SYSTEMIC TYPOLOGY AND CROSSLINGUISTIC REGULARITIES

Gertraud Fenk-Oczlon & August Fenk, University of Klagenfurt, Austria

ABSTRACT

The "systemic" approach of language typology takes into account that each language goes through selfregulatory processes optimizing the interaction between its (phonological, morphological, syntactical) subsystems. In this paper we refer at first to statistically significant crosslinguistic correlations reported in previous studies between metric properties such as size of clauses in syllables and size of syllables in phonemes. Then we present new results regarding associations between non-metric properties - e.g. a highly significant crosslinguistic correlation between number of cases and dominant adposition order – and their interrelationships with metric properties. We suggest that all these correlations reflect selfregulatory interchanges between different subsystems of language ensuring those characteristics which allow an economic handling by our cognitive and our articulatory system. Information processing limitations are discussed as possible explanations for constraints of language variation: Languages can only develop in adaptation to cognitive capacities or in "co-evolution" with these capacities.

KEYWORDS

Holistic Typology, Cognitive Constraints, Syllable Complexity, Adposition Order, Case

1 INTRODUCTION

After a short explication of the aims of systemic or holistic typology we will (in Section 2) report some crosslinguistic correlations found in a previous study between four different metric properties. These findings are the starting point for the formulation of some new assumptions regarding non-metric properties such as adposition order and number of cases. In Section 3 we will present these assumptions as well as the results of the respective statistical evaluations. Our interpretations (Section 4) refer to information processing limits constraining language variation in the sense of language universals and forcing a fine tuned coordination between subsystems of language.

Systemic Typology suggests systematic interactions between sound structure, morphology and syntax. Several authors (e.g. von der Gabelentz (1901), Skalička (1935), Lehmann (1978), Donegan & Stampe (1983), Gil (1986), Plank (1998)) already have assumed, stated or described co-variations between prosodic, phonological, morphological, and syntactic properties:

"In recent times, typologists have often confined themselves to seeking dependencies among variable language-parts WITHIN syntax, WITHIN morphology, or WITHIN phonology. As to dependencies BETWEEN levels or modules, syntax and morphology were considered essentially the only candidates showing some real typological promise. Dependencies between sound structure on the one hand and word, phrase, clause, sentence, and discourse structure, or also lexical structure, on the other were something respectable main stream typology has steered clear of. /…/ Nonetheless, the temptation to link phonological parameters of crosslinguistic variation on the one hand and morphological and syntactic ones on the other has now and again proved irresistible to the more adventurous, perhaps encouraged by the ever popular all-encompassing master maxim that languages are systèmes où TOUT se tient……"(Plank 1998:195f)

The aim of linking phonological parameters of crosslinguistic variation with morphological and syntactic parameters is the demanding program of systemic or holistic typology, or, according to von der Gabelentz (1901), of typology as such. Von der Gabelentz suggests that some of the components interacting within the system language might be more decisive than others. According to Donegan & Stampe (1983: 350) such a decisive factor might be accent: "What but accent could be behind such holism? Accent is the only factor pervading all the levels of language".

2 A PREVIOUS CROSSLINGUISTC STUDY

Our contributions to this program started with an experimental investigation of the (crosslinguistically limited) variation of the number of syllables per clause (Fenk-Oczlon 1983): A central assumption of this crosslinguistic study was that the number of syllables per "clause" will vary within the range of the magical number seven plus minus two. The clauses used were of a special quality: simple declarative sentences encoding one proposition in one intonation unit, such as *Blood is red*, *The sun is shining*, *A father looks after his family*, etc. 22 German sentences of this sort were presented to native speakers of 27 different languages. (Later on the sample has been extended to 34 languages, 18 Indo-European, 16 non-Indoeuropean. Fenk-Oczlon & Fenk 1999). Native speakers were asked to translate the sentences into their mother tongue and to determine the length of their translations in syllables. The mean number of syllables per clause, computed for each one of these languages, was found to be located almost exactly within Miller's (1956) often quoted range of 7 plus minus 2 elements: The lowest size was 5.05 syllables (Dutch), and only Japanese with 10.2 syllables per clause was located outside the hypothesized range of 5 - 9 syllables. The overall length was 6.48 syllables per simple clause.

Dutch	5.05											
French	5.32											
Czech	5.36											
Chin.	5.41											
Hebr.	5.46											
Slov.	5.50											
Germ.	5.50	Bamb.	6.46									
Icel.	5.50	Turk.	6.46									
Eston.	5.68	Alban.	6.55									
Russ.	5.68	Yoruba	6.59									
Croat.	5.77	Port.	6.64	Hopi	7.12							
Engl.	5.77	Pers.	6.64	Nava.	7.41							
Ewon.	5.77	Hindi.	6.77	Ital.	7.50	Korean	8.18					
Hung.	5.91	Pan.	6.77	Greek	7.55	Anjang	8.23					
Arab.	5.96	Mac.	6.96	Span.	7.96	Basque	8.27	Chiqu.	9.14	Japan.	10.23
5 - 5,99		6 –6,99		7 - 7,99		8 – 8,99		9 – 9,99		10 – 10,99	

Figure 1: The frequency distribution of 34 languages over different classes of the parameter "mean number of syllables per clause" (from Fenk-Oczlon & Fenk 1999:158)

A "footnote" regarding some open questions we have just started to examine. Do these results (the magical number 7 plus or minus 2) also hold true if we expand our sample and especially if we expand the proportion of non-Indoeuropean languages in this sample? Do some languages need far fewer or far more syllables to verbalize propositions? Colarusso (1983) argues that the speed of a language in delivering information depends upon the size of its phonemic inventory and upon its word or syllable canon. He suspects that languages having a very large number of phonemes - like some Caucasian languages, which can have up to 90 consonants and glides - should be "faster" than languages with a low number of phonemes like Hawaiian with only 8 consonants or glides. (Most recently, in the first months of 2004, our sample could be extended by 15 predominantly Austronesian languages. Statistical evaluations are in progress.)

Back to already existing results! The next step in our investigations was the discovery of a statistically significant negative correlation (see correlation A below) between syllable complexity and number of syllables per clause (Fenk-Oczlon & Fenk 1985). This was, as far as we can see, the first "crosslinguistic" correlation in the strict sense of the term, i.e. a computation where each one of the data-pairs (mean n of phonemes/syllable – mean n of syllables/sentence) represents one of the languages of the sample. This correlation - the more complex a language's syllables, the fewer syllables per clause - obviously reflects the general tendency to keep the duration and information of clauses constant. And so does the entire set of four significant and mutually dependent correlations found between the four variables number of phonemes per syllable, number of syllables per word, number of syllables per clause, and number of words per clause (Fenk & Fenk-Oczlon (1993)). In Fenk-Oczlon & Fenk (1999) these correlations were confirmed in a somewhat enlarged sample of 34 languages:

(A) The more syllables per clause, the fewer phonemes per syllable. $r = -0.75$ ($p < 0.1\%$)
(B) The more syllables per word, the fewer phonemes per syllable. $r = -0.54$ ($p < 0.1\%$)
(C) The more syllables per clause, the more syllables per word. $r = +0.47$ ($p < 1\%$)
(D) The more words per clause, the fewer syllables per word. $r = -0.66$ ($p < 0.1\%$)

Furthermore significant interdependencies were revealed between our quantitative parameters and the qualitative variable of word order (Object/Verb vs. Verb/Object): In predominantly agglutinative OV languages the number of syllables per clause and per word is higher and the number of phonemes per syllable lower than in VO languages.

3 A SET OF NEW CORRELATIONS BETWEEN SIZE OF SYLLABLES, NUMBER OF CASES, AND ADPOSITION ORDER

Agglutinative morphology is, moreover, often assumed to be associated with a rather high number of cases and postpositions. And OV order is not only associated with less complex syllables, but also with a tendency to postpositions (e.g. Greenberg 1966 and our sample, where 72 % of the postpositional languages showed OV order and 90 % of the prepositional languages VO order.)

3.1 HYPOTHESES

These results and considerations were the starting point for the following correlational assumptions (Fenk-Oczlon & Fenk (2003)); the correlations A1 and B1 are coupled to their partners A2 and B2 by the above mentioned significant negative correlation between the number of phonemes per syllable and the number of syllables per sentence:

A1   The fewer phonemes per syllable, the higher the number of cases.
A2   The more syllables per sentence, the higher the number of cases.

B1   A low number of phonemes per syllable is associated with a tendency to postpositions.
B2   A high number of syllables per sentence is associated with a tendency to postpositions.

The tendency to suffixing is generally stronger than the tendency to prefixing (e.g. Greenberg 1966). If postpositions get more easily attached to the stem, thus forming a new semantic case (e.g. a local case), then we may assume that

C   a tendency to postpositions is associated with a tendency to a higher number of cases.

3.2 RESULTS

These assumptions were tested on a database of 32 languages. (In 2 of our 34 languages – Annang and Ewondo – no sufficient grammatical information is available so far.) The results:

In all these assumptions the respective crosslinguistic correlations showed the expected tendency, i.e. the expected direction.
Despite the relatively small sample of languages, the correlation C turned out to be highly significant ($r = 0.494$, $p < 1\%$): The number of cases is associated with adposition order; languages with a high number of cases tend to have postpositions and languages with a low number of cases tend to have prepositions.
The coefficients regarding the assumptions B1 ($r = -0.208$) and B2 ($r = +0.314$) were somewhat lower. Only the correlations A1 ($r = -0.145$) and A2 ($r = +0.056$) were far from being significant. But correlation A1, when computed only in those 20 languages having case, was $r = -0.371$. (This is rather promising: Given the same coefficient in a sample with about ten more languages, this coefficient would already be significant.) There is every appearance that languages with relatively complex syllables tend to have a lower number of cases.

We may summarize that our assumptions regarding certain interactions between phonological, morphological and syntactic properties are corroborated by our results either significantly (C), or almost significantly (A1), or at least with respect to the sign (+ or –) of the correlational coefficients (A2, B1, B2).

4   CONCLUSIONS

If we connect, regardless of their statistical significance, present results with our previous results and theoretical inferences, we can identify two basic typological patterns (Table 1).

Table 1: A comparison of two basic typological patterns

low syllable complexity (low number of phonemes per syllable)	**high syllable complexity** (high number of phonemes per syllable)
syllable-timed	stress-timed
high number of syllables per clause	low number of syllables per clause
high number of syllables per word	low number of syllables per word
low number of words per clause	high number of words per clause
OV word order	VO word order
agglutinative morphology	fusional or isolating morphology
high number of cases	low number of cases or no cases at all
separatist case exponents	cumulative case exponents
postpositions	prepositions

Our new statistical results corroborate assumptions already partly stated by other authors such as Lehmann, Donegan & Stampe, etc (see Section 1). By showing systematic co-variations between phonological parameters such as syllable complexity, morphological parameters such as number of morphological cases, and syntactic parameters such as adposition order, they are a further step forward in the realization of the "demanding program" (von der Gabelentz (1901)) of holistic typology. These co-variations seem to reflect the dynamics of a selfregulatory system and systemic exchanges between its subsystems in order to ensure the economy of language perception and production. (According to this view of language as an open and dynamic system we tend to prefer the term "systemic" instead of "holistic" typology.)

Syllable complexity obviously determines the within-sentence segmentation and, to some degree, also the rhythmic pattern within clauses and sentences. But how should we imagine the association of such metric properties with other properties such as number of cases and adposition order?

Stress-timed languages are often (e.g. Dauer 1983, Auer 1993) characterized by their proneness to reduction processes such as the deletion of unstressed vowels, which results in relatively complex syllables. Such reduction processes will, of course, also affect (grammatical) morphemes. And if stress-timed rhythm also favours the fusion, the cumulation and deletion of morphemes, this will result in fusional and/or isolating morphology. This means, moreover, that cumulative exponents will predominantly occur in stress-timed languages. And languages with cumulative case exponents tend, according to Plank (1986), to a lower number of cases (a mean of 5.6 cases) than languages with separatist exponents (a mean of 7.3 cases). If a language has a lower number of (multifunctional) cases, then each one of these cases will be used more often; and as we know at least since Zipf, high frequency units are especially prone to reduction processes. These tendencies taken together might "explain" the associations found between certain phonological traits like syllable complexity and morphological traits such as the number of cases and the prevailing of either cumulative or separatist case exponents.

Our empirical findings suggest that it is first of all the rhythm which discriminates, or makes differences, between languages. The languages' rhythmic organization seems to be rather the determinant than a consequence or a specific aspect of different morphological types (isolating, agglutinative, fusional). One of the variables determining this rhythm is the structure of the basic unit, i.e. the syllable - its "size" or complexity in a certain language and the variability of its "size" and complexity within this language. For instance: A language having almost exclusively V- and CV-syllables represents the absolute minimum of both, the size and the variability of the basic segmentational unit.

The set of correlations presented in Section 2 reflects, first of all, time related constraints. We tend to view clauses or "intonation units" as a special case of action units. Intonation units, and action units in general, seem to have a rather "constant" size and seem to be segmented into a high number of rather small elements or a smaller number of rather complex "elements". And the duration of action units seems to correspond with the duration of succeeding steps in information processing (e.g. Schleidt & Kien (1997)). Such correspondences between general cognitive spans and capacities on the one hand and universal principles of language segmentation on the other can be understood as the result of a "co-evolution of language and cognition" (Fenk-Oczlon & Fenk (2002:215f)), i.e. of a mechanism of mutual stimulation and acceleration: Developmental changes of language stimulate and accelerate the development of cognitive capabilities and vice versa. But however we imagine the motor and the forces changing a particular language or language in general, and irrespective of the subsystem these forces act on in the first place – language will respond as a selfregulatory system maintaining its efficiency as a communicative tool and maintaining a characteristic structure enabling an "easy" or economic handling by our cognitive and our articulatory system.

REFERENCES

Auer, P.(1993). Is a Rhythm-based Typology Possible? A Study of the Role of Prosody in Phonological Typology. *KontRI Working Paper (University Konstanz) 21*

Colarusso, J. (1983). Fast vs. Slow Languages: Comments on the Structure of Discourse and the Evolution of Language. *Papiere zur Linguistik*, 27-51

Dauer, R.(1983). Stress-timing and Syllable-timing Reanalysed. *Journal of Phonetics 11*, 51-62.

Donegan, P. & Stampe, D. (1983). Rhythm and the Holistic Organization of Language Structure". In J..F. Richardson et al. (Eds.) *Papers from the Parasession on the interplay of Phonology, Morphology and Syntax*, 337-353. Chicago: CLS 1983

Fenk, A. & Fenk-Oczlon, G. (1993). Menzerath's Law and the Constant Flow of Linguistic Information. In R.Köhler & B.Rieger (Eds.) *Contributions to Quantitative Linguistics.* Dordrecht: Kluwer

Fenk-Oczlon, G. (1983). *Bedeutungseinheiten und sprachliche Segmentierung. Eine sprach-vergleichende Untersuchung über kognitive Determinanten der Kernsatzlänge.* Tübingen: Narr

Fenk-Oczlon, G. & Fenk, A. (1985). The Mean Length of Propositions is 7 plus minus 2 Syllables – but the Position of Languages within this Range is not Accidental. In Gery d'Ydewalle (Ed.) *Proceedings of the XXIII International Congress of Psychology: Selected/Revised Papers, Vol. 2: Cognition, Information, and Motivation*, 355-359. Amsterdam: North Holland

Fenk-Oczlon, G. & Fenk, A. (1999). Cognition, Quantitative Linguistics, and Systemic Typology. *Linguistic Typology 3*, 151-177

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

Fenk-Oczlon, G. & Fenk, A. (2003). Crosslinguistic Correlations between Size of Syllables, Number of Cases, and Adposition Order. Paper presented at the Fifth International Conference of the Association for Linguistic Typology. Cagliari, September 2003

Gabelentz, G. von der (1901). *Die Sprachwissenschaft: Ihre Aufgaben, Methoden und bisherigen Ergebnisse*. 2nd edition. Leipzig: Tauchnitz

Gil, D. (1986). A Prosodic Typology of Language. *Folia Linguistica 20*, 1986, 165-231

Greenberg, J. H. (1966). Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In J-H. Greenberg (Ed.) *Universals of Language*, 73-113. Cambridge, Mass: MIT Press

Lehmann, W. P. (1978). English: A Characteristic SVO Language. In W. P. Lehmann (Ed.) *Syntactic typology: Studies in the Phenomenology of Language*, 169-222. Austin: University of Texas Press

Miller, G. A. (1956). The Magical Number seven, plus or minus two: some Limits on our Capacity for Processing Information. *Psychological Review 63*, 81-97

Plank, F. (1998). The Co-variation of Phonology with Morphology and Syntax: A Hopeful History. *Linguistic Typology 2*, 195-230

Schleidt, M. & Kien, J. (1997). Segmentation in Behavior and what it can Tell us about Brain Function. *Human Nature 8*, 77-111

Skalička, V. (1935). *Zur ungarischen Grammatik*. Praha: Náklad Filosofické Fakulty University Karlovy