Glottometrics 5, 2002,22-28 *To honor G.K. Zipf*

Zipf's Tool Analogy and Word Order

Gertraud Fenk-Oczlon[1]
August Fenk

As we proceed from the artisan down the bench we shall proceed from (a) the ever smaller, lighter, and more frequently used tools to (b) the ever larger, heavier, and less frequently used tools.
G. K. Zipf, 1949: 62

Abstract. This article starts with Zipf's (1949) "Tool Analogy", where the artisan arranges and re-designs his tools in a way minimizing his total work; as a result, more frequently used tools tend to be nearer to him (better accessible), smaller and multifunctional. We then argue that short distance, small size and multifunctionality reflect not only a high overall relative frequency of usage, but in particular a high frequency of usage in the first steps of a variety of complex working procedures. Tool order – word order? This extended Tool Analogy fits to the tendency of more frequent words to obtain initial positions in frozen binomials (Fenk-Oczlon 1989) and the new finding (Fenk & Fenk-Oczlon 2002a,b) that the short, frequent and multifunctional function words tend to concentrate in the first part of sentences.

Keywords: Zipf's Tool Analogy, word frequency, word order, freezes, function words, cognitive economy, information theory

1. Zipf's Tool Analogy

Chapter Three in Zipf (1949) starts with the explication of what he calls "Tool Analogy" – "tools" in analogy e.g. to verbal expressions such as words. It is the aim of the present study to investigate if the arrangement of tools in Zipf's analogy corresponds to the arrangement of words in phrasal conjuncts and in sentences: Tool order – word order?

Before going on to some general remarks on the use of such analogies introduced into scientific communication and before investigating the potential of the Tool Analogy as an intelligent illustration of concrete empirical phenomena within the domain of linguistics, let us give a short characterization of this analogy in the words of Zipf (1949):

An artisan "must survive by performing certain jobs for us with his tools as economically as possible. Beyond that we do not care. Thus we do not care how many tools he uses, nor how he alters their size, shape, weight, and usage, nor how he arranges them on the board, as long as he performs the specific fixed job with a minimum of total work" (p.58). This total work is the product of $f \times m \times d$ (f = frequency of usage of a certain tool, m = the mass or size of the tool, and d = the distance "of a given tool to be its round-trip distance to the artisan's lap and return…" p.59). "However since the artisan is obliged to use his tools with a maximum economy *he must arrange the n tools of his shop in such a way that the sum of all the products of* $f \times m \times d$ *for each of the n tools will be a minimum*" (p.59). It is the question

[1] Address correspondence to: Gertraud Fenk-Oczlon, Department of Linguistics and Computational Linguistics, University of Klagenfurt, Universitaetsstrasse 65-67, A-9020 Klagenfurt

of "Close Packing" which is important, because "'close packing' will at all times decrease the *d* distance of the tools and thereby decrease the work of using them, regardless of the size or mass of the tools in the shop" (p.60). Furthermore, "there is an economy in a small size" of tools (p.60).

"Therefore the magnitude of the Force of Abbreviation will tend to decrease in direct proportion to the distance of the tool from the artisan; the farther that a given tool is from the artisan, the proportionately less the comparative economy will be in reducing its size by a given amount. *Hence in redesigning his tools the artisan will lay a premium upon the reduction of the sizes of all tools in proportion to their nearness to him.*

As a result of the above, we may expect to find in our artisan's shop, as a consequence of years of redesigning, that there will be a tendency for the sizes of tools to stand in an inverse relationship to their nearness to the artisan (i.e., the nearer tools will be the smaller). We shall henceforth call this inverse relationship between size and nearness the Principle of the Abbreviation of Size" (Zipf 1949:61).

"Furthermore, as the frequency of the easiest tool increases (while its mass decreases), the ever nearer to the artisan the tool will be moved because of the exigencies of the 'minimum equation'; and the ever nearer to the artisan that the tool is moved the ever greater will be the Force of Abbreviation in reducing its size" (Zipf 1949:62).

2. On the potential of the Tool Analogy as a cognitive-communicative tool

A general view on language as an "organon", or as a "mental organ" or as a cognitive-communicative "tool" is neither new nor very concrete in detail. Much more convincing is the tool character of the specific analogy introduced by Zipf and the tool character of the spatial metaphors used by Zipf when explicating his analogy. The tools being "nearer" to the artisan are such a metaphor. Powers (1998:152) identifies the term "distance" in Zipf's analogy with "access time".

Seemingly, the potential of Zipf's analogy has not been exhausted so far. Rather recent findings in quantitative linguistics can be illustrated or "explained" by this analogy – at least if we extend it in a certain respect or make it more explicit in this certain respect:

Most jobs to be done by our artisan – a shoemaker, a potter, a coppersmith – require not only a single tool, but a series of tools in a non-arbitrary order. Usually these series will start with rather common and unspecific tools of the handicraft in question before proceeding to more specified tools. Thus, the more common tools obtain not only a high overall frequency of usage, but especially a high frequency of usage in the first and basic steps or operations of a wide range of complex procedures. Special requirements and special tools, e.g. for different decorations of the product, follow – if at all – later in this procedure. Both of the following linguistic findings can be "explained" by the tool analogy, if it is extended or more explicit in this aspect, and both of them are special cases of the rule "the more frequent before the less frequent".

3. Frequency as a determinant of order: more frequent words tend to be placed before less frequent words!

3.1. The tendency of the more frequent word to obtain the initial position in frozen binomials (freezes)

More frequently used words are easier for the speaker to call up and more expectable for the hearer. In order to achieve a constant flow of linguistic information and to avoid peaks of

information, such informationally poorer elements should be placed at those positions which are *per se* associated with higher informational content. This was the main argument for the hypothesis (Fenk-Oczlon 1989) that in freezes, i.e. frozen conjoined expressions or binomials, such as *knife and fork, peak and valley, salt and pepper*, the more frequent word would tend to obtain the initial position. The predictive power of this new rule was tested on the basis of 400 freezes from English, Russian and German and was compared to the predictive power of rules previously proposed by other authors, such as "short before long", "the first word has fewer initial consonants than the second", "front vowel before back vowel", and "semantic principles" (such as the me-first principle). Token frequencies of the single words constituting the frozen binomials were taken from Thorndike & Lorge, Josselson, Meier, Ruoff. The result: With 84% of the predictions being correct (i.e. 337 of 400 freezes) the new rule achieved by far the highest accuracy.

In the context of the present paper, the most interesting rival in this competition was the rule "short word before long word". (As we all know e.g. from Zipf's work, there is a strong inverse relationship between frequency of usage and length of the respective words.) The result of the direct comparison: "High frequency before low frequency" scored with 337 hits, the rule "short before long" with only 152 hits. (To some degree, this difference is a result of a handicap of the latter rule; it is not applicable in cases of equal number of syllables of first and last word). In 145 of these 152 freezes where the word order can be "explained" by "short before long", the order can as well be explained by "more frequent before less frequent". This enormous overlap, together with the higher rate of hits of the frequency rule, is one of several arguments saying that the frequency rule represents a principle that is superordinate to the competing rules.

In his "Tool Analogy" Zipf characterizes the "dynamic" interrelationship between "ease" (small product of $m \times d$) and frequency as follows: "In short, *greater frequency makes for greater ease which makes for greater frequency and so on*" (Zipf 1949:62). This formulation suggests that in this process of a mutual build up between increase of frequency and increase of "ease" the initial impulse usually will come from the variable "frequency".

3.2. The tendency of function words to concentrate in the first part of sentences[2]

In an experimental study by Auer, Bacik & Fenk (2001) on the memory for sentences a text of Glasersfeld (1998) was presented auditorily. A tone at the end of some of the sentences (n = 10) signalled to the subjects that they should try to immediately recall as many words as possible from this sentence.

Reanalyzing the data in order to investigate word-class specific effects on recall (Fenk & Fenk-Oczlon 2002a) we made the following observation: In the sentences presented, function words (such as pronouns, articles, conjunctions,…) dominantly occurred in the first quarter of the sentences, whereas content word (nouns, verbs, adjectives, adverbs) did so rather in the last (see Figure 1, left panel). This was a problem for the statistical evaluation of the recall scores – recall scores in absolute terms were meaningless and had to be related to the proportion of function words and content words occurring in the relevant portion of the sentence – but interesting from the point of view of quantitative linguistics. Were the differences found in the within-sentence distribution of function words and content words a specific characteristic of this one author Glasersfeld or a rather general regularity?

[2] So far, the new empirical findings summarized in this section have only been "published" in conference papers (Fenk & Fenk-Oczlon 2002a, b); a full version will follow (Fenk & Fenk-Oczlon, in preparation).

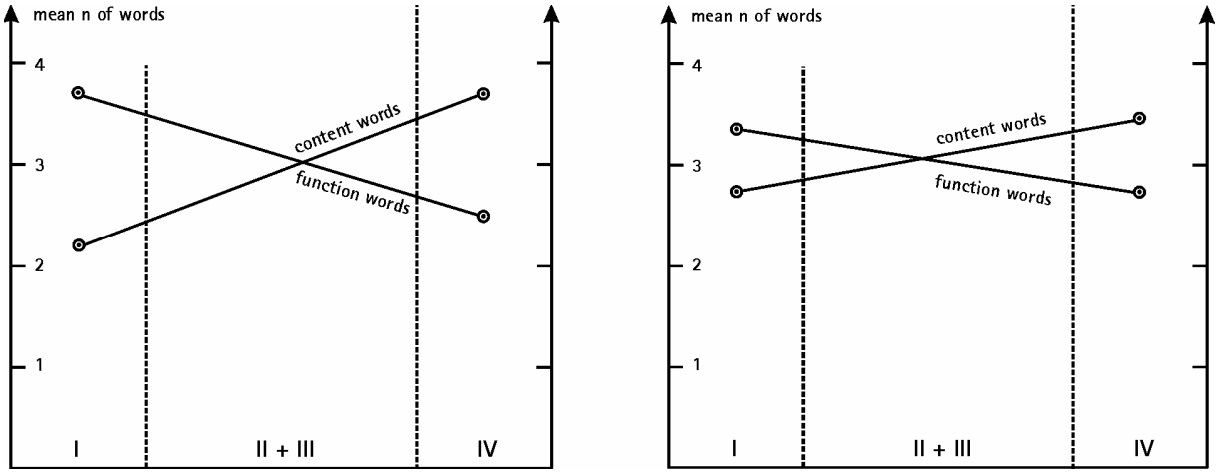

Figure 1: Mean number of function words and content words in the first quarter (I) and in the last quarter (IV) of sentences
Left panel: Mean of 10 sentences from the author Glasersfeld (1998)
Right panel: Mean of 100 sentences (10 sentences from each of 10 authors)

In order to find a first answer to this question German texts of 9 additional authors were analysed: 10 sentences (each third sentence of a text, where possible) from each of 4 scientific texts and 5 literary texts. Results are illustrated in Figure 1, right panel. Table 1 and Table 2 present the numerical values and the significance of results. These results suggest that the word-class specific within-sentence distribution can be generalized for contemporary German texts.

Table 1
Mean frequency of function words and content words in the first quarter (I) versus last quarter (IV) of 100 sentences (10 sentences from each of 10 authors)

	I		IV	differences
function words	3.36	>	2.67	significant, p<1%
content words	2.74	<	3.46	significant, p<1%

Table 2
Mean frequency of function words versus content words within the first quarter (I) and the last quarter (IV) of 100 sentences (10 sentences from each of 10 authors)

	function words		content words	differences
quarter I	3.36	>	2.74	significant, p<5%
quarter IV	2.67	<	3.46	significant, p<1%

In connected discourse many sentences will refer to what was mentioned in the preceding sentences. This reference will most commonly occur in the first part of the sentence ("Thema" before "Rhema", topic before comment, old before new), and function words might play a dominant role in this sort of reference.

If this is an appropriate explanation for the regularity found, this regularity should not be restricted to German texts, but should rather be universal in its essential respect, i.e. the

decrease of function words and increase of content words while the sentence proceeds (Table 1). The starting points of decrease and increase will, however, vary from language to language, dependent, for instance, on the proportion of function words in the specific language. All three patterns shown in Figure 2 seem to be possible. The panel in the middle of Figure 2 represents the proportions found in German, and the analysis of Müller (in progress) indicates that the left panel might represent a pattern characterizing the proportions in Roman languages.

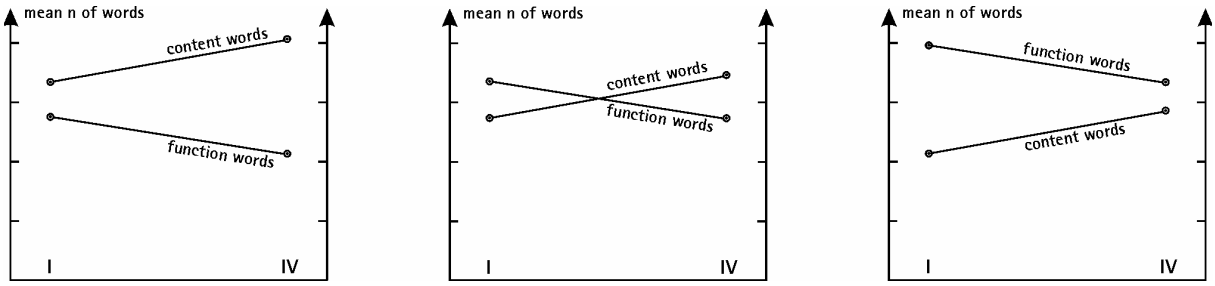

Figure 2. Three different patterns resulting from a within-sentence decrease of function words and increase of content words from changing starting points.

From our regularity – the within-sentence decrease of function words – we may derive a further regularity (Fenk & Fenk-Oczlon 2002b). In those languages, where our first regularity applies, the following regularity regarding the distribution of different word lengths within sentences will apply as well: the mean length of words will increase from the beginning to the end of sentences. The reason for this is that function words are not only extremely frequent, but also – for exactly this reason, as we know e.g. from Zipf (1929, 1949) – relatively short. The prevailing of the (very frequent and therefore) rather short function words in the first part of sentences might contribute to or even account for Behaghel's (1909) "Gesetz der wachsenden Glieder". It would be an interesting attempt to study these regularities in bigger and machine-readable text corpora.

Postponing "heavy" and "new constituents" does not only facilitate comprehension. Arnold et al. (2000: 28) put stress on the fact that it (also) facilitates "processes of planning and production". We might add that this applies not only to the activities of speakers, but the planning and production processes of "other" artisans as well. Anyway: Since we are incessantly endeavoured to anticipate how speech will continue, when we are in the role of the hearer, and since, when in the role of the producer, we are always also the hearer and controller of our own production, producing and listening will follow very similar strategies. "Ease" in active planning and production will correspond to "ease" in anticipation and comprehension.

4. Concluding remarks

It is a fascinating attempt to explore the heuristic potential and the implications of spatial analogies and spatial metaphors. The use as well as the risks of these heuristic tools – be it Zipf's Tool Analogy or one of the more actual neural network analogies – lies in their graphic quality. We have tried to describe an additional parallel between the tool repertoire of the artisan and our lexical repertoire: Tool order – word order!

In Zipf's analogy (Zipf 1949:59) the "relative frequency" of usage is a central term. In cognitive psychology, the learning of relative frequencies, or the "sensitivity" to relative fre-

quencies is also seen as a fundamental mechanism; without this mechanism we would not even be able to identify word units within the speech stream (Zacks & Hasher 2002:28f). Distributions of relative frequencies and probabilities are, moreover, the central topic of the "formal theory of communication", i.e. of the information theory. This theory lacks the concreteness of Zipf's Tool Analogy. To say it positively: it does not need any demons or any "flesh-and-blood artisan" and allows for a more general, integrative and quantitative description of relevant phenomena. In terms of this theory we may say: a higher frequency of the usage of elements goes hand in hand with both, their reduction in "size" (phonological complexity, duration) and informational content ("higher familiarity", "higher accessibility", "higher availability", "lower cognitive costs", "easier to process"). This means: less time for communicating less information! Thus, these reduction processes provide a "constant" and economic flow of linguistic information. And so does the tendency to localize those elements in the initial positions which carry – overall and/or in the specific context – a lower amount of information: the "topic" which comes before "comment", the "old" that comes before the "new", carries low information in this context and - by ways of transitional probabilities (redundancy) – lowers the information of what follows – the "comment", the "new", or the second partner in a binomial.

All the statistical laws discussed above (short before long, more frequent before less frequent) seem to contribute to an efficient communication by contributing to the principle or "covering law" of a constant flow of linguistic information. The central units of this rhythmically organized information flow are clauses with a relatively "constant" duration containing a relatively "constant" number of elements (syllables) and a relatively "constant" amount of information (Fenk-Oczlon & Fenk 2002: 224): Are these units, then, "packages" with an optimal size for cognitive "handling"? This aspect of "Close Packing" (Zipf 1949: 60) would be worthy of a separate study entitled "Zipf's Tool Analogy and the optimal size of clauses".

References

Arnold, J.E., Wasow, Th., Losongco, A. & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language 76, 1, 28–55.*

Auer, L., Bacik, I. & Fenk, A. (2001). Die serielle Positionskurve beim Behalten echter Sätze. *Paper presented at the 29. Österreichische Linguistiktagung, October 26-27 in Klagenfurt.*

Behaghel, O. (1909). Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen 25, 110-142.*

Fenk, A. & Fenk-Oczlon, G. (2002a). The decay of function words in the recall of sentences of different size. Paper presented at "Wortlängen in Texten. Internationales Symposium zur quantitativen Textanalyse". June 21-23 in Graz/Seggau.
Abstracts, http://www-gewi.uni-graz.at/quanta/programm.htm

Fenk, A. & Fenk-Oczlon, G. (2002b). Funktions- und Inhaltswörter in der statistischen Binnenstruktur von Sätzen. Paper presented at the 30. Österreichische Linguistiktagung, December 6-8 in Innsbruck
Abstracts, http://www.uibk.ac.at/c/c6/c604/abstract.html

Fenk, A. & Fenk-Oczlon, G. (in press). Within-sentence distribution and retention of content words and function words. In P. Grzybek (ed.) *Word length studies and related issues.*

Fenk-Oczlon, G. (1989). Word frequency and word order in freezes. *Linguistics 27, 517–556.*

Fenk-Oczlon, G. & Fenk, A. (2002). The clausal structure of linguistic and pre-linguistic behavior. In T. Givón & B. F. Malle (eds.) *The evolution of language out of pre-language: 215-229.* Amsterdam: John Benjamins Publishing Company.

Glasersfeld, E. von (1998). Konstruktivismus statt Erkenntnistheorie. In W. Dörfler & J. Mitterer (eds*.) Ernst von Glasersfeld – Konstruktivismus statt Erkenntnistheorie: 11-39.* Klagenfurt/Celovec: Drava Verlag.

Müller, B. (in preparation). *Die statistische Verteilung von Wortlängen und Wortklassen in lateinischen und italienischen Sätzen*. Phil. Diss., University of Klagenfurt.

Powers, D.M.W. (1998). Applications and explanations of Zipf's law. In: D.M.W. Powers (ed.): *NeMLaP3/CoNLL98: New Methods in Language Processing and Computational Natural Language Learning: 151-60.* ACL.

Zacks, R. T. & Hasher, L. (2002). Frequency processing: a twenty-five year perspective. In: P. Sedlmeier & T. Betsch (eds.) *etc. frequency processing and cognition: 21-36.* Oxford: Oxford University Press.

Zipf, G. K. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology 40*, 1-95.

Zipf, G. K. (1949). *Human behavior and the principle of least effort. An introduction to human ecology.* Cambridge, Mass.: Addison-Wesley [2nd ed. New York: Hafner 1972].

Preprint version