

PART II

.....

**THEORETICAL
DIMENSIONS OF
LINGUISTIC
TYPOLOGY**

.....

QUP UNCORRECTED PROOF – FIRST PROOF, 8/3/2010, SPi

CHAPTER 7

MARKEDNESS: ICONICITY, ECONOMY, AND FREQUENCY

JOAN BYBEE

1. INTRODUCTION

Markedness is a very important concept in linguistic theory and in cross-linguistic study. Its origins are in the theory of language developed in the Prague School and continued principally by Roman Jakobson. Building on ideas of Nikolai Trubetzkoy, Jakobson used the concept of markedness as the theoretical foundation both for a set of distinctive features for the languages of the world and for a theory of morphology based on the meaning relations within categories. The most important prediction resulting from this theory for morphology is that only the unmarked member of a category may have zero expression. In subsequent developments, many of the details of Jakobson's theory have been dropped, and we now all too often find linguists using the terms 'marked' and 'unmarked' to mean little more than unusual or not expected vs. usual or expected, both within a language and across languages.

When terms become so highly generalized, they lose their value for scientific purposes. Yet the phenomena identified by those studying markedness, in

particular Roman Jakobson and Joseph Greenberg, are real and deserve an explanation. In this chapter we examine these phenomena, which happen to differ somewhat depending upon whether they reside in the phonological, morphosyntactic, or semantic domain. Our goal is to evaluate the proposed explanations for markedness correlations, which include references to diagrammatic iconicity, economy, and frequency of use.

2. THE ORIGINS OF MARKEDNESS THEORY

.....

The concept of markedness was introduced by Trubetzkoy in the context of a large cross-linguistic study of phonological oppositions published originally in 1939. Oppositions in which one member is characterized by the presence of a certain feature and the other by its absence are called ‘privative oppositions’ (Trubetzkoy 1939: 75). In a privative opposition, the member characterized by the presence of a ‘mark’ (such as nasality) is called the ‘marked’ member, and the member characterized by its absence is considered ‘unmarked’. Examples of privative oppositions are ‘voice/voiceless’, ‘nasal/non-nasal’, and ‘round/unround’. The notion of markedness originally did not apply to gradual oppositions, in which features are present to varying degrees (e.g. vowel height), nor to equipollent oppositions, in which both members are logically equivalent (e.g. place of articulation).

Trubetzkoy observed in the languages of his corpus that when an opposition was neutralized in a certain context, it was the unmarked member that appeared. Thus, when a voicing opposition in final position is neutralized in German, it is the voiceless or unmarked member of the opposition that appears. Markedness relations for Trubetzkoy are not necessarily universal: language-specific factors, such as the types of neutralizations and the interception of oppositions, determine markedness. Yet Trubetzkoy argues that markedness has a phonetic basis: the unmarked member is the one whose production requires the least deviation from normal breathing. Trubetzkoy was also interested in phoneme frequency and predicted—based on Zipf (1935)—that the unmarked member of an opposition would be more frequent in continuous speech than the marked member, and that this would be true even if one controls for the fact that neutralization makes the unmarked member more frequent. This prediction is related to the phonetic character of the unmarked member.

Roman Jakobson developed a set of distinctive features based on Trubetzkoy’s notion of a privative opposition (Jakobson, Fant, and Halle 1952, Jakobson and Halle 1956). Jakobson reduced all oppositions to privative (binary) oppositions, and maintained the idea that each feature had a marked and unmarked value,

though the unmarked value could be represented either by a plus or a minus and could be different in different contexts (Jakobson and Waugh 1979). Binary relations with the asymmetry represented by the difference between the unmarked and the marked member in Jakobson's view pervade the structure of language and culture. In particular, Jakobson further developed his theory of markedness in the analysis of morphological systems. In some morphological categories, the unmarked member, which lacks the semantic feature, has a dual function. Battistella paraphrases the statements in Jakobson (1971b[1939], 1971c[1957]), explaining the function of the unmarked member as follows (1990: 2):

The unmarked element thus has two interpretations: it has a general interpretation in which the nonsignalization of the marked feature indicates the irrelevance of the poles of the opposition; and it has a specific interpretation in which the nonsignalization of the marked feature indicates the signalization of the opposite.

In terms of phonology, for the voiceless/voiced contrast in obstruents, this means that in a context of neutralization (e.g. in syllable final position in German), the occurrence of the voiceless obstruent means that voicing is irrelevant. In other contexts, it means that voiceless is truly signalled and contrasts with voicing.

In terms of morphology—for example, for the category of number—the singular form can signal that the number contrast is irrelevant, as in the generic use *The cow is a domestic animal*. This can be referred to as the 'neutral value interpretation' (Croft 2003a). In other contexts, the singular form actually points to a single entity. What is particularly striking is that the category member that is used in such contexts is also the one that is most likely to be marked by a meaningful zero; that is, it goes literally unmarked (Jakobson 1971b[1939]). Jakobson built a morphological theory around this form/meaning correlation. Continuing to use binary features, he analysed inflectional distinctions by decomposing them into sets of marked and unmarked features.

For example, in Jakobson's (1971c[1957]) analysis of the Russian verb, first, second, and third person are represented by two binary features. The first distinction is between 'personal' and 'impersonal', separating the first and second person as participants in the speech event from the third person which is a non-participant. 'Personal' is marked, and the absence of this value is unmarked, reflecting the tendency for languages to use zero expression in third person. The first and second person are distinguished by 'addresser' vs. all other participants, making first person the marked member and second the unmarked one. The analysis continues by designating the past as marked compared to the present, the perfective marked with respect to the unmarked imperfective, and so on.

Jakobson's theory is an excellent example of what is meant by a structural theory: the same structural relations are hypothesized to hold throughout the various domains of language despite differences in the substance (e.g. phonetic, morphological, or semantic).

3. SUCCESSES AND FAILURES OF MARKEDNESS AS A STRUCTURAL THEORY

.....

Markedness theory can be considered successful to the extent that it can be shown that a number of properties correlate with the distinction between unmarked and marked. In its application to morphology, Jakobson cited the semantic correlate: the unmarked member is used to indicate the absence of a feature as well as in cases where the feature is not relevant. In addition, he pointed out that this feature correlates with zero expression. A third property identified by Jakobson is the greater syncretization in the marked member of the category. That is, the marked member of a category may express fewer further morphological distinctions. Thus, in English pronouns, the unmarked singular distinguishes masculine, feminine, and neuter: *he*, *she*, and *it*. In the marked plural, no gender distinctions are expressed; instead, the only pronoun available is *they*.

Two other important properties of the unmarked member of a category have been discussed by Greenberg (1966a). Greenberg points to the lesser degree of morphological irregularity in the marked members of categories. It may seem counterintuitive that unmarked members have more irregularity, but it should be borne in mind that the unmarked forms have greater frequency, which allows them to maintain their irregularity while the marked members are more likely to undergo analogical regularization. As an example, consider the tense/aspect system of Spanish. In terms of tense, the present is unmarked, in opposition to the two pasts—preterite and imperfective. In terms of aspect in the past tense, the preterite is unmarked. The irregularity of paradigms in terms of vowel and consonant alternations in the stem resides almost exclusively in the present and preterite, while the imperfective is regular. Thus, in the present, a large number of verbs have vowel alternations, illustrated by *cuento* ‘tell (PRS.1SG)’ vs. *contamos* ‘tell (PRS.1PL)’; a dozen or so have consonant alternations, exemplified by *tengo* ‘have (PRS.1SG)’ vs. *tenemos* ‘have (PRS.1PL)’. In the preterite, there are more than a dozen verbs whose stem is radically different from the present or infinitive form: *poner* ‘to put’ vs. *puse* ‘put (PRET.1SG)’; *tener* ‘to have’ vs. *tuve* ‘have (PRET.1SG)’. In contrast, the imperfective stem is completely regular, with the possible exception of the verb *ser* ‘to be’, whose imperfective stem is *er-*. There are no changes to the stem within the imperfective and, with the exception just mentioned, there is no change from the infinitive stem.

The second property that Greenberg notes and explores in detail is the higher token or text frequency of the unmarked member of a category. Greenberg counts texts in Sanskrit, Latin, Russian, and French and finds that in nouns the singular is more frequent than the plural (at about a 3:1 ratio) and the dual (in Sanskrit) is much less frequent than the plural. He also finds singular pronouns more frequent than plurals

and first person more frequent than other forms. He finds direct cases of nouns more frequent than oblique cases, the positive form of the adjective more frequent than the comparative or superlative, cardinal numbers more frequent than ordinal numbers, the active form of the verb more frequent than the passive, the indicative more frequent than other moods, and the present more frequent than other tenses.

To the list of markedness criteria as composed by Greenberg, we can add the strong tendency for children to acquire the unmarked member of a category and use it for both members until the marked member is acquired (Dressler, Mayerthaler, Pangel, and Wurzel 1987). Thus, English-speaking children use singular nouns before plurals and present tense verbs before past tense.

In addition, it should be noted that when alternations in inflectional forms are levelled, it is the unmarked member whose form survives and replaces the form of the marked member. Thus, levelling of *cow*, *kine* to *cow*, *cows* favours the singular, and levelling of *weep*, *wept* to *weep*, *weaped* favours the present (Mańczak 1980, Bybee 1985).

The correlation of these properties (i.e. neutral value, zero expression, syncretization, irregularity, frequency of use, order of acquisition, and direction of analogical levelling)—if confirmed empirically and shown to be universal—presents an impressive set of relations that should be accounted for and explained in any linguistic theory. However, some problems do arise with the application of these properties to categories beyond those that constitute the best examples.

First, there are general problems with binarity itself (see the discussion of phonology in section 5), which does not seem to apply equally well to all categories. The attempt to apply binary features to categories with more than two members—such as first, second, and third person as discussed above—seems artificial. Second person is not the absence of first and third person. Nor is first person more marked because it is more specific. Indeed, considering its frequency and the fact that it has zero marking in some languages and serves as the basis for analogy in some cases, the indicators for first person seem to be mixed (Bybee and Brewer 1980).

The use of binary features for the meanings of grammatical categories depends upon being able to find a single abstract meaning for each morpheme. Current functional approaches to linguistic analysis are less likely to set the goal of arriving at one abstract meaning for a morpheme, and also less likely to view meaning as resulting only from oppositions in which one member has what the other lacks. Rather, recent work in grammaticalization emphasizes the polysemy of grammatical morphemes as well as their inherent content, which comes from their earlier lexical sources (Bybee, Perkins, and Pagliuca 1994). For instance, the binary analysis that Comrie (1976a) proposes for aspect characterizes the progressive as lacking in perfectivity and also lacking in habituality. This suggests that the progressive is a sort of default, but nothing could be further from the truth: progressives in English, Spanish, and many other languages express rather specific meaning of an agentive actively engaged in activity, in some cases in a specific location (Bybee et al. 1994).

In addition, there are many exceptions to the way the properties apply to individual categories. The semantic criterion works nicely for gender—a *poet* can be male or female, but a *poetess* must be female—and for number, but it is less clear how this criterion applies to aspect or mood. In a language with inflectional aspect, are there cases in which aspect is neutralized and one or the other is used? In a language with a subjunctive mood, the distinction between moods is neutralized in certain subordinate clauses, but it is the marked subjunctive that appears, not the indicative. This type of situation has led some linguists to propose ‘markedness assimilation’, in which a marked value occurs in a marked context, further weakening the general theory (Shapiro 1972, Battistella 1990).

Other counterexamples include the Spanish imperfective, which syncretizes the distinction between first and third person in the singular, but not in the plural. In Dutch, the first person rather than the third has zero expression. In addition, in categories—such as aspect—which interact strongly with the meanings of verbs, zeros signal different meanings according to the type of verb: in some languages, an unmarked stative verb signals present/imperfective, while an unmarked dynamic verb signals perfective (Bybee 1994: 251). Thus, it seems that the structural relations among members of categories are affected by the substance of the categories and are not all the same.

As Haiman (1985) points out, familiarity (or frequency) in some cases works against the semantic criterion. Frequency of the first person singular explains why it is often zero marked, first acquired, and the basis of analogy for other forms. Particular lexical items also show the effect of frequency of use. Tiersma (1982) shows that in Frisian the levelling of alternations in the stem of singular and plural nouns favours the singular for items that are more individuated in experience, but favours the plural in items that tend to occur in pairs or groups—horns, stairs, tears, geese, and so on.

These problems suggest that there is no one overarching explanation for all asymmetries within categories, but rather, a combination of considerations—frequency of use, diachronic source, semantics—produces the tendencies identified under markedness theory in morphology. We return in section 8 to a discussion of possible explanations for these tendencies.

4. EXPANSIONS OF MARKEDNESS

.....

Jakobson’s theory has subsequently been expanded in a number of ways. While markedness relations were originally proposed to hold only among members of the same categories whose distinctions were expressed only in binary relations, later we find the following developments.

First, markedness relations are stated in degrees; for example, singular is less marked than plural and plural is less marked than dual (Greenberg 1966a, Croft 1990). In Naturalness Theory, Preference Laws give relative markedness values to linguistic structures (Vennemann 1988), and hierarchies—such as the Animacy Hierarchy—are recast in markedness terms (Mayerthaler 1987).

Second, Greenberg (1966c) formulates certain relations among categories as implicational universals—such as the fact that in agreement systems, gender is more marked than number. Croft (1990: 92–3) interprets this as a markedness relation. In addition, Croft (1990) interprets word order co-occurrences as marked or unmarked.

Third, markedness has been applied to the contexts in which feature values occur; as in ‘markedness assimilation’, mentioned above, or in phonology, where, for instance, one could view the syllable coda as a marked position for obstruents.

Fourth, markedness has been applied to relations among non-opposing forms of expression, as in Wurzel’s (1998) proposed relations among affixing, stem-modifying, and subtractive morphology.

With this expanded set of functions for the notion of markedness, it is easy to see how the term ‘unmarked’ has come to mean very little more than ‘natural’, ‘normal’, ‘frequent’, or ‘expected’. In fact, Haspelmath (2006) suggests that the terms ‘marked’ and ‘unmarked’ should and can be avoided altogether, with one of the terms above being used instead to explain what is truly meant. In contrast, with changes in linguistic theory, those who find a role for markedness view it as a means of evaluating grammatical structures (Battistella 1990) with regard to their deviation from the innate ‘core’ (Chomsky 1981, Hyams 1986) or in terms of how much stress they put on linguistic processing capabilities (Wurzel 1998). Observe that both Trubetzkoy and Jakobson saw markedness relations as inherent to language structure and not outside of it.

5. MARKEDNESS IN PHONOLOGY

.....

As the concept of markedness has evolved in phonology, it is closely tied to the notion of naturalness and most often considered to have a phonetic explanation. Thus, in current practice, there is no expectation of a correspondence between markedness in phonology and markedness in morphology or syntax.

It was mentioned above that Jakobson came to see all the features of language as binary and exhibiting markedness relations. Binary distinctive features are also proposed in Chomsky and Halle (1968), but in the theory developed in that book, which became the basis of decades of work in generative phonology, markedness or naturalness considerations were not built directly into the theory. The authors

recognized this lacuna and paid homage to markedness by appending a chapter containing rules that stated markedness relations. However, it is not made clear upon what basis these markedness conventions are proposed, since the empirical evidence for them is not cited. Chomsky and Halle (1968: 411) do say, however, that systems made simpler by the application of universal marking conventions ‘will be more generally found among the languages of the world, will be more likely to develop through historical change, etc.’.

A general problem that arises when markedness relations are considered in sequential context is that some distributions that are patently ‘natural’, such as the voicing of intervocalic obstruents, produce marked segments. There are various proposals for two types of constraints: one which affects segments or oppositions in the phonemic inventory, in which context is irrelevant, and another which predicts outcomes in specific phonetic environments (Greenberg 1966a: 64, Bailey 1973, Stampe 1973). The first type is the classical markedness relation, since it deals with oppositions and is sometimes considered to have acoustic/perceptual motivation, as Jakobson proposed. The second type deals with natural processes, such as assimilation and lenition, and usually refers to articulation for its explanation. One could legitimately view these as competing trends.

Many marking and naturalness relations have been worked out by the proponents of Naturalness theories: Stampe (1973), Bailey (1973), Dressler et al. (1987), and Vennemann (1988). In these theories, naturalness in different domains can produce different results. Besides markedness relations among segments, various proposals about the naturalness of syllables have been worked out. For instance, Vennemann proposes several Preference Laws for syllable structure which are graded relations, such as the Coda Law:

A syllable coda is the more preferred: (a) the smaller the number of speech sounds in the coda, (b) the less the Consonantal Strength of its offset, and (c) the more sharply the Consonantal Strength drops from the offset toward the Consonantal Strength of the preceding syllable nucleus. (Vennemann 1988: 21)

Vennemann is able to show that a number of diachronic changes conspire to produce the patterns described by the Law. In *Natural Phonology*, it is proposed that diachronic change is always towards a more natural or less marked state. Marked structures arise because some ‘Preference Laws’ may be in conflict with one another (Bailey 1973, Vennemann 1993).

Unfortunately, this approach is methodologically circular: if a change produces a more marked syllable, then it is said not to be a syllable-structure change, but rather it is said to have some other motivation. In addition, this theory of change involves an undesirable teleology by proposing that languages are trying to become more natural, less marked (Dressler 1990).

Optimality Theory is quite similar to Naturalness Theory, though the proponents of OT rarely credit these earlier proposals for violable, interacting constraints

and have proposed no theory of linguistic change. In OT, the evaluative function of markedness, which is ordinarily applied at the level of the language, to phoneme inventories, rules, etc., is applied at the level of the individual form, evaluating each possible form for its relative markedness. Markedness constraints in OT are violable, and violations are handled through the language-specific rankings of constraints. The goal of this theory appears to be the specification of a universal set of constraints and their usual rankings across languages. The constraints are universal, though they can be violated, and considered to be an innate part of the language mechanisms by some practitioners (Prince and Smolensky 1997) while at the same time having phonetic motivation (Hayes 1999).

Blevins (2004: 241–4) points out a number of problems with Optimality Theory, one of which is that this theory cannot explain why, for instance, ‘constraints, like that prohibiting (syllable) coda voicing, are typically resolved by constraint rankings which result in devoicing or voice assimilation, but not by rankings which involve vowel epenthesis or metathesis’ (p. 241). Her point is that a phonological theory needs not just a theory of markedness, but also a theory of how unmarked structures arise diachronically (see below).

Another approach, based firmly on cross-linguistic evidence and phonetics, postulates that segments in a phoneme inventory can be classified as basic, elaborated, or complex. Whereas Jakobson and Chomsky and Halle would find markedness relations among even the most basic of consonants and vowels, this approach designates a set of basic segments that are commonly found in the languages of the world, and notably in languages with very small inventories, and shows that the more phonemes included in the inventory, the more elaborated and complex segments it will contain. The basic consonant inventory contains voiceless and voiced stops at three points of articulation ([p t k b d g]), voiceless fricatives ([f s h]), one voiceless affricate ([tʃ], glottal stop, three nasals ([m n ŋ]), [r], [l], [w], and [j] (Lindblom and Maddieson 1988, Lindblom, MacNeilage, and Studert-Kennedy 1984). The basic vowel system contains the five vowels ([a i e o u]). No finer distinctions of markedness are made within these inventories, but if languages contain additional consonants or vowels, these will further fractionate the phonetic space (e.g., by adding points of articulation) or be produced with secondary features, such as palatalization, rounding, glottalization, aspiration, etc. The largest inventories contain segments that combine such secondary features, yielding highly complex segments, such as a glottalized, lateralized [t]. Lindblom and Maddieson (1988) as well as Lindblom et al. (1984) argue that both perceptual and articulatory factors interact to produce the basic inventory: distribution in maximal perceptual space is modulated by articulatory effort. Willerman (1994) further explicates the articulatory parameters.

Greenberg uses Implicational Universals to state markedness relations. For instance, the statement that the presence of nasal vowel phonemes in a language implies the presence of oral vowels would mean that nasal vowels are marked and oral vowels

unmarked. In Greenberg (1978a), he argues that the explanation for this relation is diachronic in nature, since nasal vowels derive from oral vowels in nasal contexts. He generalizes that the marked features arise as contextual variants of unmarked features, which accounts both for their restricted distribution and lesser frequency and for the fact that the unmarked feature is also always present in the language.

Even in cases where the marked segment type does not conform to these generalizations, it is the diachronic processes that provide the explanation. Thus, he cites the example of long vowels which are marked vis-à-vis short vowels. There are often more long vowel phonemes than short vowel phonemes, but the reason for this is that some long vowels are formed by the contraction of diphthongs or vowel combinations. Greenberg (1978a: 87) concludes,

We see from this and similar examples that it is the nature of the process that gives rise to them that is decisive in producing marked and unmarked status rather than the inherent nature of the features as such.

Bybee (2001) and Blevins (2004), who also take this point of view, argue in addition that the underlying explanation for the sound changes that create markedness relations is phonetic in nature.

6. MARKEDNESS IN MORPHOLOGY AND SYNTAX

.....

We have already discussed how Jakobson developed a theory of morphological markedness. This theory can be applied rather directly to some syntactic phenomena. For instance, it appears to be universal that the positive version of a clause is unmarked while the negative is marked. No languages exist in which the negative has zero expression and the positive requires a marker. It also appears to be universal that the active voice is unmarked while the passive is marked. As Croft (1990: 81) points out, the passive uses more complex expression and also has a more restricted distribution, being used only with certain transitive verbs.

Greenberg's (1966c: appendix III) empirical work has turned up numerous generalizations that can be expressed as implicational universals. Croft (1990: 92–3) interprets these to mean that the marked member implies the presence of the unmarked member. Some of the statements found in this work involve more than two members of categories (e.g. singular, dual, trial, and plural), while others are implications across categories—for example, the statement that in inflected words, the presence of gender implies the presence of number. Implications across categories should probably not be taken to be markedness relations. Implicational universals, useful as they are, cover a wide range of relations between the elements in the statement and are descriptive only. The statements themselves say nothing about

the nature of the relations among the elements, nor are they explanatory in themselves (see Moravcsik, this volume).

Croft (1990: 84) argues that certain cross-linguistic patterns of word order give evidence for markedness. Because the order demonstrative—noun occurs in languages with both relative clause—noun and noun—relative clause, Croft argues that demonstrative—noun is the unmarked order for that constituent. As this is a very different interpretation of markedness, one which denotes ‘more common cross-linguistically’ and not one that contrasts elements within a language, it is not subject to the same types of criteria and explanations as the within-category relations usually discussed under markedness.

That does not mean, however, that markedness cannot apply to word orders, but it seems only appropriate within a single language. For instance, the English word order in which the auxiliary precedes the subject is a marked word order: it is used only in certain special constructions, i.e. questions, negatives, and emphatic statements. The other order, subject—auxiliary, is clearly the unmarked one.

Another extension of markedness in syntax concerns the various hierarchies, such as the Animacy hierarchy, which have been uncovered in recent research. Casting these hierarchies in terms of markedness requires once again the recognition of degrees of markedness, since such hierarchies order multiple elements (Mayerthaler 1987, Croft 1990).

7. SEMANTIC OR COGNITIVE MARKEDNESS

.....

Theories of morphological or syntactic markedness often refer to iconicity with semantic or cognitive markedness, which is also sometimes equated with semantic complexity (Clark and Clark 1978, Haiman 1985, Mayerthaler 1987; cf. Haiman, this volume). A useful point of view, though one that takes us a great distance from Jakobson’s original notions, establishes the human cognitive and physical make-up as the reference point for the unmarked. Thus, Croft (1990: 111–15) discusses the Animacy hierarchy in terms of markedness. This hierarchy, motivated by many grammatical processes, ranks the following (from least marked to most marked):

first, second-person pronouns < third person pronoun < proper names < human common noun < nonhuman animate common noun < inanimate common noun

Several different interpretations of this hierarchy have been proposed. As the name implies, one interpretation is that it represents a semantic cline of animacy (Croft 1990). Mayerthaler (1987: 41) proposes that the more accessible an entity is to the speaker and the more it resembles non-biological properties of the speaker, the less marked it is.

Clark and Clark (1978) discuss a number of markedness relations in terms of human perception and cognition (cf. van der Auwera and Gast, this volume). Basic colour terms, which refer to parts of the colour spectrum favoured by the visual cortex, are named by words that are unmarked or monomorphemic, such as *red* and *blue*. Colours residing in other parts of the spectrum have more complex expression: *dark red*, *sky blue* (Berlin and Kay 1969, Kay and McDaniel 1978). Also in the lexicon, objects can be referred to in many different ways: *an apple* might be *a fruit*, *an object*, *Corinne's lunch*, and so on, but the most useful term for it is *apple* because it is neither too general (as *fruit*) nor too specific (*Golden Delicious apple*). The nouns for this basic level of categorization tend to be morphologically simple, while more specific levels of categorization have more complex expression.

One of the most interesting markedness relations discussed by Greenberg (1966a) is the relation between spatial terms that refer to having dimension or lacking it. In pairs such as *high/low*, *long/short*, *wide/narrow*, *deep/shallow*, *thick/thin*, *far/near*, the term that designates having extent is unmarked, while the term signalling lacking extent is marked. The evidence for this is the fact that if the nominalization for these dimensions uses the stem of one of the terms, it will be the one meaning 'having extent', for example, *height*, *width*, *depth*, *thickness*. Also in neutral questions involving these dimensions, we ask, *How long is it?*, not *How short is it?* Clark and Clark (1978) explain this phenomenon, which is consistent across languages, by saying—for example, for length—that a line remains a line as it gets longer, but as it gets shorter, it will eventually disappear. Thus, the term meaning 'having extent' is more basic or unmarked. Clark and Clark also discuss basic shape terms and kinship terms (following Greenberg 1966a).

For relations that are more grammatical than lexical, Clark and Clark (1978) have a similar explanation. Referring to Greenberg's finding that if a language has expressions that differ in complexity for state, change of state, and cause of change of state, they differ in complexity in that order, Clark and Clark argue that constant states are basic and that a change of state involves added conceptual complexity, and causing a change of state adds a further complexity. They argue that language reflects these cognitive facts.

8. FREQUENCY, ECONOMY, ICONICITY

.....

The explanations for markedness in morphology and syntax become more varied as the notion is extended into more and more domains. Let us here consider explanations for the classical properties associated with markedness. We have already noted that a strong version of structural theory cannot be applied blindly

to every category; it follows that even within the classical properties, explanations might differ according to the categories involved.

First, consider zero expression. The fact in need of explanation is that within each category, there is a certain member that tends to have zero expression across languages, as mentioned above (Greenberg 1966a, Bybee 1985). Jakobson (1971g [1966]) notes the iconicity of the relation between the marked and unmarked form. In this case, it would be an instance of diagrammatic iconicity: relational properties of the form reflect the relational properties of the meaning. He cites the positive, comparative, and superlative of adjectives in Indo-European languages—for example, English *high*, *higher*, *highest* and Latin *altus*, *altior*, *altissimus*—as showing a gradual increase in the number of phonemes to ‘reflect the gradation gamut of the signata’ (Jakobson 1971g[1966]: 414). Singulars and plurals are another example: Jakobson says, ‘the *signans* of the plural tends to echo the meaning of a numeral increment by an increased length of the form’ (p. 414).

Others have generalized over categories by saying that the added morphological complexity of the marked form reflects its added semantic complexity (Vennemann 1972a, Clark and Clark 1978, Mayerthaler 1987). However, there still remain questions about what makes certain category members less complex. Why is it that a noun in the singular is less complex than a noun in the plural, and why would this relation be reversed for nouns that occur more often in the plural? It might be that for most nouns, singular is an inherent part of the meaning: for nouns that stand for entities that are easily individuated and that are frequently referred to as individuals (e.g. *a man*, *a woman*, *a table*, *a dog*), their singular designation is part of the meaning. In order to override the singular, an extra morpheme must be added. Similarly, for verbal aspect, individual verbs have inherent aspect of various sorts, and to override this aspect, additional markers must be present. Because there are many types of lexical aspect and several types of inflectional aspect, markedness relations among aspects are not as consistent cross-linguistically. As mentioned above, in some languages, the zero-marked verb signals past if the verb is dynamic and present if it is stative. However, considering the unmarked meaning to be inherent to the lexical item will not apply to deictic categories such as tense or person.

As we mentioned earlier, frequency of use correlates well with unmarked status, and frequency can certainly be invoked as explanatory. For instance, most nouns are more often used in the singular because when we talk about entities in our experience, we tend to individuate them, referring to them in the singular. Thus, it could be that the frequency with which forms are used influences their markedness relations. Conceptual bundling of singularity with individuated entities provides for economy of expression. Some would indeed argue that languages choose the most economic expression for high-frequency concepts.

Consider the important point made by Haiman (1985): while it is often the case that semantic markedness and semantic complexity are the same, as in compounds

where *room*, *bedroom*, and *master bedroom* are increasingly complex notions and increasingly marked, semantic complexity also interacts with familiarity or frequency. Haiman points out that *mare* and *female hippo* are equally complex, but they are not equally marked; the explanation being the greater familiarity of the first referent over the second. This point also relates to local markedness, as discussed above, wherein nouns that designate entities that tend to occur in pairs or groups may be unmarked—as, for instance, in the English nouns for herding animals, *deer*, *sheep*, *fish*, which are unmarked in both singular and plural. Thus, it appears that frequency or familiarity may be the underlying criterion, since it overrides semantic complexity as the correlate of unmarked status. Hay (2001) finds experimental support for the proposition that multimorphemic forms that are more frequent than their bases are viewed as less complex than those that are less frequent than their bases. (See also Witkowski and Brown 1983 for evidence from markedness reversals.)

But before becoming satisfied with a particular explanation, we need to consider how zeros develop. An explanation for a linguistic phenomenon is only valid if the factor referenced can be shown to have operated in the creation of the phenomenon diachronically (Bybee 1988). Zeros develop diachronically in two ways. Most zeros develop when the opposing member of the category undergoes grammaticization. If a language has no number marking for nouns, a zero-marked singular develops only when the overtly marked plural develops and becomes obligatory; if the overt plural is always used when plural is intended, then by inference, the unmarked noun comes to be interpreted only as singular (García and van Putte 1989). Why is it, then, more common for the overt mark to arise on the plural rather than on both members? The reason is perhaps that, as mentioned above, for most nouns, singular is inherent to the meaning and plural is something that must be specified explicitly. The repeated use of a plural morpheme will lead it to grammaticize and perhaps attain affixal status, while the singular noun continues without explicit number marking. To take another example, in English, the present form of the verb formerly signified present habitual, present progressive, and future. With the development of the future from the modal *will* and the progressive from *be + ing*, the unmarked form of the English verb signals present habitual (Bybee 1994).

A second way that zeros develop occurs less often but provides stronger evidence for the cognitive validity of markedness. In cases documented by Watkins (1962) and Bybee and Brewer (1980), a third person singular verb form with an affix is reinterpreted as zero-marked, leading to the restructuring of a sub-paradigm. For instance, in some dialects of Provençal, the third person singular preterite suffix *-t* (as in *cantét* ‘3s sang’) is taken to represent all the preterite rather than just the third singular, giving rise to a third plural form *cantéten* ‘3p sang’ replacing *cantéren*.

Turning now to explanations for syncretization or neutralization, we find that a simple frequency explanation is sufficient, even if we distinguish between cases

where certain distinctions never developed and those in which a prior distinction is lost. For example, in the marked subjunctive mood, Spanish and French do not distinguish the present from the future, though such a distinction is made in the indicative. When the synthetic future developed in these languages, no subjunctive form grammaticized. In the marked and less frequent members of categories, further distinctions will not be of very high frequency and thus may not develop. Similarly, the distinction made in Latin between the Perfect and the Imperfective in the Subjunctive is not made in Spanish; this would be a case where a distinction in a marked member was lost due to its low frequency.

The preservation of irregularity in unmarked forms is very likely due to their frequency of use. It is well attested that high-frequency paradigms maintain their irregularity more than low frequency ones (Hooper 1976, Mańczak 1980); there is no need to search for a different explanation for the unmarked parts of paradigms resisting regularization. Due to high availability in the linguistic environment, the mental representation of unmarked forms can be very strong and accessible, making them unlikely to change. The same explanation applies to the tendency for the unmarked members to serve as the basis of analogical levelling when it does occur. Their higher frequency and greater accessibility in mental representation allows them to be chosen for the basis of new formations. While it is true that regularization or analogical levelling restores the diagrammatic iconicity by which the form reflects its semantic complexity, the actual mechanism by which such a diagram is restored has to do with the higher frequency of the unmarked form.

Are frequency and economy the same thing? Most discussions of economy or economic motivation begin with references to Zipf's Law, which states that 'high frequency is the cause of small magnitude' (Zipf 1935: 29, Haiman 1985, Croft 1990). Zipf's study of vocabulary in a number of languages revealed that high-frequency words are shorter than low-frequency words. This creates a kind of economy in that the words that are used more often take less effort to produce. DuBois's (1985) slogan 'Grammars code best what speakers do most' is more general but subsumes Zipf's Law. What is missing from these statements is explicit mention of how this pattern in language systems arises. Zipf cites clipping, as when *lab* is formed from *laboratory*, as the major mechanism. Some interpretations of economy treat it as an inherent principle that guides change in language systems (Mańczak 1978a). Such interpretations invoke an unfortunate teleology that makes change seem goal-directed.

In contrast, invocations of frequency of use are not goal-directed. Rather, it is claimed that frequency of use has certain effects on cognitive representation which lead to economy in retrieval and production of linguistic units. High-frequency items are stronger in mental representation and thus easier to access, making them more available to either resist change or serve as its basis (Bybee 1985, 1995). In production, repetition leads to the automatization of neuromotor routines; further

repetition leads to the reduction and overlapping of articulatory gestures which shorten the duration of the sequence of gestures that make up a high-frequency word (Browman and Goldstein 1992, Mowrey and Pagliuca 1995). Thus, the increase in efficiency in high-frequency words results from the way the general neuromotor system operates, and is neither restricted to language nor a conscious goal-directed process.

9. CONCLUSION: STRUCTURAL VS. EMERGENTIST VIEW OF LANGUAGE

Markedness was originally proposed in a purely structuralist context, as a generalization over numerous relationships between members of categories. Given the importance of structure, the attempt to analyse all such relationships as governed by one structural principle was an admirable goal. In structural theories, issues of language use are not considered important. Usually, issues of meaning are also left aside (Chomsky 1957), but in the case of Jakobson's theory of markedness, meaning was viewed as structured in the same way as form.

Linguistic theory and practice, especially in typological studies, is no longer so cleanly structuralist. It is recognized that not all linguistic categories have the same structure. Rather, it is not controversial to attend to the substance of the category in trying to understand its structure: thus, it is recognized that some phonological features may not be truly binary, many categories contain scalar relations, and so on. In addition, due to the work of Greenberg, the way language structure is created and changed in diachrony has become an important factor in explaining synchronic structure. Also because of Greenberg's contributions, the role of language use is now considered in formulating explanations for linguistic phenomena. Current theories of language are more emergentist: certain mechanisms of change are postulated; the substance of language (including the phonetic, morphological, and semantic) is subjected to repeated applications of these mechanisms as speakers and hearers use language; and as a result, what we know of as structure emerges in the linguistic material. Both iconicity and economy arise in this way, as do the structural phenomena associated with markedness.

FURTHER READING

- BATTISTELLA, E. L. (1990). *Markedness: The Evaluative Superstructure of Language*. Albany: State University of New York Press.
- BYBEE, J. L. (1985). *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: Benjamins, ch. 3.
- CLARK, H. H., and CLARK, E. V. (1978). 'Universals, Relativity and Language Processing', in J. H. Greenberg, C. A. Ferguson, and E. A. Moravcsik (eds.), *Universals of Human Language*, vol. 1 Stanford, CA: Stanford University Press, 225–77.
- CROFT, W. (1990). *Typology and Universals*. Cambridge: Cambridge University Press, ch. 4.
- (2003). *Typology and Universals*. 2nd edn. Cambridge: Cambridge University Press, ch. 4.
- GREENBERG, J. H. (1966). *Language Universals: with Special Reference to Feature Hierarchies*. The Hague: Mouton.
- JAKOBSON, R. (1971 [1957]). 'Shifters, Verbal Categories, and the Russian Verb', repr. in *Roman Jakobson: Selected Writings*, vol. 2. The Hague: Mouton, 130–47.
- (1971 [1966]). 'Quest for the Essence of Language', *Diogenes* 51, repr. in *Roman Jakobson: Selected Writings*, vol. 2. The Hague: Mouton, 345–59.
- MAYERHALER, W. (1987). 'System-Independent Morphological Naturalness', in W.U. Dressler, W. Mayerthaler, O. Pangel, and W.U. Wurzel, *Leitmotifs in Natural Morphology*. Amsterdam: Benjamins, 25–58.