatical categories are prototypically characterized by higher morpheme counts than the unmarked members. As morphemes typically consist of segments or syllables, it is to be expected that categories with a higher morpheme count will also have a higher segment or syllable count, which makes it difficult to disentangle the effect on morphology from the one on phonology. For example, the English plural ‘dogs’ has a higher segment count than the singular form ‘dog’, but it is hard to tell whether this is the result of the Plural form having a higher count of non-null morphemes that come with their own segmental material or it is just an independent property of the (marked) Plural to be inherently bulkier that the (unmarked) Singular. Also, especially when inflection is expressed by affixation, it may be the case that affixes have different lexical shapes or allomorphs, and while it is true that considerations of frequency or economy can help explain surface distributions, there is no universally valid account for both regular behaviors (more phonological material in the less marked category) and apparent exceptions (smaller size of the marked category).
It appears therefore that a reliable account of the correlation between grammatical markedness and phonological properties has to take into consideration situations where the phonological properties of inflected forms are comparable to a certain extent (and yet different), so that the phonological differences could be attributed to language use factors such as frequency of occurrence or economy.

Consider, for an illustration of the complexity of the issues that the analysis is confronted with, the marking of the Infinitive and Causative in Somali (Saeed 1993, Bendjaballah 1998). Somali does not involve differences in size between the two categories, whose forms are underlying identical (/i/). Nevertheless, when suffixed to a base (formally identical to the Imperative), the Causative marker triggers affrication of the final (velar) stop (32a.), while the Infinitive morpheme does not (32b.):

(32) The behavior of the Causative and Infinitive in Somali

a. Palatalization of the final velar segment of the stem in the Causative:

noog    ‘be tired-Imper.’

noog + i$_{\text{Caus.}}$ $\rightarrow$ nootʃi   ‘cause to be tired’

b. Infinitive /i/ does not trigger palatalization of the velar:

noog + i$_{\text{Inf.}}$ $\rightarrow$ noogi   ‘to be tired’
The Somali case is one where similar phonological material finds a different expression in outputs that represent different grammatical categories. As Bendjaballah (1998) points out, with respect to morphological constituency (morpheme boundaries, number of morphemes) the two forms similar. The respective differences can hardly be attributed to the phonology of Somali alone, and their source lies most probably in some other property of the morphological system.

Cases like Somali potentially represent the ground where hypotheses on the correlation between morphological and phonological properties can be tested. However, there are no accounts in the linguistic typology literature on the relative grammatical markedness of the Causative and the Infinitive, but it is worth noting that the Causative is cross-linguistically a marked category (Haiman 1983). Although there is no difference in size between the Somali Causative and the Infinitive in point of segment or syllable count, it can be seen that the formation of the Causative leads a kind of marked phonological structure as compared to the Infinitive (affricate versus stop). As an additional complication, it is not entirely clear that the Causative is the phonologically marked form. One may argue that the presence of a plain stop before the front vowel [i] in the Infinitive can also be regarded as some kind of marked phonological structure. At the same time, along another dimension, a voiced obstruent is deemed more marked than a voiceless one.
Although not entirely conclusive, the Somali data are helpful in previewing the research agenda that this dissertation addresses. Recall that the claim put forth in this dissertation is that inflectional (grammatical) markedness has a phonological correlate, in that other things equal, categories that are grammatically marked are characterized by the presence of marked phonological material to an extent that is higher or equal than in the unmarked categories, a generalization that I label ‘Marked in the Marked’ (MIM).

It is important to note that the comparison between the phonological shape of outputs inflected for asymmetrically marked categories is relevant only to the extent to which the morpho-phonological markers employed are relatively similar. The Somali case does meet the similarity criterion, since the Infinitive and Causative affixes are both underlying /i/), but the MIM hypothesis cannot be tested due to the facts mentioned above (impossibility to compare the Infinitive and Causative in point of grammatical markedness and also difficulty in assessing the degree of markedness of phonological structure).

To see how important the similarity criterion is, consider the case of Hessian German Plural formation (Golston and Wiese 1996).

In Hessian German there is a class of nouns that form their Plural by apparent deletion of the final consonant of the root if it ends in an obstruent:
Hessian German Plural formation seems to contradict the claim according to which the category ranking higher on the grammatical markedness hierarchy (Plural) displays more marked phonological structure than the lower-ranking category (Singular). Indeed, as the data in (33) show, if the Singular has a complex coda, the corresponding Plural has a simple coda, or, even more dramatically, if the Singular has a closed syllable, the Plural has an open one. This behavior would constitute a counterexample to the MIM generalization.

Nevertheless, Golston and Wiese’s examination of Hessian German’s inflectional system shows that what is at first sight subtractive morphology is in fact a means of satisfying a requirement that Plural forms end in a segment that carries the [+sonorant] feature (the Singular is represented by the bare root, and no such requirement is enforced). Plural formation by apparent subtraction is only a strategy whereby final sonorancy is satisfied; final consonant deletion in the Plural

---

(33) Singular | Plural | Gloss
---

hond | hon | ‘dog’

ɛn | ɛn | ‘end’

ʃøk | ʃø | ‘shoe’

vek | vɛ | ‘way’

---

17 As we shall see in §3.1, complex codas and closed syllables are more marked phonologically than simple codas and open syllables, respectively.

18 Furthermore, Golston and Wiese (1996) suggest that the Plural formative is the [+sonorant] floating feature that needs to be aligned to the right edge of the prosodic word.
may just be a bi-product of suffixation of [+sonorant]. In fact, as Stonham (1994) points out, inflectional non-concatenative morphology (including subtraction) is in most cases an instance of concealed affixation and has no independent status in morphological theory. In sum, Hessian German employs different strategies for expressing the Singular and the Plural. The fact that the Plural is less marked phonologically than the Singular cannot be used to falsify the MIM generalization, which is assumed only for forms with relatively similar morphological marking. This is not the case in Hessian German, where the Singular has a zero marker and the Plural, a subsegmental [+sonorant] affix.

So far we have reviewed the notion of grammatical or inflectional markedness and we have discussed two widely accepted criteria for this kind of markedness, structural coding and frequency of occurrence as measured in corpora. Grammatical markedness hierarchies were introduced as an expression of structural coding. We have seen that frequency of occurrence is superior to other criteria such as iconicity and can help explain markedness reversals phenomena. Also, I introduced the Marked in the Marked generalization and I have shown that the correlation between grammatical markedness and the phonological properties of outputs inflected for the respective grammatical categories is not one of mere size, but rather one that has to do with the phonological markedness of those outputs.

Having established the substantive content of inflectional markedness and having previewed its connection with the phonology, we can now proceed to
discuss in more detail the notion of phonological markedness. The more specific issue of how markedness is dealt with in Optimality Theory (Prince and Smolensky 1993/2004), which constitutes the theoretical framework assumed for the analysis, will be discussed in Chapter 3.

3. Criteria for phonological markedness

Since the main claim of this dissertation involves the notion of ‘phonological markedness’, it is important to first establish what criteria can be employed in assessing it.

In a survey of markedness in phonology, Rice (2003) lists a number of criteria that can be used to characterize marked versus unmarked oppositions in phonology:

(34) Table 8  Marked versus unmarked oppositions

<table>
<thead>
<tr>
<th>Marked</th>
<th>Unmarked</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less natural</td>
<td>Natural</td>
</tr>
<tr>
<td>More complex</td>
<td>Simpler</td>
</tr>
<tr>
<td>More specific</td>
<td>More general</td>
</tr>
<tr>
<td>Less common</td>
<td>More common</td>
</tr>
<tr>
<td>Unexpected</td>
<td>Expected</td>
</tr>
<tr>
<td>Not basic</td>
<td>Basic</td>
</tr>
<tr>
<td>Less stable</td>
<td>Stable</td>
</tr>
<tr>
<td>Appear in few grammars</td>
<td>Appear in more grammars</td>
</tr>
<tr>
<td>Later in language acquisition</td>
<td>Earlier in language acquisition</td>
</tr>
<tr>
<td>Subject to neutralization</td>
<td>Neutralization targets</td>
</tr>
<tr>
<td>Early loss in language deficit</td>
<td>Late loss in language deficit</td>
</tr>
<tr>
<td>Implies unmarked feature</td>
<td>Implies marked feature</td>
</tr>
<tr>
<td>Harder to articulate</td>
<td>Easier to articulate</td>
</tr>
<tr>
<td>Perceptually more salient</td>
<td>Perceptually less salient</td>
</tr>
</tbody>
</table>

53
To the criteria reviewed by Rice we can add the one of predictability within a particular phonological system, as claimed by Hume (2004), who proceeds to a deconstruction of the notion ‘markedness’ in phonology and claims that it is not a primitive of linguistic description.

It appears that phonological markedness is a multifaceted property. Also, there is no consensus in the literature as to exactly how many of the properties in (34) a phonological structure should have in order to qualify as (relatively) marked. One may also wonder whether there is some ‘core’ and ‘periphery’ in the concept of markedness, so that forms that have the ‘core’ properties can be considered marked.

Among the many possible diagnostics for phonological markedness I will discuss two that are relatively uncontroversial: complexity and occurrence in phonological inventories.

3.1 Phonological markedness as complexity

Despite the difficulties in approaching phonological markedness that I hinted at above, many researchers believe that phonological markedness is robustly correlated with complexity. Phonological markedness can be viewed as complexity from a structural or articulatory point of view.

Structural complexity has been adopted as a criterion for markedness in a number of studies (Drescher and Rice 1993, Dyck 1993, 1995, Ghini 1993, 2001,
Dresher, Piggott and Rice 1994, Dresher and van der Hulst 1998, Rice and Avery 2004 etc.).

As noted by Dresher and Rice (1993), the notion of phonological complexity is relevant and can manifest itself at various levels.

For instance, at the segmental level, the ability of units to participate in phonological processes varies as a function of their internal (featural) complexity or makeup. Markedness can be said to be encoded in the representation of segments and increases with the amount of structure. A similar claim is made by Rice and Avery (1995).

Dresher and van der Hulst (1998) use the same kind of argument in discussing the relation between vowel distribution and segment complexity in a number of languages. As an illustration, in Russian strong (stressed) syllables can have one of the five vowels [i, u, e, o, a] as a nucleus, but in weak (unstressed) syllables only [i, u, a] can occur as nuclei. The authors correlate this kind of distribution with the relative degree of complexity in vocalic segments, according to the representations in (35), which view vowels as combinations of particles or elements in the spirit of theories that employ unary features, like Particle Phonology (Schane 1984) or Government Phonology (Kaye, Lowenstamm and Vergnaud 1985):
According to this metric, the relatively less complex vowels [i, u, a] are less marked than [e, o].

Segment complexity is also acknowledged in autosegmental representations. For example, Clements (1985) represents consonants with a secondary articulation by subordinating a V-Place node corresponding to the secondary articulation to a consonant’s C-Place node:

According to the representations in (36) above, consonants with a secondary articulation (36a. - d.) are more complex (and consequently, more marked) than the one with only a primary place of articulation (36e.). For complex segments involving place features see also Sagey (1986/1990).
Segment complexity relates markedness and the internal makeup of units of analysis at the same level of representation. Complexity effects can also be noted at various levels of representation, for example with respect to the Prosodic Hierarchy (Selkirk 1980, McCarthy and Prince 1986):

(37) Figure 3 Prosodic Hierarchy

Prosodic Word
  | Foot
  | Syllable
  | Mora

What is relevant for the correlation between complexity and phonological markedness is the constituency of a category, not necessarily the place of that category on the Prosodic Hierarchy. Thus at the syllable level closed CVC syllables are more marked than open CV syllables, heavy (bimoraic) syllables are more marked than light (monomoraic) syllables etc. (see Dresher and van der Hulst (1998) for a review of phonological complexity that correlates with levels of prosodic structure)\(^\text{19}\). A similar relation holds for bimoraic versus monomoraic syllables.

Articulatory and/or perceptual complexity has also been shown to closely correlate with phonological markedness. The idea of articulatory or perceptual

\(^{19}\) Closed syllables differ in structure from open syllables at the root node. There may also be differences in moraic structure.
grounding of phonological markedness is by no means new and can be seen as early as in the work of Trubetzkoy (1931), who regards articulatory complexity as an essential criterion of phonological markedness.

More recently, the relative markedness of a various segment classes has been correlated with articulatory complexity by Chomsky and Halle (1968), Lindblom and Maddieson (1988), Willerman (1994) etc. According to Chomsky and Halle, a class of segments is marked because it is harder to articulate in comparison with a class that is easier to articulate (unmarked). For example, a retroflex sound is considered more marked than an apical alveolar or a dental because retroflexes involve a raising and displacement of the tongue tip towards the post-alveolar region, whereas an apical alveolar involves only a tongue tip raising (Hamann 2003). Similar observations have been made for other instances of segmental markedness, like clicks in relation to consonants articulated with a pulmonic airstream mechanism (Engstrand 1997) or palatalized consonants in relation to plain consonants (Ní Chiosáin and Padgett 2001, Zygis 2004). With respect to the articulatory complexity of palatalized consonants (discussed in this dissertation in Chapter 5, in relation to Number expression in Romanian), they have been shown to be articulatory complex because they involve the superimposition of an [i]-like gesture upon a labial, dental, alveolar or post-alveolar (velar) consonant (Ladefoged 1971, Ladefoged and Maddieson 1996).

More generally, the phonetic and psycholinguistic underpinnings of phonological markedness as well as their optimality-theoretic implementations
have been highlighted by Flemming (1995), Steriade (1995c), Boersma (1998), Hayes (1999) and many others\textsuperscript{20}. Within a pre-OT framework, the issue is addressed by Archangeli and Pulleyblank (1994).

To conclude the discussion of complexity as a criterion for phonological markedness, it should be mentioned that structural and articulatory complexity often correlates with perceptual difficulty. There is a sizable body of evidence that points to the correlation between perceptual difficulty and phonological markedness (see, among others, Flemming (1995), Steriade (1997, 2001), Beckman (1998/1999), Boersma (1998), Kirchner (1998), Haspelmath (1999), Hayes (1999), Côté (2000), Pierrehumbert (2000), Bye (2001), Smith (2002/2005), Curtin (2002), Padgett (2003ab), Hume (2004), Walker (2005) and the contributions in Hayes, Kirchner and Steriade (2004)). As an expression of the correlation between markedness and perceptual difficulty, unmarked phonological structures, which are relatively easy to perceive due to rich perceptual cues, tend to occur in more words than marked phonological structures, characterized by weaker perceptual cues (Hume 2004). The importance of the perceptual difficulty criterion for phonological markedness is also highlighted by the fact that perception and production often go hand in hand, in that marked elements, which are perceptually difficult, may require more extreme articulation to improve their perceptibility, and conversely, reducing articulatory effort often threatens perceptual distinctness. Also, the neural systems supporting speech perception and speech production

\textsuperscript{20} Interestingly, in sign languages a correlation was recently noted between the complexity of hand shapes (‘ease of articulation’) and frequency of occurrence (‘markedness’) (Ann 2006).
partially overlap in the superior temporal lobe, as shown by Hickok and Poeppel (2000) and Hickok (2001).

A second, widely used diagnostic for markedness in phonology is the frequency of occurrence in grammars or inventories across languages, a topic which will be addressed in the following section.

3.2 Phonological markedness as (non-)occurrence in inventories

As in the case of grammatical markedness, frequency of occurrence within individual languages or in inventories has been shown to correlate with phonological markedness. The observation, made as early as in the studies of Jakobson (1932/1984, 1939/1984), Trubetzkoy (1939/1969) or Greenberg (1966a), is that marked phonological material at various levels of representation (features, segments, syllables etc.) occurs less frequently in inventories than unmarked material.

Greenberg (1966a) discusses complex articulations (glottalization, palatalization) in a number of languages, including Hausa, Klamath, Coos, Yurok, Chiricahua, Maidu and Russian, and he finds that segments with secondary articulation are significantly less frequent than plain sounds. Long versus short vowels in Icelandic, Sanskrit, Czech, Hungarian, Finnish, Karok and Chiricahua follow the same statistical tendency, and so do nasal vowels, as opposed to non-nasals.
As to occurrence in inventories, there is a similar statistical tendency for unmarked structure to occur in more individual inventories than marked material. The table in (38), compiled by Clements (2004) on the basis of the UPSID database (Maddieson 1984, Maddieson and Precoda 1989) illustrates this behavior for consonant features:

(38) Table 9 Unmarked versus marked consonant features

<table>
<thead>
<tr>
<th>all languages have:</th>
<th>some languages lack:</th>
<th>marked feature$^{21}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>obstruent consonants</td>
<td>sonorant consonants</td>
<td>[+sonorant]</td>
</tr>
<tr>
<td>coronal consonants</td>
<td>labial, dorsal, pharyngeal or laryngeal consonants</td>
<td>[labial], [dorsal], [pharyngeal] etc.</td>
</tr>
<tr>
<td>oral consonants</td>
<td>nasal consonants</td>
<td>[+nasal]</td>
</tr>
<tr>
<td>stop consonants</td>
<td>continuant consonants</td>
<td>[+continuant]</td>
</tr>
<tr>
<td>unaspirated stops</td>
<td>aspirated consonants</td>
<td>[spread glottis]</td>
</tr>
<tr>
<td>non-glottalized stops</td>
<td>glottalized stops</td>
<td>[constricted glottis]</td>
</tr>
<tr>
<td>anterior coronal stops</td>
<td>posterior coronal stops</td>
<td>[+posterior]</td>
</tr>
<tr>
<td>non-strident coronals</td>
<td>strident coronals</td>
<td>[+strident]</td>
</tr>
<tr>
<td>simple consonants</td>
<td>consonants with secondary articulations</td>
<td>features of secondary articulation</td>
</tr>
</tbody>
</table>

With respect to the presence/absence of certain items, the categories which some languages lack in (38) might be for just a particular counterpart of an element in the categories that they have. For instance a language might have [g] but lack a velar nasal, and at the same time show stop/nasal pairs at bilabial and alveolar points of articulation.

$^{21}$ I am not necessarily committed to this particular version of features. The ‘marked feature’ column serves nevertheless our demonstration purposes.
Of course, a statistical approach per se does not explain why some segments are more frequent than others, but has the empirical advantage of relating markedness to observable parameters, such as frequency distributions.

It should also be noted that frequency of occurrence in phonological inventories does not necessarily coincide with the frequency of occurrence of the relevant phonological structures in individual languages. An example for this difference is provided by Arabic, where the coronal voiceless stop /t/, although relatively unmarked cross-linguistically and present in a large number of inventories, is found only in a small number of words (Pierrehumbert 2003).^{22}

Given the diversity of diagnostics proposed for phonological markedness, the task of correlating it with grammatical markedness is by no means an easy one. In the case studies presented in this thesis I discuss instances of phonological markedness that are uncontroversial: uneven moraic trochees in Old Saxon (Chapter 4), palatalized consonants in Romanian (Chapter 5) and intervocalic voiceless stops in Mayak (Chapter 6). In each of the cases the phonological markedness of the relevant structure is assessed and the two criteria for phonological markedness discussed in this chapter (complexity and occurrence in inventories) are discussed.

Since the theoretical framework assumed for the dissertation is Optimality Theory (Prince and Smolensky 1993/2004), a word is in place at this point about

^{22} This seems at odds with Greenberg’s statistical observations. It should be noted, however, that Greenberg’s correlations were established for phonological features (features of secondary articulation, length, nasality etc.), not necessarily for individual segments in inventories.
the way in which markedness is addressed in OT. Unlike approaches to
markedness that capitalize upon phonetic or representational facts, in OT there is
only one way to show that a structure is marked, namely to show that there exists a
markedness constraint against that structure and that employing the markedness
constraint leads to the right typological predictions (see McCarthy 2002 for a
discussion of markedness in OT versus Prague School markedness). For example,
(HL) uneven trochees are marked as there is a constraint *(HL) that prohibits
them. I will return to the issue of markedness in OT in Chapter 3.

3.3 Summary: grammatical and phonological markedness

So far we have determined reviewed the most important diagnostics proposed for
grammatical and phonological markedness. For grammatical markedness structural
coding and frequency of occurrence in language corpora have emerged as the most
reliable criteria, in the sense that marked categories are characterized by at least as
much inflectional coding as the unmarked ones and have a lower frequency of
occurrence than their unmarked counterparts.

Since we will be testing the Marked in the Marked generalization, which
was stated for forms that have relative similar phonological expression of their
inflectional markers, the frequency criterion turns out to be the essential trait of
grammatical markedness. In general, for a grammatical category G with g₁ and g₂

23 See Chapter 4 for a discussion of uneven trochees as marked phonological structure. Constraints
against uneven trochees have been proposed by Prince (1990), Prince and Smolensky (1993/2004)
as its members, if \( g_1 \) and \( g_2 \) are expressed by similar phonological markers and \( g_2 \) has a lower frequency of occurrence than \( g_1 \), \( g_2 \) is more marked than \( g_1 \). For instance, in Romanian masculine and neuter nominals the expression of Number is represented by a high vowel suffix (/u/ in the Singular and /i/ in the Plural), although the Singular affix is not always expressed in outputs. Plural forms are less frequent in point of occurrence (along the lines of Greenberg 1966a)\(^{24}\), which makes the Plural (\( g_2 \)) more marked than the Singular (\( g_1 \)). For the purpose of the MIM hypothesis, grammatical (inflectional) markedness is essentially reducible to a variable of language use represented by frequency of occurrence of expressions.

For the purpose of this dissertation, the main criteria for phonological markedness are complexity (in structural and articulatory terms, and correlated with perceptual difficulty) and relative frequency of occurrence in phonological inventories across languages.

For each of the phonological structures (\( M \)) involved in the illustration of MIM it will be shown that the structure in question meets these criteria. In addition to that, since the theoretical framework assumed in the thesis is Optimality Theory, OT markedness constraints against \( M \) (*\( M \)) will be defined. The criteria of complexity and occurrence in inventories represent a way in which *\( M \) constraints are grounded in structural, articulatory or language use factors.

So far, we have determined the basic content of the notions of grammatical and phonological markedness and we have seen what problems are raised by the

\(^{24}\) See also our own text statistics in Chapter 5.
connection between them. We are now in a position to take a first step in testing the Marked in the Marked hypothesis. We have also seen (§2) that if there is a correlation between grammatical markedness and phonological properties, the correlation is not necessarily one of phonological size. Since frequency of occurrence has been shown to be the most reliable correlate of grammatical markedness such that marked categories are also less frequent tokenwise, we can paraphrase MIM as the ‘Marked in the Infrequent’. In the following section I will discuss possible phonological correlates of frequency in open versus closed class categories and affixes versus roots.

4. Frequency of occurrence and phonological markedness

4.1. Open versus closed class categories

A fertile ground for testing the correlation between phonological markedness and frequency of occurrence in actual speech is provided by the behavior of open and closed class word categories. It is well known that open class categories (also known as ‘lexical’ or ‘substantive’: nouns, verbs, adjectives etc.) have a lower frequency of occurrence than closed class categories (‘grammatical’ or ‘functional’: pronouns, determiners, auxiliaries, certain prepositions etc.)

25 See Kučera and Francis (1967) for a comprehensive study of word frequency in contemporary American English.
Observations of this sort have been made in the literature in the last several decades. As regards segmental markedness, Swadesh (1971) remarks that in languages having clicks as part of their phonological inventory, these sounds are part of ordinary verbs, nouns, and adjectives (lexical, open-class items), but not of demonstratives, pronouns, and particles (functional, closed-class items).

A more articulate account of the presence of marked segments in functional categories is provided by Willerman (1994), who compared phonological profiles of pronouns in 32 typologically different languages. She found that segments that are cross-linguistically marked, such as clicks, affricates, uvulars, ejectives and consonants with secondary articulation occurred with less than predicted frequency in the pronominal paradigms. At the same time, segments that are less marked phonologically (bilabials, glottals, nasals and approximants) occurred with greater than predicted frequency in pronouns. Working with an independently developed scale of articulatory simplicity/complexity, Willerman found that the infrequently occurring segments were those that are relatively more complex, and they were present with relatively lower frequency in pronouns. Conversely, segments that are overrepresented in pronominal paradigms are typically the phonologically unmarked ones, at least from an articulatory standpoint.

Prosodic markedness has also been noted to correlate with membership in open versus closed class categories. A statistical study conducted by Shi, Morgan and Alloopena (1998) on onset complexity shows that (less frequent) open-class

---

26 For the status of clicks as phonologically marked segments in relation to their articulatory complexity, see Sagey (1986/1990).
words are more likely to have complex syllable onsets than (more frequent) closed-class words.

Another area of morphology where we can see differences in phonological markedness between categories that occur with different frequency is the root-affix distinction.

4.2. Roots vs. affixes


The phonological reflex of the root-affix distinction has often been considered to derive from the phonetic or psycholinguistic prominence of roots as compared to affixes (see Beckman 1998/1999 and the references therein), which confers roots special faithfulness properties as compared to affixes. In optimality-theoretic terms, this property translates in the universally assumed fixed ranking labeled the Root-Affix Faithfulness Metaconstraint (McCarthy and Prince 1995), according to which faithfulness to root-sponsored phonological material outranks faithfulness to phonological material in affixes.

Less attention has been paid to the specific issue of occurrence of marked/unmarked phonological structure in roots versus affixes. However, a recent paper
by Ussishkin and Wedel (2002) lists a number of situations where affixes are attested to host unmarked phonological material, while roots can contain both marked and unmarked material.

In English, affix consonants are predominantly coronal, a relatively common, unmarked place of articulation (the situation is largely similar in other Indo-European languages). In Salish, glottalized consonants occur only in roots and lexical suffixes, never in grammatical affixes. The relatively marked mid, rounded vowels occur in Turkish only in roots, rarely in affixes. Discrepancies between the distribution of vowels are also attested in Frisian and Dutch, where affixes can contain only an unmarked subset of the vowel inventory (see also Dyck 1995 on asymmetries in vowel inventories occurring in stems versus affixes such that affixes show smaller inventories). In prosody, roots are different from affixes in Sanskrit, in that only the former allow for complex onsets\textsuperscript{27}.

The asymmetric behavior of roots and affixes parallels that of lexical (open-class) and functional (closed-class) items. From an ontological perspective, this is not unexpected, since roots, which are more likely to contain marked phonological structure, are open-class lexical elements, while affixes are closed-class functional elements\textsuperscript{28}. The frequency effects noted in the previous section for open versus closed-class items are expected to hold for roots and affixes; the former occur with less token frequency than the latter (see Segalowitz and Lane

\textsuperscript{27} For references, see Ussishkin and Wedel (2002) and the citations therein.

\textsuperscript{28} The generalization regarding the phonological markedness of roots and affixes can be extended to reduplication structures. The reduplicant, a morphological formative with an affix-like, functional status, is usually less marked than the base (a lexical element). The property is known as the Emergence of the Unmarked (McCarthy and Prince 1994).
2000 and the references therein for frequency data). The picture is reminiscent of ‘marked in the less frequent’ effects.

Such an account has positive consequences for explaining certain apparently unexpected reversals in the phonological properties of roots and affixes. Hebrew (Ussishkin 2000, Ussishkin and Wedel 2002) represents a system where affix faithfulness outranks all other instances of faithfulness, including root faithfulness, which contradicts the prediction of the Root-Affix Faithfulness Metaconstraint. In the templatic morphology of Hebrew, verbs are restricted to two syllables and are formed through the concatenation of a bisyllabic base with a bivocalic affix. The undominated bisyllabic constraint on inflected forms can only be satisfied by deleting material from the base:

(39) The verbal paradigm for gadal, ‘to grow’ (Ussishkin and Wedel 2002)

<table>
<thead>
<tr>
<th>Base form</th>
<th>+affix</th>
<th>Derived form</th>
<th>Gloss</th>
</tr>
</thead>
<tbody>
<tr>
<td>gadal</td>
<td></td>
<td></td>
<td>‘he grew’</td>
</tr>
<tr>
<td>i e</td>
<td>gidel</td>
<td></td>
<td>‘he raised’</td>
</tr>
<tr>
<td>u a</td>
<td>gudal</td>
<td></td>
<td>‘he was raised’</td>
</tr>
<tr>
<td>hi i</td>
<td>higdil</td>
<td></td>
<td>‘he enlarged’</td>
</tr>
<tr>
<td>hu a</td>
<td>hugdal</td>
<td></td>
<td>‘he was enlarged’</td>
</tr>
</tbody>
</table>

As can be seen in (39), faithfulness to affix material outranks faithfulness to root (base) material, thus contradicting the Root-Affix Faithfulness
Metaconstraint. Ussishkin and Wedel note that in Hebrew, a language with a five-vowel inventory where verbal affixes consist of two vowels, these morphemes are near neighbors of one another. Unlike in other languages, the high frequency of a Hebrew verbal affix cannot compensate for low phonemic contrast, so affix faithfulness ranking higher than root faithfulness is a mechanism by which contrast between affixes is maintained. Ussishkin and Wedel suggest that the apparent root/affix faithfulness reversal effect in Hebrew is functionally grounded in frequency and neighborhood density for roots and affixes (a property that the authors label ‘effective contrast’). The recourse to language use factors offers an elegant, straightforward account for the reversal, in a manner akin to the solution proposed by Tiersma (1982) for dealing with markedness reversals in morphological linguistic typology (§1.2.3).

5. The functional grounding of Marked in the Marked effects

One of the central ideas advocated in this dissertation is that the connection between grammatical markedness and phonological markedness (expressed by the Marked in the Marked generalization) is functionally grounded.

A first step in this direction was taken in §1.2, where grammatical markedness was shown to closely correlate with a usage factor such as frequency of occurrence. In Chapter 3, where the optimality-theoretic framework assumed for the analysis is introduced, we will see that many OT constraints can be shown to have a functional grounding and we will introduce a licensing constraint
LICENSE(M,g). The licensor of marked phonological structure M in the marked member g of grammatical category G is represented by the morpho-syntactic features of g, which are marked by virtue of their low frequency of occurrence.

The picture sketched so far suggests that the MIM generalization is rooted in language use. In this section I investigate in more depth the functional underpinnings of MIM by looking at two conflicting forces that are at work in language use and processing.

First, there are principles of economy that tend to keep linguistic expression to a minimum (like the Principle of Least Effort formulated by Zipf 1949), whose effect becomes clear when we consider phenomena of reduction affecting mainly high frequency items, as shown, among others, by Bybee 1994, 2001, Fidelholtz 1975, Hooper 1976, Mańczak 1980, Phillips 1980, Fenk-Oczlon 2001 etc.

Minimization of expression in speech production, a speaker-oriented principle, is the second economy force at work. It is inherently in conflict with the listener-oriented principle of sufficient contrast which demands that there be enough expression to facilitate perception and decoding of information.

To see how these principles work, consider two simple examples from diachrony that we owe to Haspelmath (2002). In Old High German, the Genitive Plural marker was -ono, while in Modern German it is reduced to -en, as in zungono - Zungen (‘tongues’ - tongue-genitive plural’). From the speaker’s point of view, this sort of form reduction represents an instance of economy. On the
other hand, there are cases where the amount of structure increases in time. In Old English, a noun like *dohtor* (‘daughter’) has null marking both in the Singular and the Plural, whereas in Modern English the Plural is characterized by the presence of a suffix (‘daughter-s’). The particular evolution in English illustrates a principle of clarity in communication whereby comprehension of distinct meanings is aided by formal differences that may run counter to economy.

With regard to economy in production (the speaker-oriented principle) I discuss the advantages marked in the marked phenomena can have in language use in §5.1. Such advantages are indirect, in that less marked phonological material in more frequent forms favor effort minimization in production. In §5.2 I focus on some positive consequences of MIM from a hearer-oriented perspective, and I address the issues of perception, retrieval and language acquisition.

### 5.1 Production: MIM and form minimization

As Hawkins (1983, 1990, 1992ab, 2003, 2004) notes, grammatical markedness hierarchies are in fact performance frequency rankings encoded in the grammar or instances of ‘performance shaping grammars’. The ultimate cause of frequency effects resides largely in pragmatic and/or cognitive and semantic factors that have to do with the way language users perceive and express entities in the real world. According to this assumption, groupings of two objects (Dual) can be less frequent than groupings of more than two objects (Plural), and groupings in general involve a higher level of abstractness that makes the Singular potentially more readily
available and necessary than the Plural, the Plural more available than the Dual and so on. The empirical adequacy of this take on grammatical markedness is supported by its ability to account for apparent markedness reversals, already discussed in §1.2.3.

In phonology, a number of researchers have claimed that functional grounding is internal to the grammar (Vennemann 1974, Hooper 1976, Stampe 1973, Donegan 1978, Donegan and Stampe 1979, Archangeli and Pulleyblank 1994 etc.)\textsuperscript{29}. Moreover, the particular behavior of phonetically or psycholinguistically prominent positions (Smith 2002/2005) or the mechanism of inductive grounding (Hayes 1999) offer insights into the functional grounding of phonological processes, although neither Smith nor Hayes makes explicit assumptions as to grammars being directly determined by functional factors.

Given this picture, the next question to ask is, of course, how functional factors correlate with formal complexity that can in turn translate into phonological markedness. An indirect answer to this question is provided by the Minimize Forms (MiF) principle proposed by Hawkins (2004):

\begin{align*}
\text{(40) Minimize Forms (MiF)}
\end{align*}

The human processor prefers to minimize the formal complexity of each linguistic form \( F \) (its phoneme, morpheme, word, or phrasal units) and the number of forms with unique conventionalized property assignments, thereby assigning more properties to fewer forms. These minimizations

\textsuperscript{29} It should be noted that not all researchers agree that the grammar is functionally grounded. See Smith (2004a) for a review and references on functional grounding in phonology.
apply in proportion to the ease with which a given property P can be assigned in processing to a given F. (Hawkins 2004:38)

In general, the MiF principle prefers structures with less material to those with more. This is possible if structures with less material are more predictable (and consequently, also more frequent in occurrence than structures with more material), which constitutes the expression of economy in communication along the lines of the Principle of Least Effort (Zipf 1949)\(^{30}\).

In syntax and morphology, the workings of MiF can be easily seen. Consider, for illustration, grammatical roles and case marking. If the grammatical role (subject, object etc.) of a certain NP can be reliably associated with its syntactic position within the sentence, the grammar can dispense with the presence of specific case markers. For example, in a language where there is a canonical position for the subject, the Nominative case need not be overtly expressed, while the object is usually case-marked\(^{31}\).

The explanatory power of MiF also manifests itself if we consider the grammatical markedness – phonological markedness relation that this dissertation investigates. In §2 we noted the tendency of grammatically marked forms to have greater phonological size (in terms of segments, syllables etc.) than their unmarked counterparts. Although this correlation is not without exceptions, it is

\(^{30}\) MiF is reminiscent in its effects of other principles of economy in production, like the Economy Principle (Haiman 1983) or principles of pragmatic efficiency, like Grice’s (1975) second Quantity Maxim or Levinson’s (2000) Minimization Principle.

\(^{31}\) See also Drellishak (2005) for a discussion of other syntactic phenomena that illustrate MiF, in particular, coordination.
uncontroversial that less phonological bulk in the unmarked, frequent category constitutes an advantage in production in terms of economy.

A similar advantage is noted in the cases of grammatical markedness – phonological markedness correlations where the latter kind of markedness is not simply a matter of size, but rather of phonological complexity (articulatory or structural, as discussed in §3.1). Indeed, given the criteria we have adopted for phonological markedness, marked phonological material is more complex articulatorily and is more likely to involve the presence of more numerous and/or complex articulatory gestures than unmarked material. It follows that circumscribing the marked phonological material to the less frequent, more marked categories, is bound to place less articulatory burden on the language user. For example, restricting consonants with secondary articulation to the more marked grammatical category \(^{32}\) creates an advantage in production in the sense that articulatorily complex structures are confined to a relatively smaller number of forms, and the overall articulatory effort is kept to a minimum.

Note that the activity of MiF is in effect one of optimization of form and properties that can be assigned to those forms. While forms with a high frequency of use are likely to be characterized by minimal formal marking, less frequently used forms are in a sense more conservative in that they overtly mark their corresponding properties such as grammatical category (this idea is expressed in Hawkins’s (2004) principle of Morphologization).

\(^{32}\) For an illustration of the phenomenon, see the analysis of Standard Romanian in Chapter 5.
The first to study the behavior of lexical competitors in relation to frequency in production were Landauer and Streeter (1973). Since then, numerous studies have been devoted to this issue.

With regard to the problem of errors in production, low frequency forms have been shown to be generally more vulnerable to speech errors than high frequency forms (Marquardt et al. 1979, Stemberger and MacWhinney 1986, Vitevitch 1997); the fact that low frequency forms can host articulatorily difficult, marked phonological material could make them even more prone to errors, which seems to put such forms to a disadvantage in production33.

On the other hand, the question arises whether there is some other processing advantage other than economy in production that may favor the occurrence of marked phonological structure in marked, less frequent grammatical categories that nevertheless express some relevant and necessary property P. This question will be addressed in the following section.

5.2 Perception, retrieval and acquisition

If assigning phonological complexity to grammatical categories in relation to their place on a grammatical markedness hierarchy has obvious advantages in production, where effort is minimized on the speaker’s side, it is less clear what, if any, the advantages of such behavior could be in perception and retrieval. On the

---

33 Of course, one may wonder why marked phonotactic structure does not then occur in more frequently used forms, where errors are less likely to occur. The fact that we are not seeing such ‘transfers’ seems to indicate that the overall cost of such a hypothetical ‘transfer’ would be too high on the speaker’s part.
contrary, it could be claimed that the unmarked category, with its relatively unmarked phonological structure, should be easier to perceive and retrieve given its high frequency of occurrence.

This is what psycholinguistic studies generally show (e.g. Gordon 1983, Glanzer and Eisenreich, 1979). Work by Marslen-Wilson (1990) brings evidence to bear on higher frequency words being generally faster and easier to access than lower frequency words, even when they are balanced on other features, like phonological size. For example, a frequent English word like ‘pen’ is accessed and recognized faster than the less frequent word ‘pun’, the two lexical items having the same segment count and similar phonological makeup. Another advantage of high frequency of occurrence has been studied by Balota and Chumbley (1985), who showed that the word frequency effect could be at least partially attributed to articulatory processes. Such a view implies that articulatory programs for high-frequency words may be compiled (and also executed) faster than those for low-frequency words.

Nevertheless, high frequency of occurrence *per se* is not an inherent advantage in processing. A number of studies have shown that the effect of frequency on word intelligibility is mediated by the number of competitors for a given word, and can be eliminated if the number of lexical neighbors is controlled (Havens and Foote 1963, Pisoni et al. 1985, Luce 1986).

The question that is relevant for the phenomena approached in this dissertation is, of course, whether Marked in the Marked effects in lower
frequency items present certain advantages in perception and retrieval. Without offering a direct answer to this question, there are studies suggesting that such advantages do exist.

In particular, Monaghan et al. (2005) have shown distributional information to be more useful for categorizing higher frequency words, while phonological cues provide more valid data for lower frequency words. Using corpus analyses, they found that distributional information was a highly reliable cue for high frequency words, but that reliability reduced for lower frequency words. In addition, the reliability of phonological information was highest for the low frequency words. In an artificial language experiment, Monaghan et al. (2005) found that phonological information provided most assistance to categorizing words that occurred with low frequency in the language.

The presence of marked phonological structure in the low frequency category can in principle provide exactly the desirable phonological cues and thus ease the process of decoding morphological forms. As a caveat, one should distinguish between perceptual difficulty of individual sounds or phonological structures and the perception of the output words in which they occur, although the presence of rich cues in phonological constituents can aid the overall perception of the word.

A second argument for a possible advantage of marked phonological structure in perception comes from salience effects. Perceptual salience has been long recognized to play an important role in shaping phonological inventories
(Liljencrants and Lindblom 1972, Ohala 1983, 1990, Lindblom 1986, Lindblom and Maddieson 1988 etc.) and in phonological processes (Kohler 1990, 1991, Lindblom 1990a). In particular, it has been noted that coronals and laryngeals are less perceptually salient than labials and velars, which are also more marked (Jun 1995, 2004; Hamilton 1996, Hume et al. 1999). In the case of the coronal place of articulation, the relative unmarkedness of coronals as compared to labials and dorsals (testable by higher susceptibility to place assimilation, epenthesis, occurrence in inventories) was shown to follow from poorer acoustic cues in coda position (Byrd 1992). A similar argument can be made for vowel length. Long vowels are more salient than short vowels, at least when used as a perceptual cue for word-final stop voicing (Wang and Wu 2001). As Keren Rice (p.c.) points out, acoustic cues correlate with confusability, in that the likelihood of mishearing a labial consonant, for instance, is less than for an (unmarked) coronal one. Now if the marked structure is associated with the marked, less frequent grammatical category, whose overall salience is relatively low due to low frequency of occurrence, this particular distribution can in principle have a beneficial effect for the perception of the output inflected for the relevant morpho-syntactic category.

It should be noted, however, that the relation between frequency of occurrence and perceptual salience is not yet fully understood. Although high token frequency of grammatical categories is generally assumed to correlate with higher perceptual salience, Kerswill and Williams (2002) report that some of the less well represented forms used by adolescents in a dialect study project were
nevertheless salient for them without necessarily involving particular phonological structure. Also, Hoffman (2004) maintains that low frequency complex prepositions can be both cognitively salient and involved in language change, in particular in processes of grammaticalization. This suggests that the impact of the relative infrequency of syntactic forms on their salience and susceptibility to language change requires further investigation.

As suggested by Todd Haskell (p.c.), another dimension of perception for which frequency and phonological markedness effects can be relevant is segmentation of the speech stream into words. Several researchers have looked at the potential role of transitional probabilities in helping children segment the speech stream into words (e.g. Harris 1955, Hayes and Clark 1970, Motley and Baars 1975, Morton and Long 1976, Goodsitt et al. 1993, Saffran et al. 1996, 1997, Aslin et al. 1998 etc.).

In general, for any two phonemes A and B that occur consecutively in the speech stream, transitional probability is the likelihood that, when having heard A, the next phoneme will be B. The idea is that this probability will be higher within words than between words. Naturally, this probability will also depend on the frequency of occurrence of the B phoneme. All other things being equal, a B phoneme of lower frequency will mean a lower transitional probability. If the B phoneme (or, more generally, phonological structure B) is marked, it will most likely have a lower frequency. So the occurrence of a marked phoneme could serve as some sort of indicator for a good place to try to divide the speech stream.
up into morphemes. In general, the intuitive idea is that the more a child hears something and the more frequently a given phonological structure occurs in the same place, the more likely the child will be to correctly segment the material and subsequently produce it. High frequency forms with more entrenched structure do therefore present a relative advantage over less frequent forms with more marked, less predictable structure. On the other hand, due to the fact that most inflectional processes are affixal in nature and affect word edges, transitional probability may be beneficial in singling out low frequency inflected words belonging to marked grammatical categories and thus compensate for their lower frequency of occurrence.

Word frequency and age of acquisition effects in recognition and recall have also been investigated. Using the so-called ‘remember-know’ procedure developed by Gardiner (1988), Dewhurst et al. (1998) found that recognition performance was higher for low-frequency words than for high-frequency words and higher for late-acquired words than for early-acquired words in ‘remember’ responses. Advantages for both low and high frequency words were found when the items were presented in mixed lists. The authors attribute the findings to the more distinctive encoding of low-frequency and late-acquired words. One can hypothesize that when marked phonological material is associated to low
frequency rather than high frequency words, the distinctiveness of low frequency words is enhanced.\footnote{To complicate the matter even more, it should be mentioned that word frequency should be understood in relation to phonological neighborhood effects. For example, Metsala (1997) found that school-age children were better able to recognize low frequency words with large neighborhoods than low frequency words with small neighborhoods in a gated recognition task. See also Ussishkin and Wedel (2002), cited in §4.2, for an application of neighborhood density as a functional factor in root-affix asymmetries.}

To sum up, in §5 I examined possible mechanisms of functional grounding for MIM effects. It appears that from the perspective of speech production confining marked phonological structure to less frequent, grammatically marked forms represents an advantage, since it leads to minimization of overall articulatory effort on the speaker’s part. A connection was made with Hawkins’s (2003, 2004) principle of form minimization (MiF), which opens the door to considering MIM effects as language use effects encoded in the grammar.

If from the point of view of production there is an obvious advantage of assigning less marked structure to more frequent categories, things are less clear if we consider the other side of the coin, namely perception, and also aspects of retrieval. The review of the psycholinguistic literature undertaken in this chapter seems to suggest, nevertheless, that there may be advantages to having MIM effects in natural languages. The same remark can be made to some extent with respect to language acquisition, where the phonological cues provided by marked structure have been shown to aid segmentation and increase distinguishability.
5.3. The locus and emergence of MIM effects

If it is true that functional factors can shape grammars, there are two essential issues to be addressed in relation to the way in which this influence is exerted. First, there is the issue of the exact locus where functional forces operate. Second, and equally important, is the issue of how patterns become grammaticalized, in other words, with direct application to the effects discussed in this dissertation, how MIM patterns emerge. In this section I approach the two issues mentioned above.

With respect to the locus where functional factors (specifically, frequency of occurrence of inflected forms) operate, there are, in principle, two possibilities. Starting from the general assumption made in Optimality Theory that individual grammars are particular rankings constructed from constraints in CON, one way in which functional factors influence grammars is via the constraints as such; for example, only those constraints which are functionally grounded are allowed to be part of CON. In this dissertation I propose a class of formal licensing constraints LICENSE(M,g) which license marked phonological structure M in outputs inflected for a marked member g of grammatical category G (see Chapter 3 §4 for the definition of the constraint). While I do not assume that all OT constraints are functionally grounded, I claim that the licensing constraint LICENSE(M,g), which is part of the MIM Schema, is subject to functional grounding\(^{35}\). This is seen clearly

---

\(^{35}\) For a general discussion of functional grounding of OT constraints, see Chapter 3 §1.1. For the functional grounding of the licensing constraint LICENSE(M,g), see §4 of the same chapter. The MIM Schema and its factorial typology are discussed in §5.
in the filtering mechanism which is assumed to allow only those licensing
constraints as part of CON for which, for a given marked phonological structure M,
g is a marked of grammatical category G. It is the same filtering mechanism
proposed by Smith (2002/2005) and discussed in Chapter 3 §4, but this time the
functional factor which is active in determining the range of possible constraints is
frequency of occurrence. Licensing constraints in which the value of the
grammatical category is an unmarked one are excluded from CON. As we shall see
in §5 of Chapter 3 and also from the illustrations in Chapter 5, the factorial
typology constructed on the basis of LICENSE(M,g) makes the right empirical
predictions. The essential prediction is that under general phonological similarity
of morphological marking the unmarked member g’ of G cannot license marked
phonological structure M which the marked member g does not. Had a licensing
constraint for an unmarked member g’ of G been allowed as part of CON, the
factorial typology would have predicted languages in which M occurs in g’, but
not in g. Such systems do not seem to exist, at least if we take into account the
similarity condition between g and g’ with respect to the way in which the two are
expressed phonologically.

A second conceivable possibility is that functional factors manifest
themselves by determining constraint rankings proper, for instance, as universally
fixed rankings between (some of) the constraints involved. The linguistic data
available for this thesis do not lend empirical support to such a mechanism. As
shown in Chapter 3 §5, the following patterns are predicted to exist: (a) Marked in
the Marked (M in g, but not in the unmarked category), (b) Full contrast (M both inside and outside g) and (c) Lack of variation (M prohibited, irrespective of grammatical category value). Given this diversity of linguistic patterns, among which MIM is just one particular case, there is no evidence that points in the direction of a universally fixed ranking driven by functional factors. In sum, I claim that with respect to MIM effects grammars are ‘shaped’ by functional factors in an indirect way, as a result of a substantive condition on CON, i.e. via the inventory of constraints universally available, rather than via fixed rankings between constraints.

Having addressed the issue of the locus of the activity of functional factors which determine MIM phenomena, we can now turn our attention to the problem of how such patterns become grammaticalized. There are two aspects to be taken into consideration when examining the emergence of a grammatical pattern, including MIM: language change and acquisition. The rest of this section is devoted to some considerations on the diachronic and synchronic dimensions of MIM patterns.

Within an optimality-theoretic approach to diachrony, it is commonly assumed that language change involves constraint reranking (for work on language change in Optimality Theory see, among others, Jacobs 1995, Bermúdez-Otero 1996, Cho 1998, Green 2001, Oh 2002). As for what constraints undergo promotion or demotion in the reranking process, there is no agreement among specialists. For example, while some researchers claim that language change
involves promotion of markedness constraints\textsuperscript{36} and creates unmarked structure, (Billerey 2000, Gess 2001, Green 2001, Kiparsky 2004 etc.), other researchers highlight the role of markedness demotion in diachronic change, resulting in the emergence of marked structure (Albright 2004, Deo and Sharma 2005, Morin 2005).

What is the contribution the study of MIM effects can make to the understanding of the reranking processes in language change? In Chapter 5 we shall see how Romanian, one of the case studies presented in this dissertation, allows us to shed light on the relation between constraint reranking and the diachronic emergence of MIM patterns, at least for the linguistic system in question.

An interesting question surrounding the diachronic emergence of MIM patterns is whether there are factors which can be said to favor this pattern of linguistic change. While a definitive answer to this question cannot be provided on the basis of the cases considered in this dissertation, one can entertain the possibility that such favoring factors exist. For example, considerations of economy in production favor the confinement of marked phonological structure to output words inflected for a marked grammatical category, under general phonological similarity, as we have already seen in §5.1. Another possible favoring factor is represented by a general tendency in language change noted by

\textsuperscript{36} In this respect, language change would represent the opposite of language acquisition, a process which has been claimed to involve demotion of markedness constraints (Tesar and Smolensky 1993, 1996, 1998, Gnanadesikan 1996, Hayes 2004 etc.)
Andersen (2001). Andersen formulates a principle (dubbed Markedness Agreement) according to which as a change spreads and becomes generalized in a language, “the innovated element is favored first of all in marked environments, if the innovated element is marked, but in unmarked environments if it is unmarked” (Andersen 2001:31). Although in its original statement Markedness Agreement refers to the markedness of syntactic environments as favoring the marked item, one can note the similarity with the Marked in the Marked generalization discussed in this thesis. More precisely, both of them have the effect of establishing a connection between two instances of markedness which converge in an output which is the result of linguistic change. For the Markedness Agreement principle, the two instances of markedness are represented by the marked structure of the output and the marked character of the environment in which it occurs, while in the case of MIM we are dealing with marked phonological structure in an output inflected for a marked grammatical category.

The recurring question in this section has been what mechanism allows patterns favored by language use to become grammaticalized. As shown by Kirby (1994), two kinds of explanations can be given in principle for how language use influences grammaticalization. The first type of explanation is essentially nativist: it may be the case that the language faculty (Chomsky 1988), a property of the human mind to acquire a language of a specific type, has properties that permit the acquisition and grammaticalization of certain structures, possibly those structures which are also favored by use.
The second type of explanation rests on the assumption that a language use (performance) mechanism can account for the structure of the language it processes. As Kirby points out, the nativist explanation has the drawback of being unable to account for issues of distribution of patterns across languages and, as argued by Hoekstra and Kooij (1988), is primarily devised to explain the ability of the language learner to infer grammar starting from an abundance of unstructured linguistic data (Plato’s Problem).

Returning to MIM effects, we are left with the hypothesis according to which they arise from processing preferences that become grammaticalized rather than being innate. Having already explored the emergence of MIM patterns in diachrony, I will conclude this section with a few considerations on first language acquisition. As we have seen in §5.2, there is evidence which can be used to support the idea that the presence of marked phonological structure in outputs inflected for marked grammatical categories may represent an advantage in acquisition. Such arguments include ease of segmentation of the speech stream by L1 learners and an enhanced distinctiveness of low frequency forms. It is for future research to gather further evidence for potential advantages in the acquisition of MIM patterns. From an optimality-theoretic perspective, first language acquisition presupposes an initial state in which markedness dominates faithfulness and it is through successive demotion of markedness constraints that adult language is achieved (Tesar and Smolensky 1993, 1996, 1998, Gnanadesikan 1996, Hayes 2004 etc.)
MIM patterns pose an interesting problem to the markedness demotion approach to acquisition. The MIM Schema involves LICENSE(M,g) as a top-ranked markedness constraint. Two acquisition scenarios can be envisioned for such a situation. First, it may be the case that the licensing constraint never gets demoted in languages where MIM effects exist, to the effect that the occurrence of marked phonological structure M outside marked category g is subject to the same restrictions in the L1 learner and the adult. Second, it may be the case that the licensing constraint is in fact promoted in acquisition. According to Fikkert and Levelt (to appear), such a situation may require a refinement of the notion of ‘markedness constraint’ in the sense that two kinds of markedness constraints should be distinguished, universal and emergent. Of these constraints, emergent ones may get promoted in acquisition, and it may be the case that LICENSE(M,g) is such a markedness constraint. An argument in favor of this behavior may provided by the sensitivity of emergent constraints to frequency patterns, as shown by Fikkert and Levelt, and we have already seen that licensing constraints of the type described in this dissertation are grounded in frequency effects. This issue, together with the acquisition side of MIM patterns in general, merits further investigation.

The correlations made in this chapter do not constitute a full account of the functional grounding of MIM phenomena. Nevertheless, their ultimate causes rest in functional factors and the ability of language use to shape grammars.
6. Conclusion

In this chapter I introduced and reviewed the notions of grammatical (inflectional) and phonological markedness. With respect to grammatical markedness, frequency of usage was shown to constitute main criterion for this kind of markedness. The frequentionist approach to grammatical markedness allows for a successful account of so-called markedness reversal phenomena and, as I claim in this dissertation, for the correlation between grammatical and phonological markedness that I label Marked in the Marked effects.

In phonology, a review of properties associated with markedness shows that its most reliable correlates are complexity (structural and articulatory, and its correlate, perceptual difficulty) and occurrence in phonological inventories. Such criteria form the basis for optimality-theoretic markedness constraints *M that militate against marked phonological structures M.

With respect to the connection between the two kinds of markedness, the Marked in the Marked (MIM) hypothesis was introduced, according to which under similar inflectional coding, marked grammatical categories are more (or equally) prone to sponsor marked phonological structure than unmarked categories. Since grammatical markedness is largely equated with token frequency of outputs, the hypothesis was successfully tested on linguistic objects known for their asymmetric usage frequencies, such as open versus closed class items and roots versus affixes, respectively.
Finally, I discussed the issue of functional grounding of MIM phenomena. As regards the side of language production, MIM effects represent an advantage in minimizing the speaker’s articulatory effort while assuring the distinctiveness of linguistic forms. From the point of view of speech perception, retrieval and acquisition, the picture is less clear, but there is indication that at least in some respects MIM phenomena can have a beneficial effect in facilitating word recognition and aiding speech stream segmentation.

In the following chapter I will discuss how the hypothesized correlation instantiated by MIM is implemented in Optimality Theory (Prince and Smolensky 1993/2004), which is the theoretical framework assumed for the present dissertation.