Travaux du 19ème CIL | 19th ICL papers

Congrès International des Linguistes, Genève 20-27 Juillet 2013
International Congress of Linguists, Geneva 20-27 July 2013

19 ICL

19th International Congress of Linguists
July 21-27 2013
Geneva - Switzerland

Gertraud FENK-OCZLON

Alpen-Adria Universität Klagenfurt, Austria
gertraud.fenk@aau.at

Relationships between Semantic Complexity, Structural Complexity and Markedness: Frequency Matters

oral presentation in workshop: 121 Perspectives on Complexity (Shahin NEMATZADEH)

Published and distributed by: Département de Linguistique de l'Université de Genève, Rue de Candolle 2, CH-1205 Genève, Switzerland
Editor: Département de Linguistique de l'Université de Genève, Switzerland
ISBN:978-2-8399-1580-9

19 ICL
Congrès International des Linguistes
International Congress of Linguists
21-27 July 2013 Geneva

Relationships between semantic complexity, structural complexity and markedness: Frequency matters

Gertraud Fenk-Oczlon
Alpen-Adria Universität Klagenfurt, Austria

ALPEN-ADRIA UNIVERSITÄT
KLAGENFURT | WIEN GRAZ

Outline

- frequency and cognition
- some general definitions of complexity
- definitions of semantic complexity
  - Version I: More complex meanings are expressed by more complex forms
  - Version II: A given lexeme encoding a bigger variety of meanings is semantically more complex
- semantic complexity and structural complexity
- relationships between markedness, semantic complexity, and structural complexity
  - exemplified by Russian aspectual forms
- frequency, structural complexity, and the constant flow of linguistic information

Frequency and Cognition

- Frequency is a central factor in cognitive performance.
- Our cognitive apparatus and its incidental learning shows some special sensitivity to frequency. It constructs automatically, without any specific instruction or demand, a representation of the context-relevant relative frequencies of events or elements.

Frequency and Cognition

The realization of frequency as a determinant of our cognitive processes traces back at least as far as Aristotle. In the course of memorizing, he says

> "custom takes the place of nature. Hence we remember quickly things which are often in our thoughts; for as in nature one thing follows another, so also in the actualization of these stimuli; and the frequency has the effect of nature…" Aristotle, quoted from (Suppes 2009:164)

Frequency and Cognition

- Needless to mention the frequency of the occurrence or co-occurrence of stimuli as a decisive factor in the best studied and most fundamental forms of learning, such as sensitization, habituation, and conditioning (Kandel & Schwartz 1982) .
- No wonder that our "sensitivity to frequency" (Hasher & Chromiak 1975) plays a crucial role in language acquisition (Saffran et al. 1996)

Frequency and Cognition

- In terms of information theory, *learning* means the extraction of a system's redundancy (patterns, invariants, periodicities). In Shannon's (1949) guessing game technique it is mainly the guessing person's (implicit) knowledge about the statistical structure of the respective language – frequency distributions of graphemes and words, transitional probabilities – what allows her to reduce the number of prognostic errors. In other words: She uses the redundancy of that language and she can use it the better the more familiar she is with that language.

Frequency and familarity

Frequency of linguistic segments (e.g. syllables, words, phrases…) does not exert any ***direct*** effect on language structure, but it affects, first of all, cognitive processes:

Higher frequency of use of such a segment results in higher ***familiarity*** of this segment, while the ***cognitive costs*** necessary for producing and/or perceiving these segments decrease.

Familiarity and Cognition

- The concept of familiarity is associated with availability and accessibility.
- High familiarity within a certain context manifests itself in faster and more accurate retrieval processes, in faster and more accurate identification of stimulus pattern, and in higher speed and accuracy of both psychomotor action and anticipation and prediction.

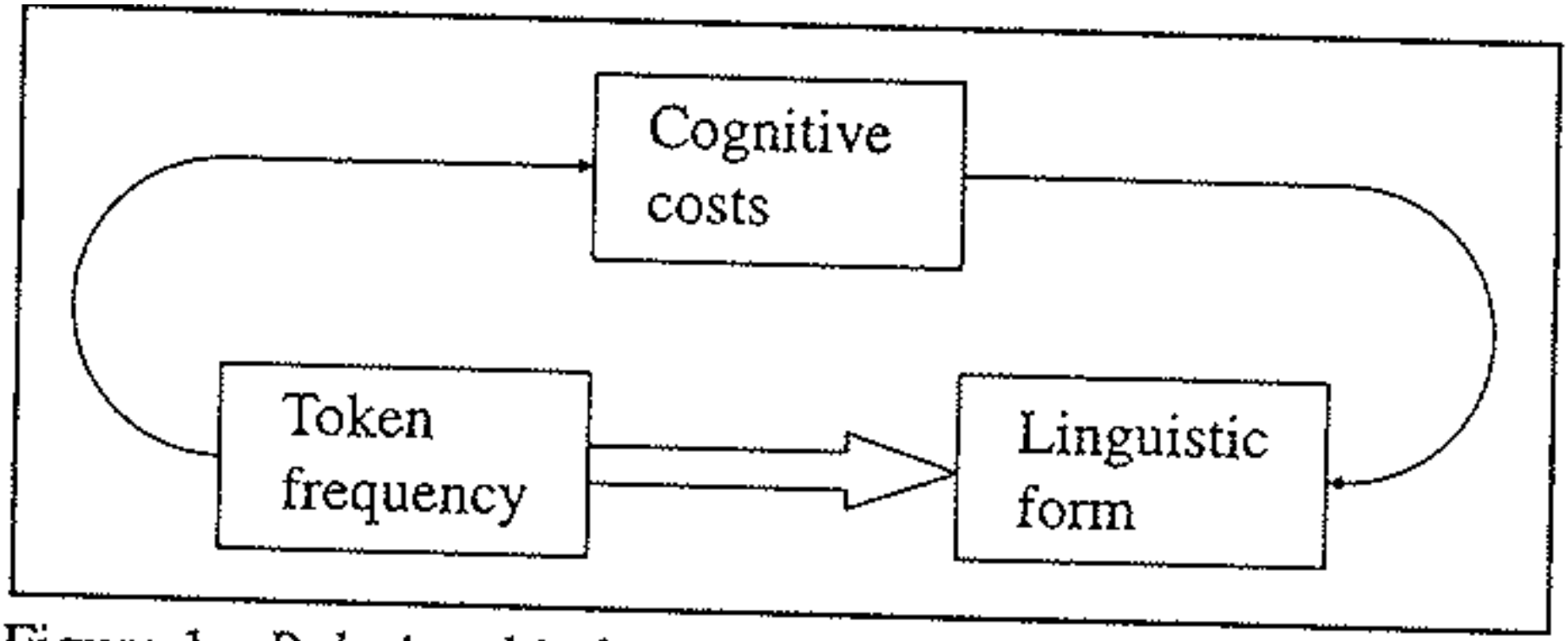


Figure 1. *Relationship between cognitive costs, token frequency, and linguistic form*

Fenk-Oczlon 2001:434

“Complexity“ in general

A rather general and unassuming definition of the complexity of a certain system or unit might refer to the

- **number of components**
- **complexity of components**
- **number of different component types**
- **number of possible interactions between compounds (Simon)**
- **number of different rules determining these interactions (Gell-Mann)**

Complexity and hierarchy

- H.A Simon, in his famous article on “The Architecture of Complexity“ (1962), views ***hierarchy*** *as an universal principle of* ***complex*** *structures,*
- and by a complex system he means “one made up of a large number of parts that have many interactions“(183f)

A hierachy of complexity in language:
number of components, complexity of components

Complexity of

- syllables: number of phonemes (1)
- words: (1) & number of syllables (2)
- clauses: (1) & (2) & number of words (3)
- sentences: (1) & (2) & (3) & number of clauses (4)

continuations: complexity of phonemes, of texts.....

An example: Relations between hierarchical steps

Previous results (Fenk-Oczlon & Fenk 1999):
Crosslinguistic correlations such as:

- The more syllables per clause, the lower the syllable complexity, i.e. the fewer phonemes per syllable. $r = -.75$ ($p < .1\%$)
- The more syllables per word, the fewer phonemes per syllable. $r = -.54$ ($p < .1\%$)

An extended sample of 51 languages

Fenk-Oczlon & Fenk 2010

- The application of our method in a meanwhile extended sample of 51 languages corroborated the previously found negative correlation between number of syllables per clause and number of phonemes per syllable: $r = -.73$ ($p < .01$)

The mean number of syllables per clause in 51 languages

4 – 4.99		5 – 5.99		6 – 6.99		7 – 7.99		8 – 8.99		9 – 9.99		10 – 10.99	
		Dutch	5.05										
		Czech	5.36										
		Mandarin	5.46										
		Sloven.	5.50	Canton.	6.05								
		German	5.50	Cham	6.32								
		Iceland.	5.59	Bambara	6.46								
		French	5.64	Turkish	6.46								
		Estonian	5.68	Albanian	6.55								
		Russian	5.68	Yoruba	6.59								
		Croatian	5.77	Portug.	6.64								
		English	5.77	Persian	6.64								
		Ewondo	5.77	Finnish	6.73	Hopi	7.12	Kadazan	8.00	Kirundi	9.05		
		Hungar.	5.91	Fukien	6.77	Navaho	7.41	Korean	8.18	Chiquit.	9.14	Hawaiian	10.05
Thai	4.64	Hebrew	5.96	Hindi	6.77	Italian	7.50	Annang	8.23	Javanese	9.18	Roviana	10.05
Vietn.	4.91	Arabic	5.96	Panjabi	6.77	Greek	7.55	Basque	8.27	Indones.	9.46	Japanese	10.23
Lamso	4.96	Polish	5.96	Macedon.	6.96	Spanish	7.96	Tagalog	8.95	Marranju	9.86	Telugu	10.96

Figure 1: The frequency distribution of 51 languages over seven classes of the parameter mean n of syllables per clause

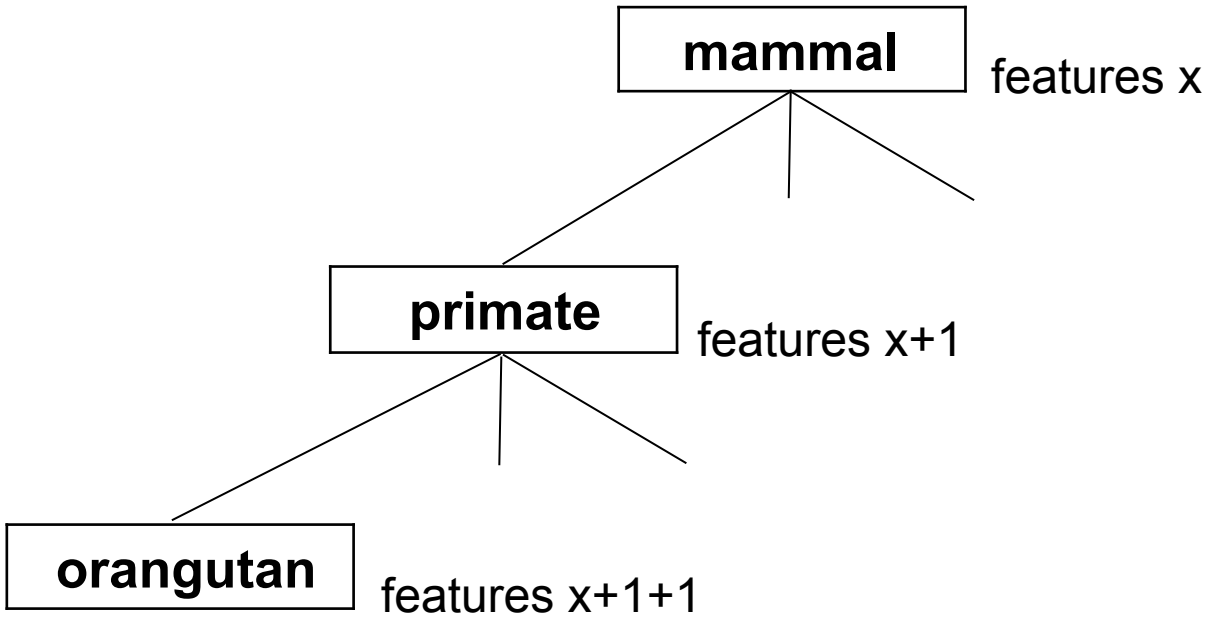
from: Fenk-Oczlon & Fenk 2010:1538

Assigning features of complexity to subsystems of language

- **Phonology:** size of phonemic inventory, articulatory complexity of phonemes, number of phonemes per syllable,...
- **Morphology:** word formation (number of morphemes and number of syllables per word), number of cases, gender distinctions...; transparency vs. opaqueness of morphological forms,..
- **Syntax:** rigid vs. free word order, SVO or others, paratactic versus hypotactic constructions,....
- **Semantics: ????????**

Semantic complexity in the linguistic literature

- Lehmann (1974): **Semantic complexity** is determined by the number of features necessary to describe the meaning of an expression. This corresponds with the "inheritance of features" in knowledge representation.



Semantic complexity in the linguistic literature

- Andersen (2001:47): "…the hyponym (e.g., "rose" (M)) is semantically more complex than its hypernym (e.g., "flower" (U)), the hyponym has more semantic features (i.e. greater intension or semantic depth) than its hypernym….."

Semantic complexity in the linguistic literature

- Mayerthaler (1981): **semantically less marked** means **cognitively less complex:** e.g. the agent is semantically less marked than the patien*t*, the nominative is semantically less marked than the accusative etc…

Semantic complexity in the linguistic literature

- Gennari & Poeppel (2003): **Semantically more complex** means **conceptually more complex**, eventive verbs for instance are semantically more complex than stative verbs

  → Semantic complexity is reflected in processing time

Semantic complexity in the linguistic literature

- Heine & Kuteva (2007:117): “Nouns and verbs are semantically more complex, while adpositions are likely to have some schematic meaning (desemanticization)”

Semantic complexity in the linguistic literature

- According to Evans (2005:34) “semantic structure derives from and mirrors conceptual structure /.../ hence linguistic polysemy reflects complexity at the level of mental representation“.

Semantic complexity in the linguistic literature

- Fenk-Oczlon & Fenk (2008:56): "... a large proportion of expressions encoding a large repertoire of different meanings, i.e., a tendency to homonymy and polysemy, may may be regarded as indicating high 'semantic complexity'."

Two versions *of semantic complexity*

Version I

- Semantic complexity means “more complex meanings” e.g. an increasing “specification” of meanings, or “conceptually more complex”….

Version II

A given lexeme encoding a bigger variety of meanings is semantically more complex.

- Encoding more than one meaning is a case of either polysemy (different related meanings) or homonymy (different non-related meanings). Since in homonymy - as compared with polysemy - more distant values are associated with one and the same word, homonymy should definitely be more complex than polysemy (Raukko 2006).

Version II

- High semantic complexity amounts to a wide range of possible contexts in which certain expressions can occur.
- The (often figurative/ metaphorical) use of an expression in many different contexts means a higher conceptual/ cognitive complexity in the sense of Evans (2005). Particularly if one assumes that polyvalent expressions have to be stored together with possible contexts.

Version II

- But we have to admit that such a concept of semantic complexity goes beyond the scope of a rather technical concept of hierarchical steps of complexity. The different meanings that can be assigned to a certain verbal expression can hardly be viewed as the parts of this expression, nor can the expression strictly speaking, be viewed as a complex of its possible meaning.

Version II

- Version II of *semantic complexity* can be related to Simon's (1962) complexity criterion **"number of interactions between components"**. Taking the word or the lexical morpheme as the component, it may be argued - in the sense of Wittgenstein - that the "different meanings" of this component come about by those linguistic contexts admitting the occurrence of this certain component. These linguistic contexts represent the possible interactions with other components.

Semantic complexity and structural complexity

- In *Version I*: Higher semantic complexity is reflected in higher structural complexity. (semantically more complex meanings are expressed by more complex forms and are less frequent (e.g. Mayerthaler 1981, Andersen 2001).
- In *Version II*: Higher semantic complexity is reflected in lower structural complexity (semantically more complex meanings are structurally less complex and more frequent).

.

Semantic complexity and structural complexity

The relationships between semantic complexity and structural complexity in Version II can be reduced to Zipf's (1935) principles:

1. "The more frequent the shorter"
2. "The number of different meanings of words increases with their frequency" (→ Principle of Economical Versatility of Words)

Correlations between homophony, structural complexity, and frequency

- Ke (2006) compared English, Dutch and German with regard to their number of monosyllables and degree of homophony: In the 5000 most frequent words, English exhibits a much higher proportion of monosyllables (32 %) than Dutch (20 %) and German (14%) and also the highest number of homophones. The correlation between homophony and number of monosyllables across 14 frequency bands was in all three languages very high, i.e. in the region from .96 - .99.

Frequency versus markedness

General assumption:

Frequency is a better predictor of the length of morphological forms than universal markedness assignments

(cf. Fenk-Oczlon 1990, 1991, Haspelmath 2006, 2008)

Frequency versus markedness

Principle of economy:

The more frequent the shorter e.g.:

- Zipf (1935: 25) "...the magnitude of words tends, on the whole, to stand in an inverse (not necessarily proportionate) relationship to the number of occurrences".

Principle of iconicity:

"The less semantically marked the shorter" e.g.:

- Mayerthaler (1981): The decisive factor for the length of morphological forms is "semantic markedness". What is "more" semantically should also be "more" constructionally.'

Frequency vs. semantic markedness

- To shed more light on this issue, an attempt was made to determine which of the two relevant dimensions – ***semantic markedness versus usage frequency*** – is the more capable predictor of the length of aspectual forms in Russian:

Frequency vs. semantic markedness: the length of aspectual forms in Russian

Hypothesis (Fenk-Oczlon 1990):

The length of aspectual forms is motivated by economy and not by (semantic) markedness

Method:

- 67 Russian aspectual were characterized
  - in terms of their frequency (frequency data from Šteinfeld and Zasorina, cited in Breu 1980)
  - in terms of word length

Frequency and length of aspectual forms

(1) Nr.	ipf	pf	St.	Za.	P.	S.
1	мочь	смочь	+1,89	+3,12	+	0
2	говорить	сказать	-1,26	-0,12	+	+
3	становиться	стать	-3,15	-1,58	+	+
4	видеть	увидеть	+0,13	+1,04	+	+
5	думать	подумать	+0,45	+1,72	+	+
6	давать	дать	-1,24	-0,40	+	+
7	делать	сделать	-0,79	-0,02	–	0
8	приходить	прийти	-1,50	-0,72	+	+
9	спрашивать	спросить	-2,01	-0,79	+	+
10	смотреть	посмотреть	0,00	+0,75	+	+
11	брать	взять	-2,58	-0,54	0	0
12	казаться	показаться	+0,85	+1,63	+	+
13	начинать	начать	-2,08	-0,46	+	+
14	выходить	выйти	-1,95	-0,09	+	0
15	проходить	пройти	-1,43	-0,31	+	0

(from Fenk-Oczlon 1990)

16	слушать	послушать	+2,10	+1,43	+	+
17	решать	решить	-3,09	-1,33	0	0
18	отвечать	ответить	-1,82	-0,50	0	0
19	уходить	уйти	-1,48	+0,09	+	0
20	получать	получить	-2,78	-0,87	0	0
21	рассказывать	рассказать	-0,26	-0,22	+	+
22	оставаться	остаться	-1,84	-0,50	+	+
23	слышать	услышать	+0,67	+1,48	+	+
24	приходиться	прийтись	-0,70	+0,41	+	+
25	находить	найти	-1,96	-1,40	+	0
26	ходить	сходить	+3,71	+2,02	+	0
27	входить	войти	-2,74	+0,82	+	—
28	писать	написать	+0,08	+0,65	+	+
29	узнавать	узнать	-2,59	-2,15	+	+
30	помогать	помочь	-0,11	-0,41	+	+
31	подходить	подойти	-2,19	-0,04	+	—
32	показывать	показать	-1,22	-0,30	+	+

33	замечать	заметить	-1,87	-0,84	0	0
34	просить	попросить	-0,05	+1,30	+	+
35	успевать	успеть	-2,34	-1,68	+	+
36	садиться	сесть	-2,39	+0,01	+	+
37	приезжать	приехать	-2,11	-1,63	+	0
38	вставать	встать	-1,82	-0,83	+	+
39	начинаться	начаться	-2,03	+0,92	+	+
40	собираться	собраться	-0,55	+0,39	+	+
41	заниматься	заняться	+1,24	+1,34	—	—
42	играть	сыграть	+1,58	+1,95	+	0
43	приносить	принести	-1,55	-1,11	+	0
44	забывать	забыть	-2,98	-1,18	+	+
45	появляться	появиться	-2,42	-0,17	+	0
46	вспоминать	вспомнить	-1,37	-0,45	+	+
47	поднимать	поднять	-2,40	-0,56	+	+
48	звать	позвать	-0,20	+1,07	+	+
49	встречать	встретить	-1,16	-0,33	0	0

50	подниматься	подняться	-1,35	+0,20	+	+
51	стараться	постараться	+2,30	+1,41	+	+
52	случаться	случиться	-1,62	-1,37	0	0
53	принимать	принять	-1,08	-1,57	+	+
54	нравиться	понравиться	-0,46	+0,93	+	+
55	улыбаться	улыбнуться	-1,43	+0,03	—	+
56	открывать	открыть	-1,89	-0,77	+	+
57	оставлять	оставить	-1,39	-1,15	+	0
58	называть	назвать	-0,04	+0,17	—	—
59	бросать	бросить	-2,25	-0,72	0	0
60	строить	построить	-1,00	-0,64	—	—
61	предлагать	предложить	-2,67	-0,36	0	0
62	класть	положить	-2,04	-1,10	—	—
63	проводить	провести	-1,01	+0,50	+	0
64	доставать	достать	-2,10	-0,84	+	+
65	поступать	поступить	-2,44	+0,07	0	0
66	получаться	получиться	-1,63	+0,24	0	0
67	снимать	снять	-2,00	-0,32	+	+

Frequency and the length of aspectual forms in Russian

Results:

In 50 out of 67 aspectual pairs the more frequent partner, perfective or imperfective, was also the shorter one. In 6 cases the inverse relationship held and in 11 cases no decision could be made.

Markedness and the length of Russian aspectual forms

Markedness theory has many more problems with defining which aspect should be shorter.

- According to e.g. Jacobson (1939) the **perfective is the marked aspect** and the **imperfective the unmarked aspect**. According to the principle of iconicity (markedness) the perfective should therefore be morphologically more complex.

- Comrie (1976) "The Perfective is the marked member of the Perfective/Imperfective opposition"(112) "…unmarked categories tend to have less morphological material than marked categories. The Perfective/Imperfective opposition in Slavonic provides little evidence here…"(114)

Frequency and the length of aspectual forms in Russian

- 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 perfective aspect (Breu 1980). And the more frequently it is used, the more likely it is to be shorter.
- The more a grammatical category is bound to a word in terms of meaning components, the more difficult it is to determine universal markedness weights.

Semantic markedness versus frequency

- It shows that in cases where (semantic) markedness and frequency diverge, as e.g. in the Russian aspect-system, frequency proved to be the better predictor variable.
- Frequency is, moreover, a tangible empirical variable whereas markedness is a theoretical construct.

“Unmarked before marked” or “more frequent before less frequent”?

- The first element of irreversible binominals is semantically unmarked (e.g. Mayerthaler 1981)
- In a former study (Fenk-Oczlon 1989) arguments were presented to support the view that the rule “more frequent before less frequent” represents a principle that is superordinate to rules previously proposed by others (including the rule “unmarked before marked”),
- I tested the new rule “more frequent before less frequent” on 400 freezes (irreversible binominals) from English, Russian and German using statistical data and comparing it with four other rules. **The frequency rule was found to achieve the highest predictive accuracy, with 84% correct predictions.**

Frequency vs. markedness

- So we may say (Fenk-Oczlon 1991) that relatively independent of its degree of markedness, that which is more frequent because of its **natural salience and/or cultural importance**
- is encoded in shorter morphological form
- occupies initial position in frozen binomials
- is more often and to a greater extent polysemous or homophonous
- is more irregular
- survives better in neutralization
- survives better in paradigm regularization
- is earlier acquired by children
- is less affected in aphasia
- is perceived and decoded more easily
- ……

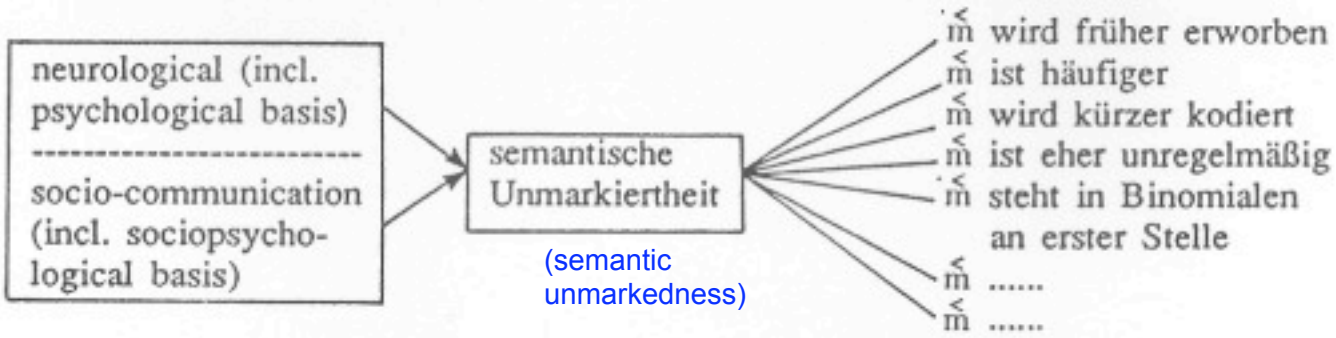
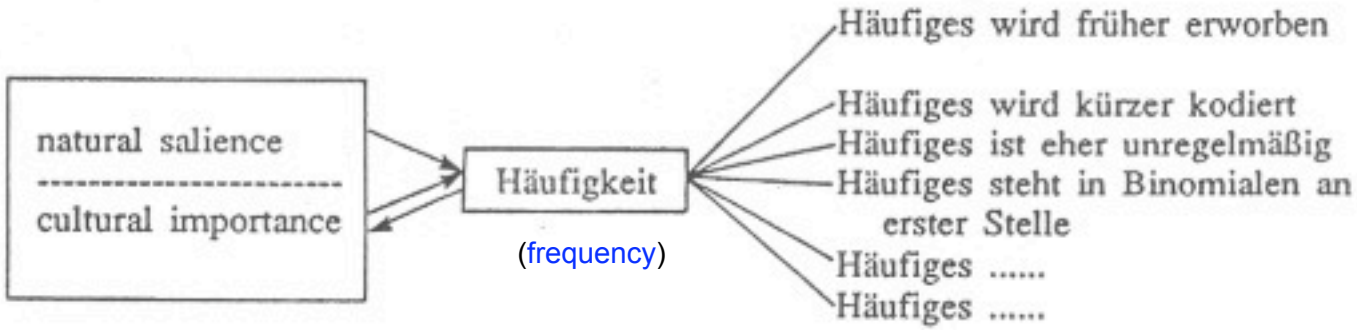


Abb. 1: Die Rolle der Häufigkeit im Modell Mayerthalers



From Fenk-Oczlon 1991: 385

Conclusions

- Relative frequency – overall token frequency as well as relative frequency within specific contexts – has strong effects on cognitive processes which on their part influence linguistic variables such as **structural complexity** or even **word order**. These wide ranging effects have in common that they contribute to a rather even distribution of information over time, i.e. to a relatively constant flow of linguistic information.

Conclusions

- An upper limit on the fluctuations of the information flow seems to result from cognitive capacity limits (e.g. the psychological present) of language users, and a lower limit from economy principles avoiding too much redundancy in communication. Thus, the effects and regularities described underscore the economy and efficiency of linguistic communication.

References

Andersen, H. (2001). Markedness and the theory of linguistic change. In H. Andersen (Ed.) Actualization: Linguistic Change in Progress, 21-58. Amsterdam/Philadelphia: Benjamins

Breu, W. (1980). Semantische Untersuchungen zum Verbalaspekt im Russischen. München: Sagner.

Comrie, B. (1976). *Aspect.* Cambridge: Cambridge University Press.

Evans, V. (2005). The meaning of *time*: Polysemy, the lexicon and conceptual structure. Journal of Linguistics 41, 33-75.

Fenk-Oczlon, G. (1989). Frequency and word order in freezes. *Linguistics 27*, 517-56.

Fenk-Oczlon, G. (1990). Ikonismus versus Ökonomieprinzip. Am Beispiel russischer Aspekt- und Kasusbildungen. Papiere zur Linguistik 40, 91-103.

Fenk-Oczlon, G. (1991). Frequenz und Kognition – Frequenz und Markiertheit. Folia Linguistica 25, 361-94.

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.

Fenk-Oczlon, G. & Fenk, A. (1999). Cognition, quantitative linguistics, and systemic typology. Linguistic Typology 3 (2), 151-177.

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.

Fenk-Oczlon, G. & Fenk, A. (2010). Measuring basic tempo across languages and some implications for speech rhythm. Proceedings of the 11th Annual Conference of the International Speech Communication Association (INTERSPEECH 2010), Makuhari, Japan. 1537-1540.

Gennari, S. & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. Cognition 89, 1, B27-B41.

Hasher, L. & Chromiak, W. (1975). The processing of frequency information: an automatic mechanism. Journal of Verbal Learning and Verbal Behavior 16, 173-184.

Haspelmath, M. (2006). Against markedness (and what to replace it with). Journal of Linguistics 42 (1), 25-70.

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

Heine, B. & Kuteva, T. (2007). The Genesis of Grammar. Oxford: Oxford University Press.

Jakobson, R. (1939/1971). Zur Struktur des russischen Verbums. In Roman Jakobson Selected Writings II. 3-15. The Hague: Mouton.

Kandel, E.R. & Schwartz, J.H. (1982). Molecular biology of learning: Modulation of transmitter release. *Science* (218), 433-443.

Ke, J. (2006). "A cross-linguistic quantitative study of homophony". Journal of Quantitative Linguistics 13, 129-159.

Lehmann, C. (1974). Isomorphismus im sprachlichen Zeichen. In H.J. Seiler (Ed.) Linguistic Workshop II: Arbeiten des Kölner Universalienprojekts 1973/4. München: Fink.

Mayerthaler, W. 1981. Morphologische Natürlichkeit. Wiesbaden: Athenaion.

Raukko, J. (2006). Polysemy as complexity? In A. Arpe et al. (Eds.) A Man of Measure: Festschrift in honour of Fred Karlsson on his 60th Birthday, 357-361. Turku: The Linguitic Association of Finland.

Saffran, J.R., Newport, E.L. & Aslin, R.N. (1996). Word segmentation: The role of distributional cues. Journal of Memory and Language 35, 606-621.

Shannon, C.F. (1949). The mathematical theory of communication. In C.E. Shannon and W. Weaver (Eds.). The Mathematical Theory of Communication. Urbana: University of Illinois Press.

Simon, H.A. (1962). The architecture of complexity. Hierarchic systems. Proceedings of the American Philosophical Societyg 106 (6), 467-482.

Šteinfeld, E. (1963). Russian Word Count. Tallin.

Suppes, S. (2009). Neuropsychological foundations of philosophy. In A. Hieke and H.Leitgeb (Eds.) Reduction. Between the Mind and the Brain. Heusenstamm: ontos verlag.

Zasorina, L.N. (1966). Avtomatizacija i statistika v leksikografii. Leningrad: Izdatel'stvo leningradskogo universiteta.

Zipf, G.K. (1935). The psycho-biology of language: An introduction to dynamic philology. Boston: Houghton Mifflin Company.