

SCALES AND COGNITIVE ECONOMY

Gertraud Fenk-Oczlon and August Fenk

ABSTRACT

Scales or hierarchies are most commonly viewed as determinants of grammatical relations. The respective principle in the domain of differential object marking (DOM): objects ranking higher on a relevant scale – e.g. the animacy scale or the definiteness scale – are more likely to be overtly case marked. In the domain of word order the higher ranking elements are expected to obtain initial positions.

We argue in favour of frequency- and economy-based explanations and present counter-examples to the above mentioned principle in DOM: In Russian, feminine and neuter animate objects in the plural are zero-coded while inanimate are overtly coded. This can be explained by the high frequency of the genitive plural that is used to code animate objects. The counter-examples from German show a determination of DOM by gender: In the singular only masculine direct objects are overtly case marked, irrespective of whether they are animate or inanimate, definite or indefinite. And though females are not less animate than males, the frequency distributions of personal pronouns and definite/demonstrative articles indicate that the feminine tends more often to be patient or object than agent or subject.

Even more cogent are the counter-examples regarding the word order in frozen binomials: In cases of conflicts between the “frequency scale” and scales in the usual sense (human>nonhuman, animate>inanimate), the regularity “more frequent before less frequent” wins (cf. *horse and rider* or *land and people*). To place components with high frequency and low informational content in those initial positions which are per se characterized by high uncertainty provides an even and economic distribution of information over time.

KEYWORDS

Scales, hierarchies, differential object marking, word order, frequency, gender, information processing

1 INTRODUCTION

Hierarchies or scales such as the animacy hierarchy and/or the definiteness hierarchy are said to be determinants of grammatical relations. They are assumed to show effects on grammatical phenomena such as differential case marking (e.g. Comrie 1981, Bossong 1983/1984, Aissen 2003), verb agreement (Comrie 1981), number distinction (e.g. Comrie 1981), constituent order (e.g. Allen 1987), and word order in frozen binomials (Cooper & Ross 1975). Cooper and Ross do not use the term *hierarchy* or *scale*, but their explanations of word order – animate before inanimate, male before female – are already in line with such principles. More recently other possibilities are considered as well (cf. call for papers, Workshop “Scales”, University of Leipzig 2008): Scales might themselves be viewed as a component of grammar or, a second possibility, rather as epiphenomena of frequency distributions and economy principles in language use.

In this paper we discuss the hypothesized influence of scales on differential object marking and word order. Concerning their status in the language system we will argue in favour of frequency- and economy-based explanations. The main explanatory principle applied here will be the concept of an economic flow of linguistic information which has already proved its value as the answer to related questions (e.g. Fenk-Oczlon 2001). All the

following principles contribute to the economy of language and especially the first two of them to a rather even distribution of information over the time:

- *Words used less frequently tend to be longer* (Zipf 1929), approximately proportionate to the longer time needed for processing their higher informational content (Fenk & Fenk 1980). Shortening more frequently used and thus informationally poorer signs contributes to a constant flow of information.
- *More frequently used units tend to be placed before less frequently used units.* As a sentence continues, the remaining words get more and more predictable – the number of possible and plausible continuations decreases, and so does the (subjective) information. In order to achieve a constant flow of information, the element that is more predictable in the actual context is placed in the sequence initial position which is *per se* associated with high uncertainty. Informationally rich elements in this position would overload our information processing capacity.
- *More frequent categories are more differentiated.* Only in categories used more frequently and in a wider range of contexts we will encounter higher levels of differentiation.

2 SCALES AND DIFFERENTIAL OBJECT MARKING

In a remarkable number of languages with overt morphological case marking not all direct objects are uniformly case marked. Bossong (1983/84: 8) termed this phenomenon “differential object marking (DOM)”. It depends essentially on two kinds of semantic properties of the transitive object (patient) – inherent features which roughly correspond to *animacy* and referential features which are based on the dichotomy of *definite* and *indefinite*. Animacy and definiteness are seen as scalar dimensions:

the animacy scale: human > animate > inanimate

the definiteness scale: pronoun > proper name > definite > indefinite specific > nonspecific

Direct objects that are high on these scales tend to be overtly case marked. In Turkish, for instance, only specific objects are case marked, in Hebrew only definite objects, and in Romanian object case marking is only obligatory for animate pronouns and proper noun objects (cf. Aissen 2003). The respective principle according to Haspelmath (2005, referring to Silverstein 1976): If any patient (P)

“is overtly case marked, then all Ps that are higher on the animacy scale, the definiteness scale, or the person scale are marked at least to the same extent”.

2.1 Counter-examples

Exceptions to the above mentioned rule can be found for instance in Russian: The animate accusative objects in the plural of the feminine and neuter are zero-coded (they got a genitive ending – therefore the name *genitive-accusative*) whereas the inanimate objects are overtly coded. For example:

- | | | |
|-----|--|--|
| (1) | <i>Ja vižu etix ženščin-Ø</i>
'I see these women' | <i>Ja vižu eti dom-a</i>
'I see these houses' |
| | <i>Ja znaju etix lic-Ø</i>
'I know these persons' | <i>Ja znaju eti knig-i</i>
'I know these books' |

The same holds for some masculine animate nouns that are frequently used in the plural. For instance:

- 2) *Ja vižu krestjan-Ø* *Ja vižu kater-a*
 ‘I see farmers’ ‘I see motorboats’
- Ja znaju etix graždjan-Ø* *Ja znaju eti most-y*
 ‘I know these citizens’ ‘I know these bridges’

In all these examples the animate objects are zero-coded and the inanimate objects overtly case marked. This is the direct opposite of the rule stated above. We can explain this phenomenon by the frequency of the genitive plural which is used to code animate objects. In Russian, the genitive plural has a lot of functions and is therefore a very frequent case (cf. Šteinfeld 1963). Zero-coding of the genitive plural is in any case an economical coding (Fenk-Oczlon 1990). The zero-coding of animate objects in Russian is a proper argument for the view of the coding of animacy as a consequence or as an “epiphenomenon” of frequency distributions in the respective language system.

In Russian DOM is also controlled by gender. In the singular, only masculine animate objects are differentially case marked, they get a genitive-accusative ending. Although there might be functional reasons – there was a case syncretism between the accusative and nominative singular – frequency again seems to be the decisive factor: Animate masculines have a higher token frequency than feminine animates, and in more frequently used categories we meet higher levels of differentiation. Taking the frequency count of Zatorina (1966) as database we got the following frequency distributions:

- (3) *otec* ‘father’ 84 *čelovek* ‘man’ 213 *syn* ‘son’ 27
mat’ ‘mother’ 68 *ženščina* ‘woman’ 41 *doč’* ‘daughter’ 17
- brat’* ‘brother’ 39 *mal’čik* ‘boy’ 24 *on* ‘he’ 2340
sestra ‘sister’ 20 *devočka* ‘girl’ 20 *ona* ‘she’ 683

In German differential object marking is exclusively controlled by gender. In the singular pronominal system only the masculine personal pronouns are case marked in the accusative. DOM does not apply to the feminine and neuter singular pronominal declension.

- (4) NOM.SG *er* ‘he’ *sie* ‘she’ *es* ‘it’
 ACC.SG *ihn* ‘him’ *sie* ‘her’ *es* ‘it’

In German case marking of nouns is primarily conveyed by the article. But again: In the singular only masculine direct objects are overtly case marked, irrespective of whether they are animate or inanimate, definite or indefinite.

- (5) NOM.SG. m. *ein Mann* *der Mann* n. *ein Haus* *das Haus*
 ACC.SG. *einen Mann* *den Mann* *ein Haus* *das Haus*
 ‘a man’ ‘the man’ ‘a house’ ‘the house’
- NOM.SG. f. *eine Frau* *die Frau*
 ACC.SG. *eine Frau* *die Frau*
 ‘a woman’ ‘the woman’

As we can see DOM in German is not influenced by animacy or definiteness, but by gender. Again frequency may be used for explaining this asymmetry. In patriarchal societies there is more often talk about men. A look on the frequency distributions of natural gender shows that the token frequency of males is higher than the token frequency of the analogue female items.

The respective frequency data for males vs. females in Ruoff's (1990) spoken corpus of 500 000 words:

(6)	<i>Mann</i> 'man'	442	<i>Vater</i> 'father'	593	<i>Bruder</i> 'brother'	202
	<i>Frau</i> 'woman'	242	<i>Mutter</i> 'mother'	296	<i>Schwester</i> 'sister'	120
	<i>Sohn</i> 'son'	102	<i>Grossvater</i> 'grandfather'	35		
	<i>Tochter</i> 'daughter'	100	<i>Grossmutter</i> 'grandmother'	20		

Regarding the question whether the masculine gender is generally more frequent we could not find reliable data. But we may infer that the masculine gender is the most frequent, at least in basic vocabulary, from two premises: Masculines are to a higher percentage monosyllabic (Köpcke 1982); shortness and particularly monosyllabism (cf. Fenk-Oczlon & Fenk 2008) go hand in hand with high token frequency.

Even more interesting for DOM is the distribution of personal pronouns and definite articles (masculine, feminine, neuter) when assigned to different cases, as in Ruoff (1990: 514f):

(7)	NOM.SG.	<i>er</i> 'he'	2187	ACC.SG.	<i>ihn</i> 'him'	417
		<i>sie</i> 'she'	292		<i>sie</i> 'her'	83
		<i>es</i> 'it'	4375		<i>es</i> 'it'	2292
	NOM.SG. m.	<i>der/dieser</i> 'the/this'	6125	ACC.SG. <i>den/diesen</i>	'the/this'	2229
	f.	<i>die/diese</i> 'the/this'	2208	<i>die/diese</i>	'the/this'	2562
	n.	<i>das/dies(es)</i> 'the/this'	3541	<i>das/dies(es)</i>	'the/this'	3541

The overall frequency of the feminine personal pronoun is much lower than that of the masculine. In the nominative the masculine pronoun is by a factor 7.5 more frequent and in the accusative by a factor 5. This means that feminine as compared with masculine pronouns appear with lower frequency in discourse and with a relatively lower frequency in the nominative than in the accusative. This difference is by far more pronounced in the case of the definite/demonstrative article. Although the proportion of the feminine definite article in the nominative only amounts to 36 per cent of the masculine definite article, it is slightly more frequent in the accusative. Obviously, the feminine tends less often to be agent and subject and more often to be object or patient. Anyway, animacy which is often assumed to be highly correlated with subject status and with definiteness (Comrie 1978: 35) applies in German predominately to masculines.

The high frequency of the German neuter *es* in the nominative singular is probably due to the fact that *es* is used very often as a 'dummy subject': *es regnet* 'it is raining', *es ist kalt* 'it is cold', *es wird getanzt* 'there is a dancing'. And the high number of neuter pronouns in the accusative could be due to the fact that the prototypical object is inanimate (Comrie 1981). An additional explanation might be seen in the frequent occurrence of *es* as an accusative object in fixed phrases such as *ich habe es eilig* 'I am in a hurry'.

3 SCALES AND WORD ORDER

Scales or hierarchies are also assumed to influence constituent order (Allan 1987). In a ranking of the hierarchies from the most powerful determinant to the least he suggests the familiarity hierarchy being strongest, followed by the given > new hierarchy, the universal sequencing conventions, the definitivity and referentiality hierarchies, the personal, social status, and role hierarchies as well as the formal hierarchies.

Familiarity, the number one in Allen's ranking, comes by frequency. High frequency is associated with low informational content and high predictability, and it seems to overrule all other determinants of word order:

In frozen binomials the elements, the concepts and the qualities that describe the prototypical speaker tend to occupy first position. The elements in first position tend to be living, human, male, animate etc. (Cooper & Ross 1975). But those elements occupying the initial position also tend to be the more frequent ones: In a corpus of 400 frozen binomials (Fenk-Oczlon 1989) there showed a large overlap between items ranked high in the animacy, male > female, or human > nonhuman hierarchies on the one hand and high frequency on the other. But in cases of discrepancies frequency turned out to be the winner.

The following examples of English and German freezes may illustrate the power of the regularity "more frequent before less frequent":

- The human > nonhuman sequence is reversed when the frequency relationships are reversed: *horse and rider, Pferd und Reiter* 'horse and rider'
- Likewise the male > female sequence is reversed when the frequency relationships are reversed: *bride and groom, Braut und Bräutigam* 'bride and groom', *goose and gander, duck and drake, Gans und Gänserich* 'goose and gander'
- The exceptions to the animate > inanimate sequence *Stein und Bein* 'stone and bone' as well as *Land und Leute* 'land and people' may also be viewed as frequency-induced reversals: *Stein* and *Land* are much more frequent than *Bein* and *Leute*.

Furthermore, the tendency to place more predictable units before less predictable units might have explanatory power in some of the tendencies usually explained by the definiteness scale: e.g. the tendency to place "definite before indefinite", "old before new", "animate NP sequences before inanimate NP sequences", and to place more frequent and therefore informationally poorer elements in the initial position of frozen conjoined expressions or binomials (Fenk-Oczlon 1989).

DISCUSSION

The regularities illustrated above can be understood as covering laws resulting from economy principles in language processing.

There are two competing tendencies in language: to be clear and to be economic (Fenk-Oczlon 1990). Our analysis of differential object marking suggests strong efforts to be clear only in prominent objects and in the respective, more frequently used categories. In several recent studies (e.g. Haspelmath 2008) the principle of differential object marking is related to economy principles: Animate elements occur less frequently than objects, and for economical reasons only less frequent elements are coded overtly.

The insight that zero-coding of grammatical categories is associated with a high frequency of the respective forms and thus is a very economic coding is not that new. It has been demonstrated on the basis of the Russian case- and aspect-system in Fenk-Oczlon (1990). This study also mentions a point even more relevant for the present topic: the fact that grammatical and lexical meaning interact. In the case of the Russian aspect system it could be shown that the use of a certain aspect depends to a great extent on the meaning of the verb. The more dynamic a verb, the more it tends to be used in the perfect aspect. And the more frequently it is used, the more likely it is to be shorter. In the case of DOM there is an

interaction between lexical and grammatical meaning as well (Bossong 1983/84). Inherent semantic features (i.e. animacy) interact with the use of a certain grammatical category. In German there is, as shown above, an interaction between gender and the frequency of use of a certain case form. Again the split between overtly case marked masculines and zero-coded feminines can be attributed to different frequency distributions. The masculine is less frequent in the accusative and therefore overtly case marked. Since the masculine seems to have a greater overall frequency, the principle of higher degrees of differentiation in more frequent categories (paradigms) might be effective as well.

In the domain of word order, the dominant rule ‘more frequent before less frequent’ seems to serve a rather ‘constant’ and economic flow of linguistic information: High token frequency of a component (in a certain context) means high familiarity, high predictability, and low informational content (in that certain context). On the other hand the initial positions of sentences and freezes are per se characterized by high uncertainty, i.e. high informational content. To place the components with lower informational content in these initial positions avoids both a cognitive overload in the initial positions and too much redundancy in the subsequent positions.

REFERENCES

1. Aissen, J. (2003). Differential object marking: Iconicity vs. economy. *Natural Language and Linguistic Theory* 21(3), 435-483.
2. Allen, K. (1987). Hierarchies and the choice of left conjuncts (with particular attention to English). *Journal of Linguistics* 2, 51-57.
3. Bossong, G. (1983/1984). Animacy and markedness in universal grammar. *Glossologia* 2/3, 7-20.
4. Comrie, B. (1979). Genitive-accusative in Slavic: The rules and their motivation. In B. Comrie (Ed.) *Classification of Grammatical Categories*, 27- 43. Edmonton: Linguistic Research.
5. Comrie, B. (1981). *Language universals and linguistic typology*. Oxford: Blackwell.
6. Cooper, W. E., & Ross, J. R. (1975). Word order. In R. E. Grossmann, L. J. San, T. J. Vance, et al. (Eds.) *Papers from the Parasession on Functionalism.*, 63-111. Chicago: Linguistic Society.
7. Fenk, A., & Fenk, G. (1980). Konstanz im Kurzzeitgedächtnis – Konstanz im sprachlichen Informationsfluss? *Zeitschrift für experimentelle und angewandte Psychologie* 27, 400-14.
8. Fenk-Oczlon, G. (1989). Frequency and word order in freezes. *Linguistics* 27, 517-56.
9. Fenk-Oczlon, G. (1990). Ikonismus versus Ökonomieprinzip. Am Beispiel russischer Aspekt- und Kasusbildungen. *Papiere zur Linguistik* 40, 91-103.
10. Fenk-Oczlon, G. (2001). Familiarity, information flow, and linguistic form. In J. Bybee, & P. Hopper (Eds.) *Frequency and the emergence of linguistic structure*, 431-448. Amsterdam/Philadelphia: Benjamins.
11. Fenk-Oczlon, G., & Fenk, A. (2008) Complexity trade-offs between the subsystems of language. In M. Miestamo, K. Sinnemaeki, & F. Karlsson (Eds.) *Language Complexity: Typology, Contact, Change*, 43-65. Amsterdam/Philadelphia: Benjamins.
12. Haspelmath, M. (2005). Universals of differential case marking. Handout.
13. Haspelmath, M. (2008). Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19 (1). 1-33.
14. Köpcke, K.-M. (1982). *Untersuchungen zum Genussystem der deutschen Gegenwartssprache*. Tübingen: Narr.
15. Ruoff, A. (1990). *Häufigkeitwörterbuch gesprochener Sprache*. Tübingen: Niemeyer.
16. Silverstein, M. (1976). Hierarchy of features and ergativity. In R.M.W. Dixon (Ed.) *Grammatical categories in Australian languages*, 112-171. Canberra: Australian Institute of Aboriginal Studies.
17. Šteinfeld, E. (1963). *Russian Word Count*. Tallin.
18. Workshop on Scales (2008). University Leipzig, call for papers <http://www.uni-leipzig.de/~va/scales.html>
19. Zazorina, L.N. (1966). *Avtomatizacija i statistika v leksikografii*. Leningrad: Izdatel'stvo leningradskogo universiteta.
20. Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology* 40, 1-95.