Phonological length of number marking morphemes in the framework of typological markedness[1]

JANA BRUNNER

Abstract: The study tests the hypothesis that morphemes marking the typologically less marked member of the conceptual category of number should have a phonological length that is less than or equal to that of morphemes marking a typologically more marked member of the same category. The hypothesis is tested for morphemes marking number in nouns, verbs, adjectives and pronouns. The cross-linguistic quantitative count includes 42 languages, 17 of them exhibiting a singular-plural-dual contrast, the others having a singular-plural contrast only. In order to measure phonological length the phonemes making up a morpheme are counted. The hypothesis is verified: Apart from rare exceptions, singular markers have for all the four parts of speech a length that is less than or equal to that of the corresponding plural markers. In languages with dual markers these are longer than or equally long as the corresponding plural and singular markers.

1   Introduction

The concept of markedness was first developed in the Prague School of linguistic theory (Trubetzkoy, 1931, 1969; Jakobson, 1932, 1939) where it described a language internal property of category values.

A marked and an unmarked member of a category, for example the category of number, are in *privative opposition*: one category value (e.g. singular for the category of number) is characterized by the absence, the other by the presence of a mark (e.g. plural). When the concept of markedness was taken over into typology by Greenberg in 1966 this involved important changes. Typological markedness is now no longer a property of a language specific category but a universal property of a conceptual category, and it is not necessarily binary (Croft, 1996, p.17). Typological markedness would thus not claim that, within the category of number, the singular is unmarked in one or the other language, but that crosslinguistically the value singular is the unmarked value of the conceptual category of number, plural is more marked and other values such as dual are even more marked.

Croft (2003, ch.4) sets up a number of *criteria of* (typological) *markedness*, i.e. *structural coding, behavioral potential* and *text frequency*. According to the first criterion, the typologically marked value of a grammatical category will be expressed by at least as many morphemes as the typologically unmarked value of that category (Croft, 2003, p.92). As a consequence of that, for the category of number, one logically possible configuration, i.e. overtly marked singular and zero-coded plural, is excluded by the markedness criterion of structural coding. And cross-linguistic comparison shows that this type is indeed very rare (Croft, 2003, p.89).

The second criterion of markedness, *behavioral potential* describes the "grammatical versatility" (Croft, 2003, p.95) of values of a category. Behavioral potential can manifest itself in the number of morphological categorical distinctions that a particular category value possesses: The typologically unmarked value of a category will have at least as many formal distinctions in a given inflectional paradigm as the marked value (Croft, 2003, p.97). To give an example from

English, there is a gender distinction in singular personal pronouns but not in plural personal pronouns, suggesting that the singular is less marked.

Text frequency presents the third criterion and at the same time the explanatory variable of typological markedness. The tokens of the unmarked value of a category occur at least as frequently as the tokens of the marked value (Croft, 2003, p.110). Greenberg (1966, p.32,37) demonstrates the greater frequency of the singular as compared to (the dual and) the plural by a quantitative count of nouns and verbs in written speech. Following this analysis the singular is in all languages the most frequently occurring value of the category. In Sanskrit, which has a dual marker, this latter marker occurs least frequently.

According to Croft (2003, ch.4), the criteria of markedness are not independent from each other, they are the result of the competing motivations of *economy* and *iconicity*. Economy suggests that expressions should be minimized where possible, whereas iconicity proposes that the structure of language should reflect the structure of human experience. This means that there should be a one-to-one mapping of conceptual values onto form (i.e. one marker for each conceptual value). Again applied to the category of number this means that overt marking for both singular and plural is maximally iconically motivated but minimally economically motivated since it reflects the structure of experience but demands a high degree of articulatory effort. On the other hand, the other three types (zero marking for one or both category values) are economically motivated but not iconically. Zero marking for both singular and plural demands a low degree of articulatory effort (therefore this type is economic); however, there are no markers in language which reflect the two category values which are experienced in the world (therefore this type is not iconic). The two asymmetric types, zero-

marking for one member of the category and overt marking for the other member are not iconic either, but moderately economic. This, however, does not explain why one asymmetric type (overt marking of singular, zero marking of plural) is so rare and the other one so frequent. Looking at the frequency of the category values, however, can explain this pattern: because the singular is much more frequent than the plural it is actually highly uneconomic to use overt marking for the singular and zero marking for the plural; it is much more economic to do it the other way round. The difference in frequency thus explains the difference in structural coding.

Other approaches have questioned the role of iconicity in the formation of linguistic structure (Haspelmath, 2008a), and even claimed that the concept of typological markedness is superfluous (Haspelmath, 2006). Haspelmath (2006) claims that the observed phenomena can be explained entirely by language external factors, particularly by frequency of use. The differences in structural coding exist because, for reasons of economy, speakers will either shorten frequent items, or they will use the base form to denote a frequent item and add a further marker for the less frequent form (Haspelmath, 2006, p.48). The differences in singular and plural marking thus come about because speakers either shorten the more frequent form or attach a marker to the less frequent form. Haspelmath claims that the concept of typological markedness is not needed in order to explain these data.

Frequency of use alone can, according to Haspelmath, also explain what has been called "markedness reversal" in the literature on typological markedness; i.e. cases in which the category value which is usually less marked (e.g. singular) occurs less often than the more marked alternative. As an example, Haspelmath (2008b, p.61) discusses singular/plural pairs such as *criterion-criteria*, where the plural is more frequent and therefore shorter than the singular.

4

Haspelmath (2008b, p.59) adds predictability as a factor shaping grammatical structure: If an item is frequent (as most singular forms are) it is usually predictable and thus receives less and less overt coding until it is finally zero coded. If items are rarer (as are plural or dual forms) they keep their overt coding because rarer forms are less predictable and therefore need overt coding.

Given the importance of frequency in shaping overt coding one could assume that a language arrives at the most frequently observed constellation (zero marking for singular and overt marking for plural) by either shortening the more frequent marker of two or more existing markers until it is finally lost for economic reasons. An alternative scenario would be that a plural marker comes into existence to distinguish between singular and plural (for reasons of iconicity or because of a lack of predictability).

Following from the assumption that the very frequent typologically less marked category markers are shortened and finally get lost over time, it has been hypothesized that structural coding would be sensitive to phonological length as well as morphological length: An overtly coded unmarked value of a category should not be phonologically longer than the overtly coded marked value (Dryer, cited in Croft, 2003: 115f). Morphologically longer means that plural forms consist of more morphemes than singular forms; phonologically longer means that they consist of more phonological material (e.g. segments).

Differences in morphological length can come about in the two ways discussed above, either by the creation of a morpheme for the more marked category (possibly due to iconic considerations) or by shortening a frequent morpheme until it is finally lost for reasons of economy. Differences in phonological length can only come about by shortening an existing marker due to economic reasons.

The aim of the present study is to test the hypothesis that typolog-

ically more marked members of the category of number are coded by markers which are at least as long as the markers of the typologically less marked members of the category. It will be investigated whether, for overt coding only, plural morphemes consist of at least as many segments as singular markers and whether dual markers consist of at least as many markers as plural and singular markers. This hypothesis will be tested on number markers on verbs, nouns, adjectives and pronouns in 42 languages from all parts of the world. In contrast to counts including morphological length the present analysis investigates the influence of economy on number marking morphemes separately from the influence of iconicity (which could lead to the creation of markers for one or the other category.

There are a number of earlier studies which have investigated phonological (and usually also morphological) length. Stolz et al. (2006) carried out a study on comitative and instrumental markers. For a sample of 217 languages with distinct markers for comitative and instrumental they found that comitative markers are longer (more segments and/or syllables) than instrumental markers in 62.7% of the cases and shorter in 27.2% of the cases (p. 173f). If the comitative was longer than the instrumental the ratio was on average 1:2 for segments and syllables. If the comitative was shorter than the instrumental the ratio was about 1:1.7.

Stolz et al. (2007) did a similar study on lexical duals in 120 languages. They argue that the definite numeral denoting two elements (engl. *both*) is semantically more complex than the indefinite numeral (engl. *two*), since it not only quantifies the noun, as does the indefinite numeral, but also determines it. Following from that, the definite numeral should also be morphologically and phonologically more complex. They counted the number of morphemes, segments and syllables in indefinite numerals denoting two elements

(engl. *two*) and definite numerals (engl. *both*). The analysis shows that in the overwhelming majority of cases the definite numeral consisted of more segments (76.3%), more syllables (77% of the cases) and more morphemes (60.8%, Stolz et al. (2007, p.171).

Stolz et al. (2010) investigated the markedness relation between comitative/instrumental and abessives. Starting from an analysis of semantic complexity of these markers the hypothesis was set up that the abessive should be the marked category value and should therefore have at least as much structural coding. A count of morphological and phonological length on 212 morphemes in 74 languages (number of morphemes, syllables and segments) showed that the abessive is phonologically and morphologically at least as long as the comitative and instrumental markers. 80% of the abessive morphemes consisted of more segments than the corresponding comitative or instrumental morphemes, 85% had more syllables and 75% had more morphemes (Stolz et al., 2010, ch. 3.2.1.).

2 Methods

2.1 Aim of the study

The aim of this cross-linguistic study is to find out whether typologically less marked members of the category of number are expressed by phonologically shorter or equally long markers than more marked members of the category; more specifically, whether the phonological length of singular markers is less than or equal to that of plural and dual markers and whether the length of plural markers is less than or equal to that of dual markers. Because trial and paucal are extremely rare in the languages of the world and it is consequently very hard to get a sample large enough for this analysis they will be ignored in the present study. Phonological length will be measured

in two different ways: as the number of segments (phonemes) and number of syllables. The method is similar to earlier studies (Stolz et al., 2010), except that, in order to measure phonological length only, morphemes are not counted. The following data will be obtained:

- the percentages of markers of a member of the category of number that are longer, shorter or equally long as compared to markers marking other members of the category (syllables and segments).

- the absolute number of cases in which a marker marking one member of the category is longer, shorter or equally long as compared to markers marking other members of the category (syllables and segments).

- the average difference in phonemes and syllables for cases in which the morphemes are not equally long.

2.2 Language sample

The sample consists of number marking morphemes from 42 geographically distant languages which are not closely related, 17 of them showing a singular/plural/dual contrast and 25 a singular/plural contrast only. Sampling closely followed the method proposed by Dryer (1989, 1992): Only one language from a genus was taken and the languages were apportioned about equally to six large linguistic areas (Dryer, 1992, p.83). Languages were classified following Ruhlen (1991). Table 1 gives the language sample of the present study. The languages are preceded by their genus.

Table 1. Language sample

Africa
1 Semitic: Arabic
2 Gur: Supyire
3 Southern West Atlantic: Kisi
4 Nilotic: Turkana
5 Bantu: Tswana
6 Central Khoisan: Nama
7 Biu-Mandara: Hdi'

Eurasia
8 Turkic: Karachay
9 Ugric: Hungarian
10 Basque: Basque
11 Celtic: Breton
12 Slavic: Russian
13 Dravidian Proper: Malayalam
14 Munda: Santali

South-East Asia and Oceania
15 Tibetic: Athpare
16 Oceanic: Tinrin
17 Burmese-Lolo: Bisu
18 Central Malayo Polynesian: Kambera
19 Sulawesi Austronesian: Tukang Besi
20 Aslian: Semelai
21 Atayalic: Atayal

Australia and New Guinea
22 Gunwinyguan: Wardaman
23 Mabuso: Amele
24 Dani-Kwerba: Lower Grand Valley Dani
25 Nor-Pondo: Yimas
26 Wapei-Palei: Olo
27 Pama-Nyungan: Guugu Yimidhirr
28 Solomons: Lavukaleve

North America
29 Eskimo-Aleut: West Greenlandic
30 Athapaskan-Eyak: Slave (Hare dialect)
31 Coast Salish: Upriver Halkomelem
32 Algonquian: Blackfoot
33 Muskogean: Koasati
34 Aztecan: Classical Nahuatl
35 Mayan: Tzutujil

South America
36 Mura: Wari
37 Peba-Yaguan: Yagua
38 Arawan: Paumari
39 Guaicuruan: Toba
40 Paya: Paya (Pech)
41 Araucanian: Mapudungun
42 Waicari-Qinigua: Kadiwé

2.3 Choice of morphemes and determination of length differences

The term *morpheme* refers to bound morphemes as well as free morphemes. With regard to verbs, nouns and adjectives, only bound morphemes are considered in the present study. The verbal morphemes used in this study mark agreement with the subject or object, the nominal morphemes mark the number of the denoted element, the adjectival morphemes mark agreement with the noun modified by the adjective. Pronominal markers used in this study are all free morphemes. The bound morphemes sometimes called "pronouns" in the grammars which can be attached to verbs as person/number markers would be treated as verbal morphemes.

As stated in the introduction, number marking usually involves zero marking for at least one member of the category. Since the present study does not compare morphological but phonological length, however, only overt markers will be considered. Cases where there is a zero and an overtly marked form (e.g. English number marking on nouns) will not be considered. The reason for this is that when a value is coded by zero length, one cannot differentiate between morphological length and phonological length.

For languages with a singular-dual-plural contrast, if there is zero marking for one value of the category (usually the singular), but overt marking for the other two categories, the two overt markers will be compared and the zero marker will be ignored. An example is given for Arabic (Holes, 1995, p.134, example 1). Here the dual and plural are overtly marked and will be included in the analysis, but the singular is zero marked and will therefore be ignored.

1.
   SG: *mudi:r*
   DU(Nominative): *mudi:ra:ni*

PL(Nominative):   *mudi:ru:na*

Ideally, the morphemes used in this study should only mark number. However, for reasons of economy, those morphemes are very rare. Often, morphemes marking number mark at the same time person, tense or aspect on verbs, or case on nouns. There are only three languages in the sample with number markers only marking number: Tinrin, Bisu and Wardaman. To give just one example, Wardaman has the dual markers *–guya* or *–wuya,* which only mark dual and can be affixed to either noun or verb, and the plural markers *-mulu* and *–bulu* (Merlan, 1994, p.90):

2. *ngayi-ga-n-guya*
   1INDU-take-PRES-DU
   'Let's you and I take it.'

3. *dan-mulu   goyogba-mulu   wurre-mulu   yibiyi-wagbwun*
   this-PL   orphan-PL-ABS[2] child-PL-ABS father-lacking-ABS
   'These are orphans, [they have] no father.'

In contrast to these languages, however, most languages have number markers which mark something else apart from number. If there are more of these morphemes available than are needed for the present study morphemes which mark the unmarked category members of other categories were preferred over morphemes marking a marked member (for example, indicative markers rather than subjunctive markers, markers of subject agreement rather than object agreement, active voice rather than passive). Tzutujil, for example, has nominal number markers which at the same time mark absolute or ergative case and person (table 2). Here the absolutive person markers are used rather than the ergative markers.

Sometimes not a complete paradigm can be considered. In Russian number marking on male nouns, there is zero marking in the nominative and accusative singular. Therefore, only the other cases were

Table 2. Number marking on Tzutujil nouns (Dayley, 1985, p.62)

CASE/PERSON	SINGULAR	PLURAL
ABSOLUTIVE		
1st	*in-*	*og-*
2nd	*at-*	*ix-*
ERGATIVE, PRECONSONANTAL		
1st	*nuu-*	*gaa-*
2nd	*aa-*	*ee-*
3rd	*ruu-*	*kee-*
ERGATIVE, PREVOCALIC		
1st	*w-*	*q-*
2nd	*aaw-*	*eew-*
3rd	*r-*	*k-*

Table 3. Declination of Russian masculine nouns (Morris et al., 1996, p.442)

	SINGULAR	PLURAL
Nominative	*stol* 'table'	*stol-y*
Genitive	*stol-a*	*stol-ov*
Dative	*stol-u*	*stol-am*
Accusative	*stol*	*stol-y*
Instrumental	*stol-om*	*stol-ami*
Prepositional	*stol-e*	*stol-ax*

12

Table 4. Number marking on countable nouns in Turkana (Dimmendaal, 1983, p.210)

	SINGULAR	PLURAL
Feminine	*a-*	ŋ*a-*
Masculine	*e-*	ŋ*i-*
Neuter	*i-*	ŋ*i-*

used (table 3).

Very often one finds more than one singular form for only one plural form. In the Turkana example (table 4) there are only two forms in the plural as opposed to three in the singular. This is because the singular has a greater behavioral potential (see introduction). In those cases all the singular forms are compared with the respective plural forms separately. In the Turkana example this means that the two feminine forms were compared with each other, the masculine singular form was compared with the masculine/neuter plural form and the neuter singular form with the masculine/neuter plural form. This of course applies not only to gender contrasts but also to all other cases in which the greater behavioral potential of a value of the category of number shows up.

As stated above, the category of number can have different numbers of members depending on the language. The singular always denotes one referent and the dual two. The value plural, however, can have different meanings in the different languages. It can determine more than one element, but in dual languages possibly only more than two. In Wardaman the plural which refers to more than one element is called *non-singular*, the plural which denotes more than two elements is called *plural* (table 5).

Those differences are ignored here; *plural* as it is used here includes both meanings. The non-singular in the Wardaman example is treated as plural.

Some languages mark one value of the category of number by an

13

Table 5. Wardaman pronouns, base form (Merlan, 1994, p.122)

SINGULAR		DUAL		PLURAL		NONSINGULAR	
1	*ngayug-bi*	1IN	*yawung-guya-wi*	1IN	*ngarrug-bi*	1EX	*yurrug-bi*
2	*yinyang-bi*	2	*nurrug-(g)uya-wi*	2	*nurrug-bulu-wi*	2	*nurrug-bi*
3	*narnaj-bi*	3	*narnay-guya-wi*	3	*narnay-bulu-wi*		

umlaut. In Olo, for example, one finds plural marking on nouns with affixes and umlaut (McGregor and McGregor, 1982, p.30):

4. *sengke*   'pig'
   songkou   'pigs'

Those cases are ignored since there is no clear boundary between stem and number marker and it is not possible to count differences in phonological length (Croft, 2003, p.94).

Inflectional paradigms often have cases of suppletion. Suppletion is a phenomenon which does not occur equally often in all areas of the world (Veselinova, 2005a,b). Ignoring suppletive forms completely could therefore have led to a bias in our corpus. However, similarly to the cases with umlaut marking, for bound morphemes it is often not possible to determine the boundary between stem and affix in suppletive forms. Therefore, cases with stem suppletion are not included in the sample. For pronominal markers, which are free morphemes, suppletive forms were included because the complete word was counted. For the same reason, number marking by inner modification and suprasegmentals was excluded.

Number can furthermore be marked by reduplication. One could say that in this case phonological material is added to the stem. However, what is added is not an independent morpheme (Croft, 2003, p.95) so its status is questionable. This led to the exclusion of these cases from the present count.

Sometimes the form of a morpheme which is added to a stem depends on the semantics of the word. Those systems can be very

Table 6. Number marking on Kisi nouns (Childs, 1995, p.148)

-ó	singular of all animates, some inanimates
-á	plural of animates
-léŋ	singular of inanimates
-láŋ	plural of inanimates
-é	singular of collective plants
-óŋ	plural of collective grains, etc.
-áŋ	liquids

complex. In these cases the most frequent type(s) were chosen. To illustrate this, an example from Kisi is given (table 6). Each Kisi noun belongs to one of seven classes which are defined semantically. Some classes are singular, some are plural, and some seem to be neither one nor the other. The first column in the following list gives the suffix of the nouns in the class, the second a semantic characterization.

Some of these noun classes can be paired, so that a pair consists of a singular class denoting certain elements and a plural class denoting the same class of elements. Childs (1995) also gives the frequency of the respective pairings (table 7). As one can see, some pairings are very rare and are therefore ignored in the present study. Only the first, second and third pairing are included in the analysis since they occur most frequently. This method differs from Stolz et al. (2010)'s investigation of the markedness relation between comitative/instrumental and abessive. Their sample includes on the one hand languages with different markers for the first two categories but only one for the abessive, and on the other hand languages with only one marker for both comitative and instrumental. For the first group, Stolz et al. selected the shorter marker from the first two categories. Since, however, the aim of the present study was to investigate this very parameter, the length of the markers, it was decided to take all markers except for the very rare ones.

Table 7. Number marking on Kisi nouns (Childs, 1995, p.149)

SINGULAR	PLURAL	PERCENTAGES OF TOTAL PAIRING
-ó	*-láŋ*	43.4%
-ó	*-á*	27.3%
-é	*-óŋ*	15.3%
-léŋ	*-é*	4.3%
-léŋ	*-láŋ*	3.2%
-léŋ	*-áŋ*	3.1%
-ó	*-é*	3.0%
-léŋ	*-óŋ*	0.3%

Following Haspelmath (2006, p.56), frequency can result in articulatory simplicity of the item. As a consequence, articulatorily difficult phonetic segments should occur more often in rarer morphemes than in frequent ones. So the articulatory difficulty of the segments should actually be taken into account in the present study. However, for practical reasons it was decided not to do so because it was impossible to set up a ranking of sounds as to their articulatory difficulty. Whereas most people will agree that certain classes of sounds (such as ejectives) are more complex than others, for the purposes of this study it would have been necessary to decide whether the difference in complexity between two given sounds is greater than the difference between two other given sounds, which is not easily possible. Articulatory complexity certainly influences shortening of the morphemes marking less marked category values, but the present study restricts itself to a discussion of phonological length.

The phonemes which make up a morpheme were counted. A singleton consonant was counted as one unit, as was a monophthong. Vowels with length markers, diphthongs and consonantal geminates were counted as two units. This partly follows Stolz et al. (2010, ch.3.2.1.) who counted diphthongs as two elements. The reasoning behind this is that segment length plays a role in language change,

so one can assume that it plays a role in the process which is under investigation here, namely, whether morphemes marking more marked members of a category are shortened more easily than less marked morphemes of the category.

2.4 Determining phonological length within a language

In most cases not all the parts of speech of a chosen language exhibit overt number marking by suitable morphemes. Therefore very often markers for one, two or three parts of speech only are considered. Moreover, sometimes only very few morphemes are chosen from one part of speech of one language, in other cases very many morphemes are taken because very many suitable morphemes could be found. In order to avoid a biasing of the data due to different numbers of morphemes for different parts of speech or languages phonological length is measured separately first for each language (or genus, since one language represents one genus), second for each area and only finally for all the areas together. The parts of speech are counted separately throughout the whole study.

An example for the count within a language and part of speech will be given. In Russian nominal inflection (masculine, inanimate nouns, see table 3) the nominative and accusative are zero-marked in the singular. Therefore, they are not considered any further. In each of the remaining pairs of morphemes the singular form is one phoneme shorter than the plural. In one case the singular is also one syllable shorter. Furthermore, the average difference in number of segments (or syllables) between the plural and singular forms is 1 and 1, respectively. The result of this count is summarized in table 8.

The first column in this table gives the relationship between two

Table 8. Result for number marking on nouns in Russian. First column: relationship between two members of a category, second column: percentage of cases in which this relationship was found and – in parentheses – absolute number of cases. Third and fourth column: Results for syllables.

	Segments		Syllables	
	Percentage	Average difference	Percentage	Average difference
Sg<Pl	100% (4)	1	25% (1)	1
Sg=Pl	0% (0)		75% (3)	
Sg>Pl	0% (0)		0% (0)	

members of the category. The second and third columns refer to segments, the last two columns to syllables. The second and fourth column give the percentages of cases and – in parenthesis - the absolute number of cases. The third and fifth column show the average difference in the number of segments/syllables.

All the languages from the different genera are now grouped according to the linguistic areas they belong to (see the list of languages table 1). For each part of speech the average percentages of morphemes are counted which are longer, shorter or equally long in one value as compared to the other(s). The numbers of cases are added and the average difference in length is counted.

Finally, a mean over the six linguistic areas is calculated (percentages and numbers of average difference). The numbers of cases are added.

3 Results and discussion

Tables 8 to 12 show the results for each part of speech (verbs, nouns, adjectives, pronouns). In the overwhelming majority of cases the phonological length of the singular is less than or equal to the length of the plural or the dual, and the length of the plural is less than or equal to the one of the dual. The data thus support the hypothesis.

Table 9. Result for verbs. Columns as in table 8.

	Segments		Syllables	
	Percentage	Average difference	Percentage	Average difference
SG<PL	41% (48)	1.3	18% (21)	1
SG=PL	54% (61)		81% (91)	
SG>PL	5% (4)	1	1% (1)	1
SG<DU	59% (19)	1.9	36% (14)	1
SG=DU	41% (14)		64% (19)	
SG>DU	0% (0)		0% (0)	
PL<DU	45% (16)	1.36	10% (8)	1.06
PL=DU	49% (25)		90% (35)	
PL>DU	6% (2)	1	0% (0)	

There are more interesting facts that can be seen from those figures which all show that markedness (and with it the explanatory variable frequency) strongly influences phonological length. The first observation concerns the average difference in length which is overall greater in cases that prove the hypothesis right than in cases that prove it wrong. To give an example, for verbs the average difference for cases in which the length of the singular is less than or equal to that of the plural (predicted case) is 1.3 segments whereas in the unpredicted case (singular longer than plural) it is only 1.0 (table 9). The only exception to that is the average difference in syllables for verbs which is 1.0 for both the predicted and the unpredicted case (table 9). Relating this observation to frequency one can say that if a more frequent form is longer than a less frequent form (unpredicted case) the difference in length is usually smaller than if a less frequent form is longer than a more frequent form (predicted case). The reason for that could again be that more frequent forms are shortened more easily than less frequent forms. This is consistent with Stolz et al. (2006, p.174); they showed that if the less marked instrumental is shorter, it is half as long as the more marked comitative, but

Table 10. As table 9 but for nouns.

	Segments Percentage	Average difference	Syllables Percentage	Average difference
Sɢ<Pʟ	57% (42)	1.62	35% (26)	1.11
Sɢ=Pʟ	36% (26)		59% (43)	
Sɢ>Pʟ	7% (8)	1.57	6% (7)	1
Sɢ<Dᴜ	88% (12)	1.8	75% (12)	1.22
Sɢ=Dᴜ	12% (1)		25% (1)	
Sɢ>Dᴜ	0% (0)		0% (0)	
Pʟ<Dᴜ	59% (7)	1.2	27% (7)	1.0
Pʟ=Dᴜ	41% (13)		73% (13)	
Pʟ>Dᴜ	0% (0)		0% (0)	

if the comitative is shorter, the difference between instrumental and comitative is smaller.

The average difference between singular and dual is greater than the average differences between singular and plural or plural and dual: The mean value of the averages for singular<dual is 1.93 across parts of speech (segments) and 1.12 (syllables). The numbers for singular<plural (1.48/1.04) and plural<dual (1.51/1.11) are lower. The reason for that could be that there is a greater difference in frequency between singular and dual than between singular vs. plural and dual vs. plural.

Looking a little closer at the cases that run counter to the hypothesis one can find even more evidence for the singular<plural<dual hierarchy in typological markedness. Following from the greater inequalities in frequency between singular and dual than between singular vs. plural and dual vs. plural, there should be fewer unpredicted cases in the singular vs. dual comparison than in the singular vs. plural and plural vs. dual comparisons. In fact this tendency can be seen. Whereas in up to 8% for segments and 11% for syllables (tables 12 and 11) the singular is longer than the plural and in up to

Table 11. As table 9 but for adjectives.

	Segments Percentage	Average difference	Syllables Percentage	Average difference
Sg<Pl	27% (9)	1.25	27% (7)	1
Sg=Pl	68% (17)		62% (18)	
Sg>Pl	5% (2)	1	11% (3)	1
Sg<Du	100% (6)	1.88	75% (4)	1
Sg=Du	0% (0)		25% (2)	
Sg>Du	0% (0)		0% (0)	
Pl<Du	83% (6)	1.75	75% (5)	1.17
Pl=Du	17% (2)		25% (3)	
Pl>Du	0% (0)		0% (0)	

Table 12. As table 9 but for pronouns.

	Segments Percentage	Average difference	Syllables Percentage	Average difference
Sg<Pl	53% (67)	1.73	34% (45)	1.03
Sg=Pl	39% (49)		62% (75)	
Sg>Pl	8% (8)	1	4% (4)	1
Sg<Du	75% (30)	2.13	74% (29)	1.27
Sg=Du	23% (13)		26% (16)	
Sg>Du	2% (2)	1	0% (0)	
Pl<Du	51% (20)	1.2	25% (14)	1.21
Pl=Du	32% (22)		75% (39)	
Pl>Du	17% (11)	1.13	0% (0)	

17% (table 12, segments, no cases for syllables) the plural is longer than the dual, there are only two cases (2%, table 12) in which the singular is longer than the dual (looking at segments, not looking at syllables).

From all these observations the following conclusion can be drawn. There seems to be a connection between differences in the phonological length of the forms of two values and differences in markedness. If there is a great difference in markedness (like in the relation singular vs. dual) then there is a great difference in phonological length (average length and percentages of cases). If the difference in markedness is smaller (dual vs. plural and singular vs. plural) then the difference in phonological length is smaller.

Regarding the differences between segments and syllables, there is a tendency to have no difference in syllables but a difference in segments. In other words, typically the plural has more segments than the singular but not more syllables, and the same applies to the relation dual vs. plural. For the relation singular vs. dual the dual has typically both more syllables and more segments. Moreover, the average differences in segments are greater than the differences in syllables.

Looking at the parts of speech the following preferences can be discovered. Nouns and pronouns typically exhibit singular $\leq$ plural (depending on whether one looks at segments or syllables), singular $<$ dual and plural $\leq$ dual, (tables 10 and 12). Verbs and adjectives support the hypothesis less than nouns and pronouns. Verbs have singular = plural, singular $\leq$ dual and plural = dual (table 9). For adjectives the relations are singular = plural, singular $<$ dual and plural $<$ dual (table 11). One could again try to find a reason for that by looking at frequency. Nouns should, following from this result, occur more often in the singular than for example verbs. However, according to the counts given in Greenberg (1966, p.32,37) the

opposite seems to be the case. According to Greenberg, percentages for singular and plural forms in different parts of speech vary across languages, and for three out of four languages the percentages for singular forms of verbs are higher than the ones for nouns. For Latin, for example, Greenberg gives the following figures for singular: 91% of the verbs and 85.2% for nouns.

Pronouns have the highest percentage of cases contradicting the markedness hierarchy (SG>PL: 8/4%, SG>DU: 2/0%, PL>DU 17/0%). In some cases these results can be explained by the existence of several alternatives (inclusive and exclusive) in dual and/or plural. In Hdi' pronouns, for example, there are inclusive and exclusive forms in the plural but not in the dual. The exclusive plural form is longer than the dual form. An explanation for that could be that if there are two alternatives each of those has a lower frequency than a single form would have. So each form will be shortened less than it would have been if it was the only form.

4   Summary

The aim of the present study was to test the hypothesis that typologically more marked (and thus less frequent) markers of the category of number are phonologically longer than or equally long as markers of typlogically less marked (and thus less frequent) members of the category. The results of this study demonstrate that even when there is no difference in morphological length, there is a difference in phonological length between the typologically marked values and typologically unmarked values. Furthermore, the results are in agreement with an explanation of structural coding in terms of economy. There is a greater economic motivation to shorten the more frequent and less marked forms than to shorten the less frequent and more marked forms.

The count included 42 languages, each of them from a different genus. The languages/genera were divided into six large linguistic areas which were analyzed separately first before calculating means over areas.

Number marking in nouns, verbs, adjectives and pronouns was analysed. In the great majority of cases the lengths of morphemes marking a less marked category value were less than or equal to the one of morphemes marking a more marked category member. The results were especially clear for nouns and pronouns. The average difference in length between two values of the category of number is greater in the cases that support the hypothesis than in the few cases that run counter to it. There are more differences in segments than in syllables. Due to greater differences in frequency between singular and dual than between dual and plural or singular and plural, the result is most obvious for the difference between singular and dual. Here only two cases could be found which disprove the hypothesis (i.e. where the dual marker was shorter than the plural marker).

The present analysis shows that support for typological markedness in the category of number cannot only be found in morphological length but also in phonological length. It therefore supports the hypothesis that zero marking can be a result of truncation due to the high frequency of a marker.

The present study concentrated on overt coding and therefore on frequency induced effects, for example shortening of an existing marker of a less or unmarked category value. Cases in which a morpheme marking a more marked category value came into existence for iconic reasons are not the subject of the present study.

In the present study number marking by inner modification and umlaut was excluded for all bound number markers because it would have been impossible to determine the boundary between stem and affix. It is possible that by doing so the effect frequency has on struc-

tural coding was underestimated. This is because coding via umlaut or inner modification only does not lengthen the word and is therefore probably used predominantly in very frequent forms.

Another factor which was not investigated in this study is articulatory complexity of single segments. It is possible that some of the long markers of the less marked values additionally consist of articulatorily complex segments.

5 Acknowledgments

References

Andrews, J. (1975). *Introduction to Classical Nahuatl.* Austin/London: University of Texas Press.

Asher, R. and Kumari, T. (1997). *Malayalam.* Routledge, London/New York.

Bromley, H. (1981). *A grammar of lower Grand Valley Dani.* Dept. of Linguistics, Research School of Pacific Studies, Australian National University.

Carlson, R. (1994). *A grammar of Supyire.* Mouton de Gruyter, Berlin/New York.

Chapman, S. and Derbyshire, D. (1991). *Paumari.*, volume 3, pages 161–352. Mouton de Gruyter, Berlin/New York.

Childs, G. (1995). *A grammar of Kisi: a southern Atlantic language.* Mouton de Gruyter, Berlin/New York.

Cole, D. (1955). *An introduction to Tswana grammar.* London Longmans and Co., London/Cape Town/New York.

Croft, W. (1996). *'Markedness' and 'Universals': From Prague School to Typology.*, pages 15–21. Nodus Publikationen.

Croft, W. (2003). *Typology and Universals.* Cambridge: Cambridge University Press.

Dayley, J. P. (1985). *Tzutujil Grammar.* Linguistics 107. University of California Press.

Dimmendaal, G. J. (1983). *The Turkana Language.* Dordrecht/Cinnaminson: Foris Publications.

Donohue, M. (1999). *A grammar of Tukang Besi.* Mouton de Gruyter, Berlin/New York.

Dryer, M. (1989). Large linguistic areas and language sampling. *Studies in Language*, 13(2):257–292.

Dryer, M. (1992). The Greenbergian word order correlations. *Language*, 68(1):81–138.

Ebert, K. (1997). *A grammar of Athpare.* Lincom Europa, München/Newcastle.

Everett, D. and Kern, B. (1997). *Wari: the Pacaas Novos language of western Brazil.* Routledge, London/New York.

Foley, W. (1991). *The Yimas language of New Guinea.* Standford University Press, Stanford, California.

Fortescue, M. (1984). *West Greenlandic.* Routledge Kegan & Paul.

Frajzyngier, Z. and Shay, E. (2002). *A grammar of Hdi.* Mouton de Gruyter, Berlin/New York.

Frantz, D. G. (1991). *Blackfoot Grammar.* University of Toronto Press, Toronto/Buffalo/London.

Galloway, B. (1993). *A grammar of Upriver Halkomelem.* Linguistics 96. University of California Press.

Greenberg, J. (1966). *Language universals: with special reference to feature hierarchies.* The Hague/Parin: Mouton & Co.

Hagman, R. (1977). *Nama Hottentot Grammar.* Bloomington: Indiana University.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of Linguistics*, 42(01):25–70.

Haspelmath, M. (2008a). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics*, 19(1):1–33.

Haspelmath, M. (2008b). Reply to Haiman and Croft. *Cognitive Linguistics*, 19(1):59–66.

Haviland, J. (1997). *Guugu Yimidhirr.*, volume Volume 1, pages 26–180. Handbook of Australian Languages. Amsterdam: John Benjamins B.V.

Holes, C. (1995). *Modern Arabic.* Longman, London/New York.

Jakobson, R. (1932). *Zur Struktur des russischen Verbums.*, pages 74–84. Charistera Guglielmo Mathesio. Prague.

Jakobson, R. (1939). *Signe zéro.*, pages 143–152. Georg.

Kenesei, I., Vago, R., and Fenyvesi, A. (1998). *Hungarian.* Routledge, London/New York.

Kimball, G. and Abbey, B. (1991). *Koasati grammar.* U of Nebraska Press, Lincoln/London.

Klamer, M. (1998). *A grammar of Kambera.* Mouton de Gruyter, Berlin/New York.

Manelis Klein, H. E. (2001). *Toba.* Lincom Europa, München.

McGregor, D. and McGregor, A. (1982). *Olo language materials.* Dept. of Linguistics, Research School of Pacific Studies, Australian National University.

Merlan, F. (1994). *A grammar of Wardaman.* Mouton de Gruyter, Berlin/New York.

Morris, G. W., Vyatyutnev, M. N., and Vokhmina, L. (1996). *Russian Face to Face. A Communicative Program in Contemporary Russian.* National Textbook Company, Lincolnwood, Illinois.

Müller-Ott, D. (1982). *Russische Grammatik.* Ott-Verlag, Tulln.

Neukom, L. (2001). *Santali.* Lincom Europa.

Osumi, M. (1995). *Tinrin grammar.* Univ of Hawaii Pr, Honolulu.

Payne, D. and Payne, T. (1990). *Yagua.*, volume 2, pages 249–474. Mouton de Gruyter, Berlin/New York.

Press, I. (1986). *A grammar of Modern Breton.* Mouton de Gruyter, Berlin/New York/Amsterdam.

Rau, D.-H. V. (1992). *A grammar of Atayal.* Cornell University.

Roberts, J. (1987). *Amele.* Croom Helm, London/New York/Sydney.

Ruhlen, M. (1991). *A guide to the world's languages: Classification.* Stanford Univ Pr.

Saltarelli, M. (1988). *Basque.* Croom Helm, London/New York/Sydney.

Sandalo, F. (1997). A grammar of Kadiwéu: with special reference to the Polysynthesis Parameter.

Seegmiller, S. (1996). *Karachay.* LINCOM Europa, Munich/Newcastle.

Shixuan, X. (2001). *The Bisu Language.* Lincom Europa, München.

Stolz, T., Khizanishvili, T., and Levkovych, N. (2007). Lexikalische duale. typologische gedanken zu den Übersetzungsäquivalenten von deutsch beide in crosslinguistischer sicht. *Lingua Posnaniensis*, 44:167–196.

Stolz, T., Stroh, C., and Urdze, A. (2006). *On Comitatives and Related Categories: A Typological Study with Special Focus on the Languages of Europe.* Mouton de Gruyter, Berlin, New York.

Stolz, T., Stroh, C., and Urdze, A. (2010). With(out). on the markedness relation between comitatives/instrumentals and abessives.

Terrill, A. (1999). *A grammar of Lavukaleve: A Papuan language of the Solomon Islands.* PhD thesis, Australian National University.

Trubetzkoy, N. (1931). Die phonologischen systeme. *Travaux du Cercle Linguistique de Prague*, 4:96–116.

Trubetzkoy, N. (1939 [1969]). *Principles of phonology.* (Translation of Grundzüge der Phonologie, Prague, 1939). Berkeley and Los Angeles: University of California Press.

Veselinova, L. (2005a). *Suppletion in Tense and Aspect Categores.*, pages 322–325. World Atlas of Language Structures. Oxford: Oxford University Press.

Veselinova, L. (2005b). *Verbal Number and Suppletion.*, pages 326–329. World Atlas of Language Structures. Oxford: Oxford University Press.

Zúñiga, F. (2000). *Mapudungun.* Lincom Europa, München.