

(Note: In *Papiere zur Linguistik* 62/63, 2000, pp 3-14. In case of any discrepancy with the printed version, the printed version will be the ‘authorized’ version.)

The magical number seven in language and cognition: empirical evidence and prospects of future research

Gertraud Fenk-Oczlon and August Fenk, University of Klagenfurt

The first part of this paper is a collection of more or less confirmed occurrences of the magical number 7 in language and of more or less explicit assumptions regarding limits such as the following:

In crosslinguistic comparison the mean number of syllables per clause seems to be restricted to a range of 5 to 10 and the mean length of lexemes to a range of 7 plus minus 2 segments. The upper limit of phonemes per syllable is said to be 9 and languages with more than 9 basic vowels are quite uncommon. 9 or 10 seems to be the upper limit of gongs per phrase in drum languages and of syllables per word in whistled speech. Even the mean number of the languages’ cases, gender and person distinctions is likely to be located in this range.

The second part offers a theoretical framework for a unified theory of such phenomena. It is argued that these limits in the range of about 7 plus minus 2 first of all reflect capacity limits of our working memory (immediate memory span, focus of attention). Further considerations regard some possible ways in which these constraints are also manifested in long-term memory materials, i.e. categories such as case, gender, and person systems.

Introduction

This paper does not represent the state of the art within an elaborated domain of linguistic research, nor is it a typical empirical study starting with hypotheses, operationalizations, and so forth. It is rather a “collection”:

- A collection of occurrences of the magical number 7, plus or minus 2, in linguistic studies.
- A collection of open questions and ideas for future research comprising statistical examinations of hypothesized limits in the region around 7 as well as possible classifications and a more or less unified theory of relevant phenomena.

In the psychology of information processing the number 7 is a somewhat “magical” invariant: According to Miller (1956) it manifests itself as a constraint of the span of absolute judgment, the span of immediate memory, and of the span of attention. This limit (or these limits?) of about seven, plus or minus two, has (or have) since figured prominently in information processing theories. Miller (1956: 91) warned against assuming “that all three spans are different aspects of a single underlying process”. But “generality” is a relevant dimension of empirical progress. Thus it is still tempting to search for **one** “underlying process” or for **one** “covering law” for several regularities corresponding with each other in a certain respect.

Arguments presented in Fenk-Oczlon & Fenk (2001a,b) point to the behavioral relevance of the magical number 7 in general - independently of the sense modality of the input or even independently of whether the respective activity is of rather afferent/perceptual or efferent/motoric nature, and independently of whether one analyzes activities of human beings or non-human animals. As suggested by the experimental findings by Köhler (1952) with birds and by Brannon & Terrace (2000) with non-human primates, this limit of about 7 is no specific characteristic of human information processing. A fascinating finding is reported by Kareev (2000): Small series of about 7 plus minus 2 data pairs produce stronger correlations between the respective variables than the population. This would mean that a span of comprehension comprising about 7 elements or chunks of elements does not reflect a rather arbitrary cognitive limit, but that there must have been a selective advantage and selective pressure for pushing up the limit to this region where minimal indications and minimal contingencies can be detected with a minimum of "computational" work.

All these findings were discussed as indicating some general, extralinguistic or pre-linguistic cognitive preconditions of language, i.e. some sort of "matrix" allowing for as well as constraining the evolution of our complex language system. Is it possible to see some more specific communalities, if one concentrates on those limits that seem to occur in different fields of **linguistic** research?

Linguistic information is a special type of message processed by our cognitive apparatus. If the number seven marks some general limits of this apparatus, it should also show in languages, because language must have developed in adaptation to the general constraints of this apparatus (Fenk-Oczlon & Fenk 2001a).

Relevant observations and assumptions

The following list of manifestations comprises cases where the magical number seven marks the middle of the range of variation (from 5 to 9) as well as cases where it rather marks the upper end of variation. And it includes cases where the distribution within the respective range was already subject to statistical examinations as well as cases where the limits of variation have the character of as yet uninvestigated assumptions.

a) In crosslinguistic comparison ($n = 34$ languages) the mean number of syllables per proposition seems to be restricted to a range of 5 - 10 syllables (Fenk-Oczlon 1983, Fenk-Oczlon & Fenk 1999)

Moreover, we found a significant crosslinguistic correlation (Fenk-Oczlon & Fenk 1985): the higher a language's mean number of phonemes per syllable, the lower the mean number of syllables per simple declarative sentence. This negative correlation between number and complexity (duration) of syllables indicates time related constraints determining this span from 5 - 10 syllables, as does the whole set of significant crosslinguistic correlations (Fenk & Fenk-Oczlon 1993) found between the dimensions number of phonemes per syllable, number of syllables per word, number of syllables per sentence, and number of words per sentence.

Kien et al. (1991) have described comparable action units in humans and non-human primates, and we have suggested (Fenk-Oczlon & Fenk 2001a) that intonation units be viewed as a special case of action units. From this point of view it seems tempting to identify and measure this segmentation in recordings of, on the one hand, special human languages not only in the acoustic mode (like drum language and whistled speech) but also - probably more difficult

- of sign languages. On the other hand it would be interesting to study and analyze the size (in terms of duration as well as in terms of number of elements) of the chimpanzees' tonal patterns or segments of vocalization.

b) 9 syllables per word is likely to be a maximum in German (Menzerath 1954).

It would be interesting to study if this is a maximum value in general.

c) The maximum number of phonemes per syllable seems to be restricted to the region of about 9 phonemes.

According to Menzerath (1954) the maximum number of phonemes per syllables is 7 in Rumanian and 8 in English and German. Gil (1986) studied a list of 170 languages in the Stanford Phonology Archive and found that "the most complex syllable structures observed were assigned value 9 - for example, the (C)(C)(C)V(V)(C)(C)(C)(C) template of English". (Gil 1986:204)

d) More than 80% of the languages investigated by Crothers have from 3 to 7 vowels, and languages with more than 9 basic vowel qualities seem to be quite uncommon (Crothers 1978: 104, 113).

Only 6 languages in his list (Crothers 1978:143) of 209 languages have more than 9 vowels. Four of these 6 languages have 10 vowels, one has 11 (French), and another one 12 (Pacoh).

e) According to Miller (1956), referring to Jakobson et al. (1952), there are about 8 or 10 dimensions (distinctive features) that distinguish one phoneme from another.

f) The average length of a word is approximately 7 plus minus 2 segments (Nettle 1995).

Nettle found a tradeoff between word length and inventory size in a sample of ten languages: The bigger the segmental inventory size in a language, the shorter the average length of a word. Nettle obtained this finding by determining the word length of isolated lexical entries in dictionaries, e.g. the infinite form of verbs. It would, however, be interesting to determine the mean length of words within spoken or written sentences with all their grammatical affixes.

g) Drum languages: about 6 to 10 gong-phrase syllables are produced (sent, needed) for every syllable in a spoken word.

In the Cameroons, for instance, "in order to signify a dog, the gong-beaters sent a six-syllable signal that they rendered phonetically as *kukutotokulo*. Their spoken word for *dog*, however, was *mbo*." (Carrington 2001:1) In spoken Lokale the word for *up above* is *hikolo*, the equivalent gong phrase is *hikolokondause*, which means "up above in the sky". The word for *leopard* is *ngoi* which is beaten out as *alonga losambo*, i.e. "he tears up the roof"; Carrington 2001:2).

h) Whistled speech: The word length seems to be restricted to about nine syllables.

Charalambakis (1994) studied whistled speech in Antia on the Greek island of Euboea. He reports that e.g. "an informant readily whistled Greek words of up to 9 syllables and was

accurately understood by a second informant".

i) The average number of cases in languages seems to be restricted to 5.6 terms with cumulative exponents and to 7.3 terms with separatist exponents (Plank 1986).

According to Plank (1986:32) "*cumulative* exponents simultaneously express at least two co-occurring inflectional categories without being formally segmentable into two or more parts, while *separatist* exponents express only one inflectional category of a word form."

In a later study, Plank (1999:321f) again separates two (dichotomic) parameters of the description of case - separation versus cumulation and invariance versus variance. He assumed that "cumulation and variance are both inherently uneconomical" and could not see any

cogent immediate reason why cumulation should go with variance and separation with invariance in the first place, rather cumulation ought to come with (economical) invariance and (economical) separation with (uneconomical) variance. (Plank 1999:322)

We think that **frequency** (token frequency) is the key concept that offers a rather simple explanation: If a language has predominantly multifunctional cases (one case "accumulates" two or more functions), then this language will manage with a rather low number of case forms but will need and use each one of these case forms very often (high token frequency). And signs with high token frequency tend to both high variance and short coding for obviously economic reasons (e.g. Zipf 1929, Mandelbrot 1954, Fenk-Oczlon 2001). This explanation should also hold for "split morphology" - cumulation/variance and separation/invariance - within languages: The most frequent cases in a given language will tend to cumulation and variance. By the way: We assume that a language's speech rhythm interacts with both parameters, cumulation/variance and separation/invariance, and that there is, on the one hand, a covariation between stress rhythm, cumulation and variance, and, on the other hand, between syllable rhythm, separation and invariance.

The highest number of cases we found so far is 15 (Udmurt). The often reported "giant" number of case categories of the East Caucasian languages is according to Plank "somewhat misleading, and quantitatively fairly exaggerated." (Plank 1986: 44)

j) The number of crosslinguistic aspectual/temporal gram types seems to be restricted to six (Bybee & Dahl 1989).

k) The number of gender distinctions seems to be restricted to about 13.

Corbett (1991:55) reports that Yimas, a Papuan language, "has eleven noun classes or genders", and in Arapesh there seem to be thirteen genders (Corbett & Frazer 2000:316). These are, with the exception of Fula (see below), the highest values we could find so far.

Bantu languages are said to have extensive gender systems comprising between 10 and 20 noun classes. But actually they have only between 5 and 10, because singular and plural are counted (cf. Corbett 1991:47) as separate classes.

An exception to this limit of about 11 might be Fula (a West African language) which has, according to Corbett, about twenty genders in the singular, depending on the dialect, and five in the plural. "There are evident semantic principles involved in gender assignment, but many unclear cases remain" (Corbett 1991:191f). Maybe it is also possible in this case to identify a factor reducing this extraordinary large repertoire.

I) The average number of personal pronouns (excluding gender specifications) of 71 languages is about 8.

When analyzing the person systems of 71 languages from data presented by Forchheimer (1953) and Ingram (1975) we found a mean number of 8.154 and a median of 7.36.

Among these 71 languages the six-person system was the most frequent, followed by eleven-person system, seven-person system and nine-person system. The maximum was a fifteen-person system.

And it is at least remarkable that in some American Indian languages the pronouns of third person have again a sevenfold classification (Forchheimer 1953).

Discussion

The (hypothesized) constraints a - h refer to phonological properties and constraints i - l to grammatical categories. But only few of them can be regarded as a direct manifestation of immediate memory span (see section A below). Nevertheless an attempt is made (in section B) to explain at least some of the remaining universals on the basis of organisational processes of memory. One of the relevant assumptions: "We will assume that information in semantic memory is highly organised, and that categorical clustering reflects this underlying organisation." (Eysenck 2000:327, referring to Mandler 1967).

(A) The magical number seven and the immediate memory span

We will at first focus on considerations regarding the rather direct influence of "focal attention" and "immediate memory span". The planning of a whole clause as well the comprehension of a sentence in its entirety requires a sufficient size of our immediate memory, or, to put it the other way round, presupposes that clauses have to be short enough to fit into this span.

Five very complex or ten very simple syllables forming a clause (constraint a) or nine syllables forming an extraordinary long word (constraint b) seem to fit into such a span. Constraint g (regarding drum languages) and constraint h (regarding whistled speech) can be explained by such a span as well:

If the repertoire of the most elementary signs is very small - e.g. only three elements in the Morsealphabet -, then the equivalence to our graphemes, syllables, words, sentences (supersigns of different levels) will be proportionately longer. The segmentation of an auditorily transmitted message has again to take into account our immediate memory span: About ten gongs in drum language (constraint g) or "syllables" of whistled speech (constraint h) seem to mark the upper limit of such segments or "clauses" in these special languages.

Miller (1956) postulated a "constant" capacity of short term memory when its capacity is measured in terms of number of chunks. Baddeley et al. (1975) argued that there is no valid operationalization of "chunk" and explored in a number of experiments

the hypothesis that immediate memory span is not constant, but varies with the length of the words to be recalled. Results showed: (1) Memory span is inversely related to word length across a wide range of materials; (2) When number of syllables and number of phonemes are held constant, words of short temporal duration are better recalled than

words of long duration; (3) Span could be predicted on the basis of the number of words which the subject can read in approximately 2 sec (Baddeley et al. 1975: 575)

To cut the arguments of Baddeley et al. short: **constant** is nothing but the duration of about 2 sec, not the number of any “chunks” or any better identifiable units such as words or syllables. If measured in terms of the number of items (syllables, words), the span of immediate memory (1986: 44) depends on the articulation rate, i.e. the number of syllables or words that a person can articulate within about 2 sec.

In a later study Baddeley (1994) at first confirms his theory of a time based span: “The fact that span is strongly influenced by the spoken duration of the words suggests a system that is time based rather than chunk based.” (Baddeley 1994:355) But then he admits:

Zhang and Simon concluded that there is a need to assume effects of both the spoken duration of the items, as proposed by the Baddeley and Hitch (1974) working memory model, and also of number of chunks. I accept this and suggest that the chunking effects may be dependent on the operation of the central executive component of working memory. (Baddeley 1994:355)

So far Baddeley’s position regarding the controversy “constancy of time” versus “constancy of number of units”.

But how about the “constancy of information” which is in some respects suggested in Miller 1956 and is explicitly postulated in Miller & Selfridge (1950). Baddeley (1994:354) admits that information theory was rather successful in the area of language, because it emphasizes the redundancy of language. In this context he mentions as an example the study by Miller & Selfridge (1950). Within psychology, however, “the precise measures of information-processing capacity have proved to be much less valuable.” (Baddeley 1994: 354).

From our point of view, this is a rather artificial differentiation, because it was exactly the aim of Miller & Selfridge to demonstrate, by varying the redundancy of the word strings to be recalled, that our memory’s capacity in terms of information remains constant or invariant despite changes of the redundancy of materials.

More generally we would like to argue that the scepticism regarding the potential of information theory of e.g. Baddeley (1994) or Shiffrin & Nosofsky (1994) comes from some misconception of *information*. Our points in this respect (Fenk 1986):

- The sceptics stick to the concept of **objective** information.
- Maybe it is nowhere possible to determine objective information. Assertions regarding “objective” information would afford that objective probability values are available, which is not really possible in empirical science. Anyhow, concerning fixed capacity limits in psychology, only *subjective information* (see next point) can be relevant.
- Whatever might have been the original goal of Shannon’s (1951) measuring instrument, the guessing game technique - what it really and objectively measures is, like “item-difficulty” in psychology, a **relation** between certain problem solvers (guessing persons) and problem (reconstruction of a text): a higher number of errors in the guessing procedure, does not reflect higher information *per se*, but a higher information (higher uncertainty) of the text for the respective guessing subject or collective of subjects.

Some consequences of this approach:

a) If Miller & Selfridge had tested the subjective information of their different word sequences and defined the y-axis of their famous diagram (1950:181) in terms of bits recalled, the empirical data would form a straight horizontal line. More central for the present topic:

b) The controversy between “constancy of chunks” and “constancy of information” disappears:

From this point of view, cognitive mechanisms or strategies like chunking and semantic clustering are methods that enable us - in spite of strict limitations concerning the subjective information - to expand our capacity as defined in terms of “objective” information (Fenk 1985:362; for a broader discussion see Fenk 1986: 212 ff.)

In summary: There is no theoretical or empirical reason for denying the possibility of linguistically relevant constraints in all three dimensions - time (in sec), information (in bits), and number of units (“chunks”, syllables). From this one may expect “constancy” principles in all three dimensions. Crosslinguistic computation of already condensed data (mean values per single language) should show per clause not only a relatively constant number of syllables, but also a relatively constant duration and informational content.

(B) How does the magical number seven get from immediate to long term memory?

If it is the immediate memory span which determines the range of about 7 plus minus 2 (7 plus 8 or minus 4) units - how can it have any effects on the restrictions for the repertoires of gender, of case, and of person?

It is at least remarkable that 15 was not only the maximum in the person, case and probably also gender systems but is also reported (by Miller 1956) to be the highest number of categories a subject can simultaneously handle in absolute judgment experiments: the mean number of categories that a person could simultaneously handle was 6.5, the range was from 3 to 15 categories. “Seven plus or minus two” is perhaps only a catchy and simplifying description of this distribution.

Norman’s considerations about the relations between primary or working memory and secondary memory offer some possible answers to our question why the organisation of secondary memory (or long-term memory) reflects some of the constraints we usually ascribe to working memory:

Die begrenzte Kapazität des Primär-Gedächtnisses mag eine entscheidende Rolle bei der Bestimmung der Organisation des mit großer Kapazität ausgestatteten Sekundär-Gedächtnisses spielen. Wenn das Primär-Gedächtnis die Funktion eines Arbeitsspeichers hat, in dem neu angekommene Information so lange aufzuhalten wird, bis sie erfolgreich in die Struktur des Sekundär-Speichers integriert werden kann, dann gehen die Beschränkungen des ersten auf das zweite über. Dies würde bedeuten, daß Material im Dauer-Gedächtnis in Gruppen von nicht mehr als fünf bis sieben Items kategorisiert oder in anderer Form eingeteilt wird, da mehr durch den Flaschenhals des Primär-Gedächtnisses gleichzeitig nie hindurchgehen.

Wir dürfen auch die Möglichkeit nicht außer acht lassen, daß die zwei Gedächtnissysteme verschiedene Eigenschaften derselben physischen Struktur darstellen können.(Norman 1973:230f.)

In some respects similar is another possible explanation that is “nearer” to linguistic categories in “secondary memory” and has, moreover, the advantage of being sufficient without the metaphor

of the restricted bottle-neck of consciousness:

Most apparently, there are two reasons for rapid language acquisition in children: Some sort of neuronal predisposition or program for the “implementation” of language, and, more generally, the extremely high efficiency of learning because of its active hypothesis-testing nature (Fenk 1986:214). If one is generating and testing rule based hypotheses about how a sentence might continue or what might be a well formed sentence, one always has to reactivate or memorize the “repertoire” of possibilities regarding gender, case, person, etc. Since reactivating and memorizing are characteristics of the activities of our working memory, such repertoires should not exceed this working memory’s capacity. Natural languages could only develop in a way which allows for such cognitive constraints.

References

- Baddeley, Alan, Thomson, Neil & Buchanan, Mark 1975. Word Length and the Structure of Short-Term Memory. *Journal of Verbal Learning and Verbal Behavior* 14: 575-689.
- Baddeley, Alan 1994. The Magical Number Seven: Still Magic after all these Years? *Psychological Review* 101 (2):353-356.
- Brannon, E.M. & Terrace, H.S. 2000. Representation of the numerosities 1-9 by Rhesus Macaques (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes* 26 (1): 31-49.
- Bybee, Joan & Dahl, Östen 1989. The Creation of Tense and Aspect Systems in the Languages of the World. *Studies in Language* 13 (1): 51-103.
- Carrington, John F. 2001. The Talking Drums of Africa. *Scientific American*.
http://www.brainforest.org/the_talking_drums_of_africa.htm, 08.06.01
- Corbett, Greville G. 1991. Gender. Cambridge: Cambridge University Press.
- Corbett, Greville G. & Fraser, Norman M. 2000. Gender Assignment: a Typology and a Model. In G.Senft (ed.) *Systems of Nominal Classification*. 293-325. Cambridge: Cambridge University Press.
- Crothers, John 1978. Typology and Universals of Vowel Systems. In J. H. Greenberg (ed.) *Universals of Human Language*, 93-152. Stanford: Stanford University Press.
- Charalambakis, Christopher 1994. A Case of Whistled Speech from Greece. In I. Philippaki-Warburton, K. Nicolaidis & M. Sifianou (eds.) *Themes in Greek Linguistics: Papers from the First International conference on Greek Linguistics*, Reading.
- Eysenck, Michael W. 2000. Psychology. A Student’s Handbook. Hove, East Sussex: Psychology Press.
- Fenk, August 1985. Is the reduction of subjective information (per unit of time) independent from the starting level? In G.D’Ydewalle (ed.) *Cognition, Information Processing, and Motivation*, 361-373. XXIII International Congress of Psychology (Selected/revised papers). Amsterdam: North-Holland, Elsevier Science Publishers B.V.
- Fenk, August 1986. Informationale Beschränkungen der Wissenserweiterung? *Zeitschrift für experimentelle und angewandte Psychologie* 33 (2): 208-253.
- Fenk, August & Fenk-Oczlon, Gertraud 1993. Menzerath’s Law and the Constant Flow of Linguistic Information. In R. Köhler & B. Rieger (eds.) *Contributions to Quantitative Linguistics*, 11-31. Dordrecht: Kluwer Academic Publishers.
- Fenk-Oczlon, Gertraud 1983. Bedeutungseinheiten und sprachliche Segmentierung. Eine sprachvergleichende Untersuchung über kognitive Determinanten der Kernsatzlänge. Tübingen: Gunter Narr.

- Fenk-Oczlon, Gertraud 2001. Familiarity, information flow, and linguistic form. In J. Bybee & P. Hopper (eds.) Frequency and the Emergence of Linguistic Structure, 231-448. Amsterdam: John Benjamins.
- Fenk-Oczlon, Gertraud & Fenk, August 1985. The Mean Length of Propositions is 7 Plus Minus 2 Syllables - but the Position of Languages within this Range is not Accidental. In G.d'Ydewalle (ed.) Cognition, Information Processing, and Motivation, 355-359. XXIII International Congress of Psychology (Selected/revised papers). Amsterdam: North-Holland, Elsevier Science Publishers B.V.
- Fenk-Oczlon, Gertraud & Fenk, August 1999. Cognition, Quantitative Linguistics, and Systemic Typology. *Linguistic Typology*, 3(2): 151-177.
- Fenk-Oczlon, Gertraud & Fenk, August 2001a. The Clausal Structure of Linguistic and Pre-linguistic Behavior. Paper presented at the Interdisciplinary Symposium "The Rise of Language out of Pre-Language" Eugene (Oregon), May 4-6.
- Fenk-Oczlon, Gertraud & Fenk, August 2001b. What Language Tells us about Immediate Memory Span. In K.W. Kallus, N. Posthumus, P. Jiménez (eds.) Current Psychological Research in Austria, 175-178. Graz: Akademische Druck-und Verlagsanstalt.
- Forchheimer, Paul 1953. The Category of Person in Language. Berlin: Walter de Gruyter.
- Gil, David 1986. A Prosodic Typology of Language. *Folia Linguistica* XX(1-2): 165-231.
- Ingram, David 1978. Typology and Universals of Personal Pronouns. In J.H. Greenberg (ed.) Universals of Human Language, 213-247.
- Jakobson, Roman, Fant C.G.M. & Halle, M. 1952. Preliminaries to speech analysis. Cambridge, Mass. Acoustics Laboratory, Massachusetts Institute of Technology. Tech.Rep. No.13
- Kareev, Yaakov 2000. Seven (Indeed, Plus or Minus Two) and the Detection of Correlations. *Psychological Review* 107 (2): 397-402.
- Kien, Jenny, Schleidt, Margret & Schöttner, Bernd 1991. Temporal Segmentation in Hand Movements of Chimpanzees (*Pan troglodytes*) and Comparison with Humans. *Ethology* 89: 297-304.
- Köhler, Otto 1952. Vom unbenannten Denken. *Verhandlungen d.dt. Zool. Ges.* 16: 202-211.
- Mandelbrot, B. 1954. Structure formelle des textes et communication. Deux etudes. *Word* 10: 1-27.
- Mandler, George 1967. Organisation and Memory. In K.W. Spence & J.T. Spence (eds.) The Psychology of Learning and Motivation: Advances in research and theory. Vol 1. London: Academic Press.
- Menzerath, Paul 1954, Die Architektonik des deutschen Wortschatzes. (*Phonetische Studien*, 3) Bonn: Dümmler.
- Miller, George A. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychological Review* 63: 81-97.
- Miller, George A. & Selfridge, J.A. 1950. Verbal Context and the Recall of Meaningful Material. *American Journal of Psychology* 63:176-185.
- Nettle, Daniel 1995. Segmental inventory size, word length, and communicative efficiency. *Linguistics* 33: 359-367.
- Norman, Donald A. 1973. Aufmerksamkeit und Gedächtnis. Weinheim und Basel: Beltz Verlag. Original Edition 1969. Memory and Attention. New York: John Wiley & Sons.
- Plank, Frans 1986. Paradigm Size, Morphological Typology, and Universal Economy. *Folia Linguistica* 20: 29-48.
- Plank, Frans 1999. Split Morphology: How Agglutination and Flexion Mix. *Linguistic Typology* 3 (3): 279-340.
- Shannon, Claude E. 1951. Prediction and Entropy of Printed English. *The Bell System Technical Journal* 30: 50-54.

- Shiffrin, Richard & Nosofsky, Robert M. 1994. Seven Plus or Minus Two: a commentary on Capacity Limitations. *Psychological Review* 101 (2): 357-361.
- Zipf, George K. 1929. Relative Frequency as a Determinant of Phonetic Change. *Harvard Studies in Classical Philology* 40: 1-95.