Descriptiveness, Activity and Nominality in Formalized Text Sequences

Peter Zörnig

in cooperation with

Kamil Stachowski
Ioan-Iovitz Popescu
Tayebeh Mosavi Miyangah
Panchanan Mohanty
Emmerich Kelih
Ruina Chen
Gabriel Altmann

2015
RAM-Verlag

Studies in quantitative linguistics

Editors

Fengxiang Fan	(fanfengxiang@yahoo.com)
Emmerich Kelih	(emmerich.kelih@uni-graz.at)
Reinhard Köhler	(koehler@uni-trier.de)
Ján Mačutek	(jmacutek@yahoo.com)
Eric S. Wheeler	(wheeler@ericwheeler.ca)

1. U. Strauss, F. Fan, G. Altmann, *Problems in quantitative linguistics 1*. 2008, VIII + 134 pp.
2. V. Altmann, G. Altmann, *Anleitung zu quantitativen Textanalysen. Methoden und Anwendungen*. 2008, IV+193 pp.
3. I.-I. Popescu, J. Mačutek, G. Altmann, *Aspects of word frequencies*. 2009, IV +198 pp.
4. R. Köhler, G. Altmann, *Problems in quantitative linguistics 2*. 2009, VII + 142 pp.
5. R. Köhler (ed.), *Issues in Quantitative Linguistics*. 2009, VI + 205 pp.
6. A. Tuzzi, I.-I. Popescu, G.Altmann, *Quantitative aspects of Italian texts*. 2010, IV+161 pp.
7. F. Fan, Y. Deng, *Quantitative linguistic computing with Perl*. 2010, VIII + 205 pp.
8. I.-I. Popescu et al., *Vectors and codes of text*. 2010, III + 162 pp.
9. F. Fan, *Data processing and management for quantitative linguistics with Foxpro*. 2010, V + 233 pp.
10. I.-I. Popescu, R. Čech, G. Altmann, *The lambda-structure of texts*. 2011, II + 181 pp
11. E. Kelih et al. (eds.), *Issues in Quantitative Linguistics Vol. 2*. 2011, IV + 188 pp.
12. R. Čech, G. Altmann, *Problems in quantitative linguistics 3*. 2011, VI + 168 pp.
13. R. Köhler, G. Altmann (eds.), *Issues in Quantitative Linguistics Vol 3*. 2013, IV + 403 pp.
14. R. Köhler, G. Altmann, *Problems in Quantitative Linguistics Vol. 4*. 2014, VIII+148 pp.
15. Best, K.-H., Kelih, E. (eds.), *Entlehnungen und Fremdwörter: Quantitative Aspekte*. 2014. VI + 163 pp.
16. I.-I. Popescu, K.-H. Best, G. Altmann, G. *Unified Modeling of Length in Language*. 2014, VIII + 123 pp.

17. G. Altmann, R. Čech, J. Mčutek, L. Uhlířová (eds.), *Empirical Approaches to Text and Language Analysis dedicated to Luděk Hřebíček on the occasion of his 80th birthday*. 2014. VI + 231 pp.
18. M. Kubát, V. Matlach., R. Čech,, *QUITA Quantitative Index Text Analyzer*. 2014, VII + 106
19. K.-H. Best, *Studien zur Geschichte der Quantitativen Linguistik.* 2015. III + 158 pp.

ISBN: 978-3-942303-31-6

RAM-Verlag
Stüttinghauser Ringstr. 44
D-58515 Lüdenscheid
RAM-Verlag@t-online.de
http://ram-verlag.de

Preface

In the present book we study characteristics of language based on formalized text sequences. The study of text as a sequence of various entities is rapidly developing in form of articles, omnibus volumes and monographs. In fact, our linguistic study can be considered as a part of a very fertile interdisciplinary research activity devoted to the analysis of information sequences. Such sequences occur also in computational biology (e.g. in form of DNA strings), in coding theory and data compression. While qualitative linguistic analysis searches for rules which are important for language learning, quantitative analysis tries to capture hidden mechanisms which are not necessary for the understanding of language. Except for certain poetic phenomena, e.g. rhythm which can be produced consciously, these mechanisms cannot be learned and do not represent the core of standard linguistics.

In the present book, a group consisting of mathematicians and linguists – specialists for a certain language – attempts to discover textual phenomena which may seem to be strange for the "normal" linguistics but whose deciphering may help to reveal candidates for laws. Laws are the highest aim of science because without them no theories and no explanations are possible. Unfortunately, in linguistics the testing of a hypothesis is never finished, one can at most validate it to a certain degree. In practice, this validation will never terminate because one would be forced to analyze all languages and, in case of text laws, as many texts as possible. Here no corpuses can help because none of them contains the complete history of language, the evolution of an individual speaker or a complete collection of text sorts.

Hence our attempts merely reveal a few of the infinite number of facets of a text. We try to collect data, find models of their behavior in form of hypotheses, test them, compare the results in texts of eleven languages available to us and try to create a research domain which will never be satisfactorily explored.

We present all observed data in order to enable other researchers to analyze them applying other methods or other characterizations, and to formulate and test other hypotheses. We reduced the whole field to specific phenomena of description, activity and specifying, otherwise the study would be too extensive. Nevertheless, we show at some places the possibility of going into the depth of the hierarchy of phenomena.

Peter Zörnig

Contents

1. Introduction 1

2. Descriptiveness vs. activity 4

2.1. Definitions and tests 4
2.2. Sequential measurement 19
2.3. Runs of two elements 33

3. Nominality 40

3.1. Nominality vs. predicativity 40
3.2. Variability 45
3.3. Triads 46
3.4. Runs of three elements 49

4. Distances 53

5. Some further aspects 72

5.1. Predicativity motifs 72
5.2. Length and frequency 74
5.3. Rank-frequency of predication motifs 83

6. Conclusions 97

References 99
Appendix 102
Author index 118
Subject index 119

1. Introduction

Every linguistic entity has an uncountable number of properties. Their number does not depend on the entity itself – as has been supposed for centuries in the philosophy – but on the *status quo* in linguistics as a science. The researchers define the entities, establish some classifications according to the aim of their research, search for the links between the properties and seek the forces that bring them about. Usually the links between properties are substantiated linguistically – as shown in synergetic linguistics – and are based on the assumption that language is a dynamic system. The text, as the most complex linguistic entity, has the most properties of all, comprising both those of hierarchically lower composing entities and its own ones. While lower entities (except for clause and phrase) are static or local constructions that can be found in dictionaries, the text is in addition a *sequence* of lower units and is able to display a special aspect of the course of any given property.

The fact that texts are written differently because they follow different (conscious or unconscious) aims is well known. There are disciplines like text-type and style classification, language development based on texts of the youth, frequency dictionaries, metrics, speech act, psycholinguistics, sociolinguistics, etc. following quite different aims. Some of these disciplines – or better, some aspects – have already been partially quantified and some mathematical models can be found in this research (cf. e.g. Janda 2013). The history of quantifying linguistic phenomena with mathematical models is more than one hundred years old and the bibliography is very extensive (cf. Köhler 1995). However, mathematical models are no bearers or warrants of truth; they merely reflect our striving for more understandable and more exact capturing of the research object, and yield us the possibility of operating formally with the "facts" discovered. Disciplines using mathematics develop faster than other ones.

Here we shall restrict ourselves to two domains: the expression of text descriptiveness vs. its expression of activity concerning only adjectives and verbs, and the nominality vs. predicativity/specification which is restricted here to the comparison of noun, adjective and verb occurrences. Descriptiveness is expressed by the use of adjectives specifying a noun, and some adverbs specifying both the adjectives and the verbs. Here adverbs and adverbial expressions will be omitted. The adjectives are usually parts of the nominal phrase (*the nice girl*) but they can be added also to the verb (*the girl is nice*; Hungarian: *a szép kislány; a kislány szép*; Russian: *krasivaja devuška; devuška krasivaja*) with or without copula according to the grammar of the given language. Activity is expressed (mostly) by verbs and can even be scaled. We shall not do it here and take into account all forms of the verb "to be" only if it is expressed overtly, e.g. in Indonesian, in stressed forms one uses *ada*, otherwise it does not exist; in other languages it may be quite complex, e.g. the personal forms of *to be* exist but as copula it is omitted. We omit also the modal and the other auxiliary verbs if they

accompany the main verb. A text translated from an Indo-European language into Indonesian would be here automatically less active if we counted also "to be". The cases of Odia and Turkish are described below. Here we are not interested in language typology but in text properties. Verbs consisting of several parts, e.g. in sentence like Slk. *Bol by som býval chcel urobiť*, E. *I would like to work*, Hu. *Szerettem volna megcsinálni* will be considered as 1 verb. Gerunds, gerundives and participles may be interpreted according to the official grammar. In some languages they have different forms, e.g. Slk. *tancoval spievajúc* (he danced singing) but *spievajúci muž tancoval* (the singing man danced). In the first case there are two verbs, in the second one there is an adjective and a verb. In some languages a decision will be necessary in several cases. For a survey of English see Krug (2001), Quirk et al. (1985).

Nominality is both a matter of style and text sort, perhaps also a matter of language. One can express the same subject either by a mere verb, e.g. *I inform you* or one can express this subject using also a noun, e.g. *I convey to you the information*, as is usual in information-theoretical texts. As to nouns, we consider nominal compounds as one noun even if they contain a blank or a conjunction or other joining morphemes and ignore the rest, e.g. *United States; light velocity; bottle filling machine; Natur- und Kulturschutz*, full personal names (*Franz Liszt*), titles (*Der Vorsitzende des internationalen Kommittees*), etc.

In general, one supposes that lyrical poetry is rather descriptive and epical rather active but this need not be the case (cf. Popescu, Čech, Altmann 2013). Further, one supposes that scientific and judicial text-types are rather nominal than active, but this must be tested separately.

It has been shown in the literature that these three word classes may give a text a special character: the adjectives emphasize the descriptiveness, the verbs show the activity, and the nouns may be characteristic of the nominalized expression, e.g. in scientific or judicial texts. The numbers of occurrences of these classes may be combined, their sequences can be scrutinized and help to disclose some aspects of the text dynamics.

The study of predicativity/specification could be continued taking into account logical predicates of second, third, … order, e.g. adverbs are predicates of both adjectives and verbs, but this way of seeing the text has not been studied up to now. In the same way, the trees developed in some grammars (dependency and generative grammar) may be reinterpreted in this sense: for each word the downward number of steps in the hierarchy (tree) will be stated and an indicator can be constructed taking into account the numbers obtained. It would be more appropriate to speak about specification because it is easier to state semantically which word specifies another word than to get problems with the philosophical concept of predicate. The problem may be considered also from the topic-comment or thema-rhema points of view.

A slightly more complex task is the scaling of word classes; at a deeper level even the entities of an individual class may be scaled; for example, the verbs according to the degree of the activity they express, e.g. *to run* expresses

more activity than *to sleep*; or to the history of the rise of an activity in the biological development of Man, e.g. *eat, feel, move, play, think, speak* arose in different periods of our development – but this task needs the cooperation of biologists and anthropologists; the adjectives may be scaled according to the level of the properties (e.g. *nice, pretty, beautiful, magnificent, splendid*, etc.) or by gradation expressed grammatically or lexically. Nouns can be scaled according to the abstract/concrete scale, specific/generic scale, imagery (cf. e.g. Darley, Sherman, Siegel 1959; DeVito 1967; Flesch 1950; Paivio 1979; Pikas 1966; Kisro-Völker 1984; Ballmer, Brennenstuhl 1986), etc. The same holds for all other word classes. Some of the categories have been scrutinized by psycholinguists, child language specialists, grammarians, semanticists, etc. In general linguistics, it is rather a task for the future, even if one finds a great number of trials both in books and on the Internet.

In the present book we shall directly analyze or take into account the results concerning some languages, even if the counting had been performed using different principles. We restrict ourselves to the given aspects and shall not search for their interrelations with other viewpoints. Such an enterprise would be infinite and must be left to future research. It can be performed only stepwise. We consider merely modern journalistic texts; automatically, one could extend the research to the development of journalistic texts historically or scrutinize other text types.

Quite different approaches to sequences in texts can be found in Mikros, Mačutek (2015).

2. Descriptiveness vs. activity

2.1. Definitions and tests

In order to measure the descriptive-active (dis)equilibrium we use the slightly modified Busemann-indicator (1925) defined as

$$(2.1) \quad Q = \frac{V}{A+V},$$

where V is the number of verbs in the text and A the number of adjectives. The indicator in this form represents a simple proportion used several times for this purpose (cf. Altmann 1987, 1988; Popescu, Čech, Altmann 2013; Ziegler, Best, Altmann 2002; Popescu, Lupea, Tatar, Altmann 2015). It has been used in psychology and linguistics both for text, style, as well as characterization of persons, and has a long history beginning with Busemann (1925) and continuing with Antosch (1953, 1959), Goldman-Eisler (1954), Bakker (1965), Fischer (1969), Schlissmann (1948/49), to mention only some of the older works.

If $Q > 0.5$, we consider the text as "active"; if it is smaller than 0.5, we consider it as "descriptive" one. However, a much finer classification is possible. If Q is significantly greater than 0.5, we may consider the text as strongly active; and, on the contrary, if Q is significantly smaller than 0.5, we consider the text as strongly descriptive. Further, texts may exist in which there is no adjective, hence $Q = 1$ can occur. We may consider it extremely active; on the contrary, if the texts contain adjectives but the only verbs are omitted copulas, we obtain $Q = 0$, and consider the text as extremely descriptive.

The adequacy of the simplistic indicator (2.1) can be verified by the following a little more sophisticated approach. Assume that a writer selects n times between a verb and an adjective. Let X be the number of verbs obtained in the n selections. If the verbs and adjectives are chosen with equal probability, then X is binomially distributed with $p = 1/2$, i.e.

$$P(X = x) = \binom{n}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{n-x} = \frac{1}{2^n}\binom{n}{x}.$$

The probability that the number of verbs is smaller or equal to the observed number V is then

$$(2.2) \quad P(X \le V) = \frac{1}{2^n}\sum_{x=0}^{V}\binom{n}{x}.$$

This sum can also be expressed by means of a hypergeometric series.

If the probability (2.2) is smaller than 0.05, we consider the text as strongly/extremely descriptive (SD), if it is greater than 0.05, as descriptive (DE).

If $V > A$, one computes

$$P(X \geq V) = \frac{1}{2^n} \sum_{x=V}^{n} \binom{n}{x}. \tag{2.3}$$

If P is smaller than 0.05, then we consider the text as strongly/extremely active (SA), if it is greater than 0.05, merely active (A).

The test may be performed also asymptotically, without much computation, using the chi-square test shown in Altmann (1988: 26ff) and Altmann, Köhler (2015) and computing

$$X^2 = \frac{(V - A)^2}{V + A} \quad , \tag{2.4}$$

which is distributed as a chi-square with 1 degree of freedom. It can easily be set up if we consider the deviations of A and V from the expectation which is $(A+V)/2$. The conditions are the same as above. The equivalent normal test yields

$$u = (2Q - 1)\sqrt{V + A}. \tag{2.5}$$

It can be shown that $u^2 = X^2$ and the binomial tests yield almost identical probabilities (if n is large). For the sake of illustration, consider a short text in which one finds the sequence: *A, A, V, A, A, V, V*. Here we have $A = 4$, $V = 3$. The descriptiveness ratio yields $Q = 3/7 = 0.43$. Since $V < A$, or $Q < 0.5$, the text is descriptive. In order to test the significance, we compute

$$X^2 = \frac{(3-4)^2}{3+4} = 0.14.$$

Since this is much smaller than 3.84 (= chi-square for $\alpha = 0.05$ with 1 DF) the text is descriptive but not significantly descriptive. Using the normal test (2.5) with the critical value ±1.96 we obtain

$$u = (2*3/7 - 1)\sqrt{7} = -0.38,$$

whose square yields almost exactly the X^2. The respective probability can be found in tables of the chi-square distribution. The asymptotic tests work well if $A+V$ is large but for classification purposes even small values may be used.

Alternatively, by using the binomial distribution we obtain

$$P(X \le 3) = \frac{1}{2^7}\sum_{x=0}^{3}\binom{7}{x} = \frac{1+7+\binom{7}{2}+\binom{7}{3}}{2^7} = \frac{1}{2}$$

Since this probability is very large, there is no reason to reject the hypothesis that the decision between a verb and an adjective is purely accidental.

The texts processed up to now have been analyzed partially under different conditions, but this is the price paid to any analysis in social sciences. Nevertheless, one can perform comparisons or make statements about the indicator. In Table 2.1 we present a survey of some published results and add some new ones. The last column contains abbreviations as follows:

SA = significantly active ($V > A$, $X^2 > 3.84$)
AC = active ($V > A$, $X^2 < 3.84$)
N = neutral ($Q = 0.5$)
DE = descriptive ($V < A$, $X^2 < 3.84$)
SD = significantly descriptive ($V < A$, $X^2 > 3.84$).

Table 2.1
Some adjective-verb indicators for journalistic texts in 11 languages and 86 texts

Text	A	V	Q	X^2	Type
Brazilian-Portuguese					
1	41	168	0.80	77.17	SA
2	32	61	0.66	9.04	SA
3	40	130	0.62	47.65	SA
4	208	174	0.46	3.02	DE
5	115	127	0.52	0.60	AC
6	82	193	0.70	44.80	SA
7	114	154	0.57	5.97	SA
8	147	142	0.49	0.09	DE
9	71	91	0.56	2.47	AC
10	54	36	0.40	3.60	DE
11	132	137	0.51	0.09	AC
12	32	128	0.80	57.60	SA
13	139	168	0.55	2.74	AC
14	132	156	0.54	2.00	AC
15	181	128	0.41	9.09	SD
16	96	97	0.51	0.01	AC
17	83	85	0.51	0.02	AC
18	46	97	0.68	18.19	SA
19	84	83	0.50	20	N

20	59	141	0.71	33.62	SA
21	43	80	0.65	11.13	SA
Portuguese					
1	45	47	0.51	0.04	AC
2	30	28	0.48	0.07	DE
3	30	45	0.60	3.00	AC
4	28	41	0.59	2.45	AC
5	39	54	0.58	2.42	AC
6	47	54	0.53	0.49	AC
7	52	48	0.48	0.16	DE
8	44	56	0.56	1.44	AC
9	45	70	0.61	5.43	SA
10	41	63	0.61	4.65	SA
11	61	45	0.42	2.42	DE
12	68	61	0.47	0.38	DE
13	46	45	0.51	0.01	DE
14	27	38	0.58	1.86	AC
15	39	35	0.47	0.22	DE
16	44	66	0.60	4.40	SA
17	27	39	0.59	2.18	AC
18	29	24	0.45	0.47	DE
19	43	53	0.55	1.04	AC
20	34	37	0.52	0.13	AC
Slovak					
P 1	35	39	0.53	0.22	AC
P 2	44	73	0.62	7.18	SA
P 3	41	66	0.62	5.84	SA
P 4	29	34	0.54	0.40	AC
P 5	47	55	0.54	0.63	AC
Hungarian					
P 1	27	35	0.56	1.03	AC
P 2	59	29	0.33	10.23	SD
P 3	37	48	0.56	1.42	AC
P 4	41	29	0.41	2.06	DE
P 5	63	43	0.41	3.77	DE
Croatian					
P 1	8	41	0.83	22.24	SA
P 2	8	29	0.78	11.92	SA
P 3	32	52	0.62	4.76	SA
P 4	46	66	0.59	3.57	AC
P 5	31	52	0.63	5.31	SA
Chinese					
T 1	52	225	0.81	108.05	SA

T 2	91	470	0.84	256.04	SA
T 3	**33**	436	0.93	346.29	SA
T 4	70	382	0.85	215.36	SA
T 5	48	362	0.88	240.48	SA
Persian					
T 1	150	135	0.47	0.79	DE
T 2	154	110	0.42	7.33	SD
T 3	130	70	0.35	18.00	SD
T 4	147	115	0.44	3.91	SD
T 5	222	145	0.40	16.16	SD
German					
T 1	24	46	0.65	6.91	SA
T 2	36	114	0.76	40.56	SA
T 3	38	66	0.63	7.54	SA
T 4	37	55	0.60	3.52	A
T 5	42	61	0.59	3.50	A
Odia					
T 1	49	55	0.53	0.35	AC
T 2	37	43	0.54	0.45	AC
T 3	46	59	0.56	1.61	AC
T 4	68	56	0.45	1.16	DE
T 5	59	70	0.54	0.94	AC
Russian					
T 1	15	24	0.62	2.08	AC
T 2	18	25	0.58	1.14	AC
T 3	9	19	0.68	3.57	AC
T 4	31	26	0.55	0.53	AC
T 5	37	29	0.44	0.97	DE
Turkish					
T 1	62	58	0.48	0.13	DE
T 2	84	84	0.50	0.00	N
T 3	188	73	0.28	50.67	SD
T 4	125	45	0.26	37.65	SD
T 5	159	52	0.25	54.26	SD

The *Brazilian-Portuguese* and *Portuguese* data were taken from Ziegler (1998, 2001) as ready results but without the presentation of the sequences. Nevertheless, they can be used for some purposes as shown below.

The Chinese, Croatian, German, Hungarian, Odia, Persian, Russian, Slovak and Turkish texts were taken from the current press (cf. Appendix).

Our way of obtaining data was not the same in all cases but a certain degree of uniformity has been attained. The problems connected with data col-

lection will be here touched using the cases of Odia, one of the Indian languages, and of Turkish, as a representative of the Turkic family.

Odia occupies a strategic position among the Indian languages, i.e. in the middle of the Indo-Aryan and Dravidian speaking areas. The state of Odisha is actually the only state in India where there is the largest number of Dravidian and Munda languages speakers. Hence though Odia has been widely accepted as an Indo-Aryan language, it is full of Dravidian and Munda characteristics. In fact, a detailed analysis would persuade us to accept it as a creole. The most important piece of evidence comes from the use of the copular 'be' verb in it. The Dravidian languages' genius is not to use the 'be' verb in the equational sentences. Consider the following Telugu example:

(1) ra:muDu manci pillawa:Du
Ram good child
'Ram is a good boy.'

Notice that this sentence is verbless. But Hindi must have the 'be' verb in a similar sentence. The following example is illustrative:

(2) ra:m accha: laDka: hai
Ram good boy is
'Ram is a good boy.'

Due to its millennia-old contact and convergence with Dravidian, the Odia language does not use the 'be' verb (copula) in similar structures, e.g.:

(3) ra:ma bhala pila:
Ram good child
'Ram is a good boy.'

Though some highly traditional people forcibly use an inflected 'be' verb /aTe/ (from the root /aT-/) in written Odia, it is used neither in the standard spoken variety nor in the texts written by conscientious writers.

Another significant point regarding the use of the copula in Odia is that it makes a clear distinction between nouns and adjectives. Consider the following examples:

(4) ra:ma pila:Ta: bhala
Ram child good
'As a boy Ram is a good.'

Notice that (4) has the adjective /bhala/ 'good' sentence-finally. Therefore, it is also quite acceptable to use:

(5) ra:ma pila:Ta: bhala achi
Ram child good is
'As a boy Ram is a good.'

In other words, if there is an adjective at the end of a sentence like (4), the 'be' verb can be used optionally as in (5) whereas its use is not allowed if the sentence ends in a noun like (3).

There is another intriguing characteristic of Odia that has never been discussed by any Odia grammarian or linguist. Though Odia, like most other

Indian languages, is predominantly a verb-final language, when it comes to the use of the 'be' verb it is either verbless or obligatorily verb-medial. For example:

(6a) se mo ba:pa:
he my father
'He is my father.'

(6b) se hele mo ba:pa:
he is my father
'He is my father.'

If the verb in (6b) is moved to the sentence-final position it will be ungrammatical in the intended sense, e.g.:

(7) se mo ba:pa: hele
he my father became
'He became my father.'

It means 'he' is not 'my' real father, but 'he' became 'my' father due to some reason. It should be mentioned here that the Proto-Munda language was most probably verb-medial and Khasi, an Austroasiatic language and a sister of the Munda languages, verb-medial even today. So it can be argued that the verb-medial Odia example in (6b) is, in fact, an instance of retention of the Munda characteristic. Thus, the use of the verb 'be' in Odia follows either the Dravidian or the Munda pattern.

The structure of Turkish is somewhat different from that of many of the Indo-European languages, and the traditional Graeco-Roman distinctions do not always apply to it as readily as they do to Greek or Latin. This poses a problem to the present work, in which it is possible to remove that multiple ways. The one chosen here can be described very briefly as practical and functional.

What we mean by this is that in order to achieve the practical goal of meaningfully comparing Turkish with Indo-European and other languages, we are more concerned with the function the specific words have in the specific context, than with their morphological structure or any other property. This is not to say that we look at Turkish as if it were an Indo-European structure dressed in Turkic words. We extract from it a set of features that are already there, only it is a different set than the one that comes to the fore in the most natural of ways.

Perhaps the most contentious is the treatment of the so-called *izafet* constructions, and of participles, especially those in *-dık* and *-acak*. For a detailed explanation, the reader must be referred to one of the grammars of Turkish (cf. Banguoğlu 1986; Ersen-Rasch 2001; Swift 1963; Stachowski 2009). Here, we will only adduce, as an illustration, two and a half of the sentences upon which our results are based.

(8)

Kur'an'ın	ilk	ayetlerinin	vahyedildiği
Koran-Gen.	first	verse-Pl.-Px3Sg.-Gen.	to reveal-Passive-*dık*-Px3Sg.
N	A	N	A

Kadir	Gecesi	bu	ayın
Qadr	night-Px3Sg.	this	month-Gen.
A	N	A	N

içindedir.
inside-Px3Sg.-Locative-predicative suffix
V

The Night of Qadr, revealed in the first verses of the Koran, is in this month.

(9)

Festivalin	ülkenin	ve	dünyanın	edebiyat
festival-Gen.	country-Gen.	and	world-Gen.	literature
N	N		N	A

gündeminin	nabzını	tuttuğunu
agenda-Px3Sg.-Gen.	pulse-Px3Sg.-Acc.	to hold-*dık*-Px3Sg.-Acc.
N	N	N

söyleyebiliriz.
to say-Potential-Aorist-1Pl.
V

We can say that the festival keeps a finger on the pulse of the country's and world's literary agenda.

As a general, though not exceptionless rule, the attributive element of the first and second izafet was counted as an adjective; of the third izafet as a noun. Hence *dünyanın edebiyat gündemi* in (9) is reduced to the sequence N A N. Similarly, *Kadir Gecesi* in (8) is represented as A N, regardless of the clearly nominal translation into English.

For the participles in *-dık* and *-acak*, there was no general rule. In (7), *vahyedildiği* was considered an adjective because that is its function in this context, while *tuttuğunu* in (9) is clearly employed as a noun. The majority of participles were most often considered adjectives; cases such as (9) were less frequent, and rarer still were situations such as the one in (10). Note also that the nominal elements in compound verbs are not counted separately.

(10)

… satın	aldığınız	kitabın	parasını
purchase	take-*dık*-Px2Pl.	book-Gen.	money-Px3Sg.-Acc.
A		N	N

dürüstlük	kutularına	bırakıp	kitap
honesty	box-Pl.-Px3Sg.-Dat.	to leave-*ıp*	book
A	N	V	A

keşfine	devam	ediyorsunuz.
discovery-Px3Sg.-Dat.	continuation	to do-Present-2Pl.
N	V	

… leaving the money for the books you bought in the honesty box, you [pl.] continue the discovery of books.

The same reasoning, and the ideas of practicality and functionality, apply at the level of sentences. We follow the native speakers' intuition of the authors of the texts, and do not break sentences in what they saw as the middle of an utterance, merely because we happen to run into a conjugated verb. The notion that the end of the sentence is the only place where a Turkish verb may and must be, is already rather rebutted by the use of participles such as in (10), and the very examples one encounters in spoken and written texts.

In every language there are some problems rendering our results relative – the fate of all scientific enterprises. Nevertheless, we are sure that an appropriate analysis can unveil the laws concealed somewhere in the background.

The style, the text type and the language can be characterized by means of the activity-descriptiveness vector defined as (QV = vector of qualified *Q*s)

(2.6) QV = [SA, AC, N, DE, SD],

where the elements represent the number of texts having the above properties. For the analyzed data we obtain the results presented in Table 2.2.

Table 2.2
QV-vectors of activity-descriptiveness

	Language and texts	**SA**	**AC**	**N**	**DE**	**SD**
1	Brazilian-Portuguese	9	7	1	3	1
2	Portuguese	3	10	0	7	0
3	Slovak	2	3	0	0	0
4	Hungarian	0	2	0	2	1
5	Croatian	4	1	0	0	0
6	Chinese	5	0	0	0	0
7	Persian	0	0	0	1	4
8	German	3	2	0	0	0
9	Odia	0	4	0	1	0
10	Russian	0	4	0	1	0
11	Turkish	0	0	1	1	3

One could, of course, present the vectors in Table 2.2 in relative values yielding a better optical survey especially if the numbers of texts are quite different, but here we shall consider only those languages for which we analyzed 5 texts.

As can be seen in Table 2.1 and 2.2, languages may have an expressed trend. Only a comparison with other text types in the given language could show whether the trends are properties of texts types or of language. The identity of Odia and Russian in Table 2.2 does not mean a final result: increasing the number of texts or taking other text types would surely change the result.

Let's consider the well known formula of the cosine of the angle between two vectors

$$(2.7)\quad \cos\alpha_{ij} = \frac{\vec{V}_i \cdot \vec{V}_j}{\left|\vec{V}_i\right| \cdot \left|\vec{V}_j\right|},$$

where α_{ij} (expressed in radians) is a measure of dissimilarity between the considered vectors. Consider, for example, the Brazilian-Portuguese (9,7,1,3,1) and Portuguese (3,10,0,7,0) texts. Computing (2.7) we obtain

$$\cos\alpha_{ij} = \frac{9(3) + 7(10) + 1(0) + 3(7) + 1(0)}{\sqrt{9^2+7^2+1^2+3^2+1^2}\sqrt{3^2+10^2+0+7^2+0^2}} = \frac{118}{11.8743(12.5698)} =$$

$$= 0.7906$$

and the radian is *arccos*(0.7906) = 0.6590. Numerical values for the comparison of vectors considered above are given in Table 2.3 below.

Table 2.3
Dissimilarity angles α_{ij} in radians

	Language	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	**Slovak**	x	0.9828	0.7378	0.5547	1.5708	0.4253	0.6529	0.3948
2	**Hungarian**	0.9828	x	0.1617	1.5708	1.0643	0.9404	0.4473	1.1920
3	**Croatian**	0.7378	0.1617	x	0.2452	1.5708	0.4985	1.1324	0.3430
4	**Chinese**	0.5547	1.5708	0.2452	x	1.5708	0.7107	1.3298	0.5889
5	**Persian**	1.5708	1.0643	1.5708	1.5708	x	1.4273	1.4353	1.5708
6	**Braz.Port.**	0.4253	0.9404	0.4985	0.7107	1.4273	x	0.6590	0.2921
7	**Portuguese**	0.6529	0.4473	1.1324	1.3298	1.4353	0.6590	x	0.8765
8	**German**	0.3948	1.1920	0.3430	0.5889	1.5708	0.2921	0.8765	x

	Odia	**Russian**	**Turkish**
Slovak	0.6314	0.6314	1.5708
Hungarian	0.6293	0.6293	1.0443
Croatian	1.3333	1.3333	1.5708
Chinese	1.5708	1.5708	1.5708
Persian	1.5119	1.5119	0.3155
Braz.-Port.	0.8851	0.8851	1.3921
Portuguese	0,4350	0,4350	1.4201
German	1.0026	1.0026	1.5708
Odia	x	0.0000	1.4976
Russian		x	1.4976

Note that 1.5708 = π/2. The numbers in Table 2.3 can be used for classification or at least for ordering the texts/text types. If we compute the mean radians for each class expressing the difference *to all other text sets*, we obtain

Turkish:	1.3450
Persian:	1.2949
Chinese:	1.2732
German:	1.1920
Croatian:	1.0034
Odia:	0.9999
Russian:	0.8888
Hungarian:	0.8533
Brazilian:	0.8227
Slovak:	0.8153
Portuguese:	0.7917

Needless to say, many more texts would be necessary in order to perform a first typological description based on Busemann's indicator. By adding a new language or new texts, the similarities change but we conjecture that the more data are collected the more stable will be the ordering/classification. On the other hand, journalistic texts may change their image in the course of time not only on political grounds. Thus the study of the history of journalistic texts shown from this point of view could tell us something about the language itself. The above ordering does not show any typological or genetic connections.

The unity of the style of journalistic texts in a language can be stated in a different way. One compares all *Q*-values of a language with another using the asymptotic normal test

$$(2.8) \quad u = \frac{|Q_1 - Q_2|}{\sqrt{Var(Q_1) + Var(Q_2)}}.$$

Since *Q* is a proportion, one can use the above simplified asymptotic test (2.5). Again, if $|u| < 1.96$, the texts can be considered similar. Further, if there are *n* texts and more than $n(n\text{-}1)/4$ of them are similar, the given text sort can be considered uniform; or the style of the writer in the given texts is uniform – of course, merely concerning his activity/descriptiveness. In our terms, in the above criterion (2.7), $Q = V/(A+V)$ and $Var(Q) = Q(1\text{-}Q)/(A+V)$ because *Q* is a proportion.

Performing the above test we obtain the results as follows: For Slovak, the results are displayed in Table 2.4, for Hungarian in Table 2.5, for Croatian in Table 2.6, for Chinese in Table 2.7, for Persian in Table 2.8, for German in Table 2.9, for Odia in Table 2.10, for Russian in Table 2.11, for Turkish in Table 2.12.

Table 2.4
u-tests for differences of *Q* in Slovak texts

Text	1	2	3	4	5
1	x	1.23	1.21	0.12	0.13
2	1.23	x	0.0	1.04	1.20
3	1.21	0.0	x	1.02	1.17
4	0.12	1.04	1.02	x	0.0
5	0.13	1.20	1.17	0.0	x

Table 2.5
u-tests for differences of *Q* in Hungarian texts

Text	1	2	3	4	5
1	x	2.85	0.0	1.74	1.90
2	2.85	x	3.13	1.04	1.16
3	0.0	3.13	x	1.88	2.08
4	1.74	1.04	1.88	x	0.0
5	1.90	1.16	2.08	0.0	x

Table 2.6
u-tests for differences of *Q* in Croatian texts

Text	1	2	3	4	5
1	x	0.58	2.79	3.38	2.65
2	0.58	x	1.85	2.30	1.74
3	2.79	1.85	x	0.43	0.13
4	3.38	2.30	0.43	x	0.57
5	2.65	1.74	0.13	0.57	xx

Table 2.7
u-tests for differences of *Q* in Chinese texts

Text	1	2	3	4	5
1	x	1.06	4.55	1.38	2.45
2	1.06	x	4.62	0.44	1.79
3	4.55	4.62	x	3.90	2.51
4	1.38	0.44	3.90	x	1.29
5	2.45	1.79	2.51	1.29	x

Table 2.8
u-tests for differences of *Q* in Persian texts

Text	1	2	3	4	5
1	x	1.18	2.68	0.70	1.79
2	1.18	x	1.54	0.46	0.50
3	2.68	1.54	x	0.02	0.01
4	0.70	0.46	0.02	x	1.00
5	1.79	0.50	0.01	1.00	x

Table 2.9
u-tests for differences of *Q* in German texts

Text	1	2	3	4	5
1	x	1.54	0.31	0.78	0.87
2	1.54	x	2.13	2.62	2.81
3	0.31	2.13	x	0.53	0.63
4	0.78	2.62	0.53	x	0.51
5	0.87	2.81	0.63	0.51	x

Table 2.10
u-tests for differences of *Q* in Odia texts

Text	1	2	3	4	5
1	x	0.12	0.48	1.17	0.21
2	0.12	x	0.33	1.20	0.07
3	0.48	0.33	x	1.67	0.29
4	1.17	1.20	1.67	x	1.45
5	0.21	0.07	0.29	1.45	x

Table 2.11
u-tests for differences of *Q* in Russian texts

Text	1	2	3	4	5
1	x	0.37	0.51	0.68	1.82
2	0.37	x	0.85	0.30	1.44
3	0.51	0.86	x	1.18	2.24
4	0.68	0.30	1.18	x	1.22
5	1.82	1.44	2.24	1.22	x

Table 2.12
u-tests for differences of *Q* in Turkish texts

Text	1	2	3	4	5
1	x	0.28	3.81	3.85	4.35
2	0.28	x	0.46	4.59	5.21
3	3.81	0.46	x	0.34	0.82
4	3.85	4.59	0.34	x	0.41
5	4.35	5.21	0.82	0.41	x

In Slovak, the texts have a quite uniform *ductus*; there is no significant difference. In Persian, there is only one significant difference (between texts 1 and 3); the same holds true for Russian (texts 3 and 5); in Odia there is none. The greatest non-uniformity can be found in Turkish. Taking the *mean u* of 10 comparisons in every language, we obtain the "non-uniformity"-ordering as follows:

Turkish	2.412
Chinese	2.399
Croatian	1.642
Hungarian	1.578
German	1.273
Russian	1.062
Persian	0.988
Slovak	0.709
Odia	0.699

This ordering is surely not language-dependent, but text-dependent. Perhaps, the analysis of other text types could emphasize language families.

The test for similarity can be performed also using the chi-square test for a 2 x 2 table or applying Fisher's exact test. The test statistic is

$$(2.9)\quad \chi^2 = \frac{(A_1V_2 - A_2V_1)^2 n}{n_1 n_2 AV},$$

where $n_1 = A_1 + V_1$, $n_2 = A_2 + V_2$, $n = n_1 + n_2$, $A = A_1 + A_2$, $V = V_1 + V_2$. The result has the chi-square distribution with 1 degree of freedom and is approximately the square of *u* in formula (2.8). Consider, for example the Turkish texts T 1 and T 2. We have $A_1 = 62$, $V_1 = 58$, $n_1 = 120$; $A_2 = 84$, $V_2 = 84$, $n_2 = 165$, $A = 62 + 84 = 146$, $V = 58 + 84 = 142$, $n = n_1 + n_2 = 120 + 165 = 285$. Inserting these numbers in (2.9) we obtain $X^2 = 0.07838$. The square root yields 0.28 which is exactly the value given in Table 2.12 obtained in form of the *u*-text.

The indicators *Q* of individual texts can be considered a variable and the test for difference between two text groups or languages can be constructed also

using the mean *Q*-values. For German texts we obtain e.g. *Q*(German) = (0.65, 0.76, 0.63, 0.60, 0.59) yielding a mean $\bar{Q}$ = 0.65 and a variance *Var*(*Q*) = 0.00372. Hence, the variance of $\bar{Q}$ is *Var*($\bar{Q}$) = *V*(*Q*)/(*n*-1), in our case 0.00372/4 = 0.00093.

In order to perform an asymptotic normal test for difference of two languages (restricted to journalistic texts) we use

$$(2.10)\quad u = \frac{|\bar{Q}_1 - \bar{Q}_2|}{\sqrt{\frac{Var(Q_1)}{n_1 - 1} + \frac{Var(Q_2)}{n_2 - 1}}},$$

where *n* is the number of texts of the given group. Comparing all groups with one another we obtain the results presented in Table 2.13.

Table 2.13
Differences between mean Busemann coefficients in text groups

	Brazilian	**Portuguese**	**Slovak**	**Hungarian**	**Croatian**	**Chinese**	**Persian**
Brazilian	x						
Portuguese	1.4843	x					
Slovak	0.2494	1.2940	x				
Hungarian	2.1806	1.5420	2.0726	x			
Croatian	1.8719	2.7869	2.0542	3.1837	x		
Chinese	8.2732	12.3110	9.0437	7.2985	2.9476	x	
Persian	4.7658	4.5383	4.7926	0.6809	4.7025	13.9307	x
German	1.5833	3.0149	1.8528	3.1291	0.6918	5.2777	5.6366

	Odia	**Russian**	**Turkish**
Brazilian	1.7016	0.0963	3.3289
Portuguese	0.4895	0.0912	2.8412
Slovak	1.5411	0.0869	3.2489
Hungarian	1.2837	1.8509	1.2401
Croatian	2.9103	1.7324	4.0800
Chinese	11.3714	6.2646	7.6475
Persian	3.6542	3.4450	0.9344
German	3.1256	1.3721	4.1080
Odia	x	1.1288	2.6043
Russian		x	2.9689
Turkish			x

The test is two-sided because there is no reason to propose one-sided trends.

Usually, one performs a classification or expresses the results by means of a graph. However, if there are many languages/text types, one obtains results which are not quite lucid and do not help theory construction. First many texts must be analyzed, then a hypothesis concerning text types and languages must be formulated and tested; the resulting graph can be illuminating in the end. Using factor analysis one can obtain a kind of grouping of the languages.

2.2. Sequential measurement

The activity of the texts can be captured also dynamically. There are two possibilities: (1) One counts the proportion of *V* stepwise; in this way one obtains either a beta function or a Morse function (cf. Popescu, Čech, Altmann 2013); the curves need not be monotone increasing but they can easily be integrated into the unified theory. (2) One counts the number of verbs (*X*) up to the Y^{th} adjective. The sequence is non-decreasing and the resulting curve is a characteristic of the text (cf. Ziegler, Best, Altmann 2002). Mostly it is a power function but it may be both convex and concave, it can even display a sigmoid character. Consider, for example, the sequence VAAVVAAAV, then up to the first *A* there is 1 *V*; up to the second *A* there is 1 *V*; up to the third *A* there are 3 *V*'s, etc. We obtain a sequence

A	1	2	3	4	5	6
V	1	1	3	3	3	4

The non-existing last (sixth) *A* has been added in order to include all *V*'s. Texts can be classified according to the course of the function (convex, concave, sigmoid, wave form) or according to the values of the parameters of the applied functions. Texts can be compared with one another in different ways using different methods.

Consider first the Slovak text T 1 where we obtain the sequence

[A,V,V,V,A,V,V,A,V,A,V,A,V,A,V,V,A,V,A,A,A,V,A,A,V,V,V,A,A,V,A,V,A,
V,A,A,V,A,A,A,A,V,V,V,A,V,A,A,V,V,A,V,A,V,V,A,V,V,A,V,V,A,V,V,A,A,
V,V,A,A,V,V,A,V]

A	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
V	0	3	5	6	7	8	10	11	11	11	12	12	15	15	16	17	18	18	19	19	19	19	22	23

25	26	27	28	29	30	31	32	33	34	35	36	
23	25	26	28	30	32	34	34	36	36	38	39	

Fitting the power function $y = ax^b$ to this data we obtain $y = 1{,}0976x^{0{,}9818}$ with $R^2 = 0.9736$. The parameter *a* shows merely the beginning of the course; the para-

meter *b* is the measure of activity. The greater is *b* in the exponent of the power function, the stronger is the course of activity.

The parameters of the curves for the data in which it is adequate are shown in Table 2.14.

Table 2.14
Slovak. Sequential measurement
The power function $y = ax^b$ for the course of activity in journalistic texts

Text	***a***	***b***	$\mathbf{R^2}$
T 1	1.0976	0.9818	0.9736
T 2	3.2768	0.8096	0.9704
T 3	0.5767	1.2638	0.9835
T 4	1.6436	0.8787	0.9618
T 5	1.0426	1.0350	0.9766

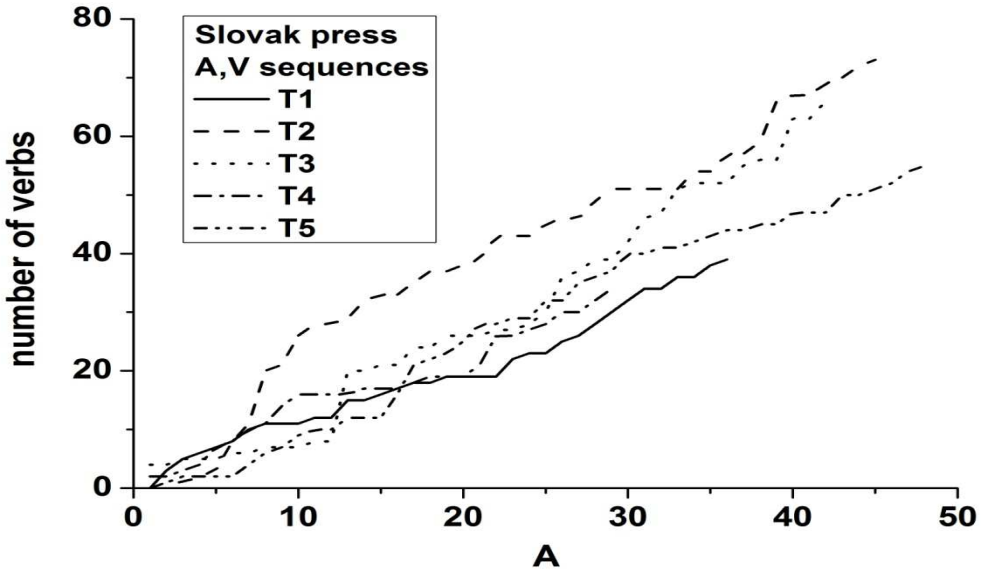

Figure 2.1. Increase of activity: Slovak

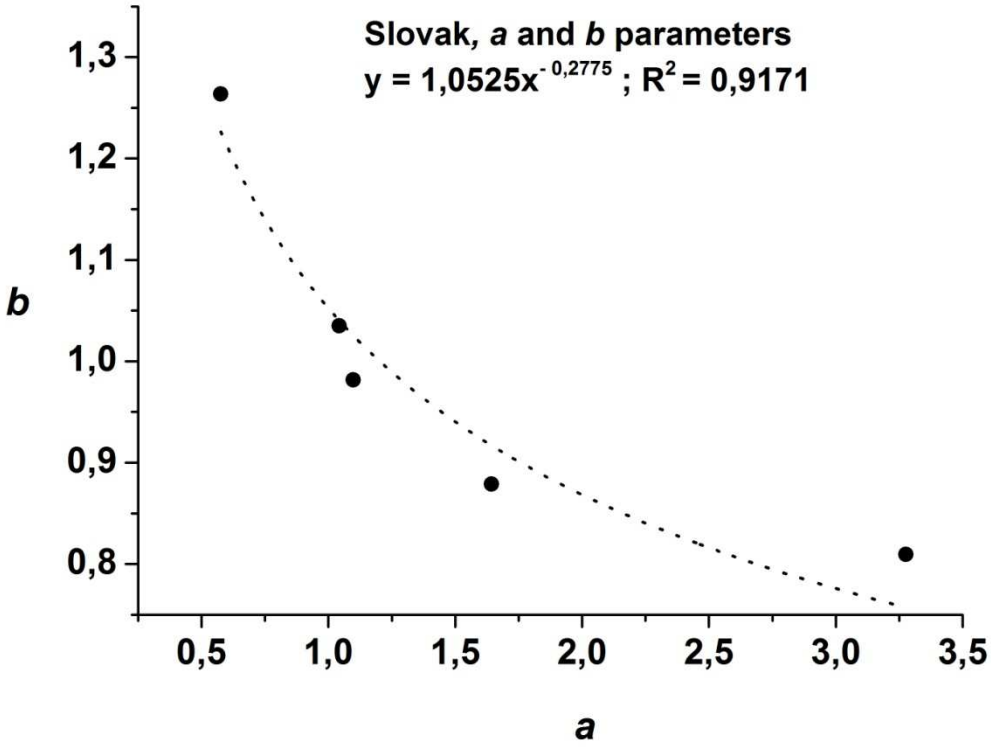

Figure 2.2. Dependence of *b* on *a*, Slovak

Table 2.15
Croatian. Sequential measurement
The power function $y = ax^b$ for the course of activity in journalistic texts

Text	***a***	***b***	**R^2**
T 1	2.4672	1.3296	0.9258
T 2	1.4117	1.3492	0.9507
T 3	1.7701	0.9831	0.9846
T 4	1.2099	1.0510	0.9929
T 5	0.1526	1.7207	0.9417

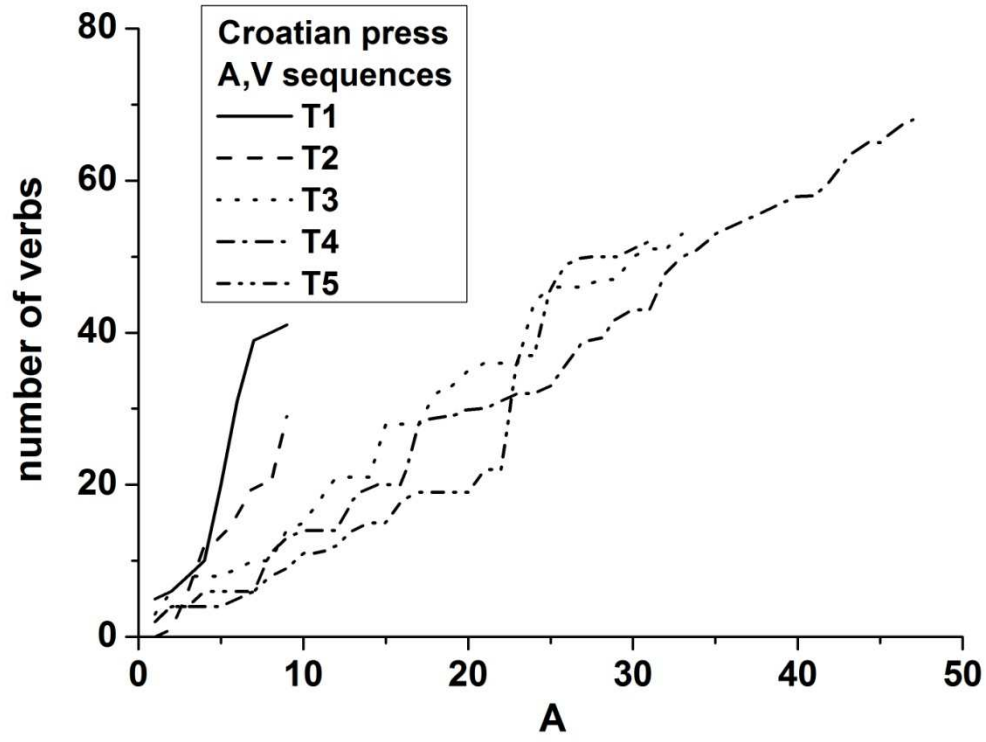

Figure 2.3. Increase of activity in Croatian

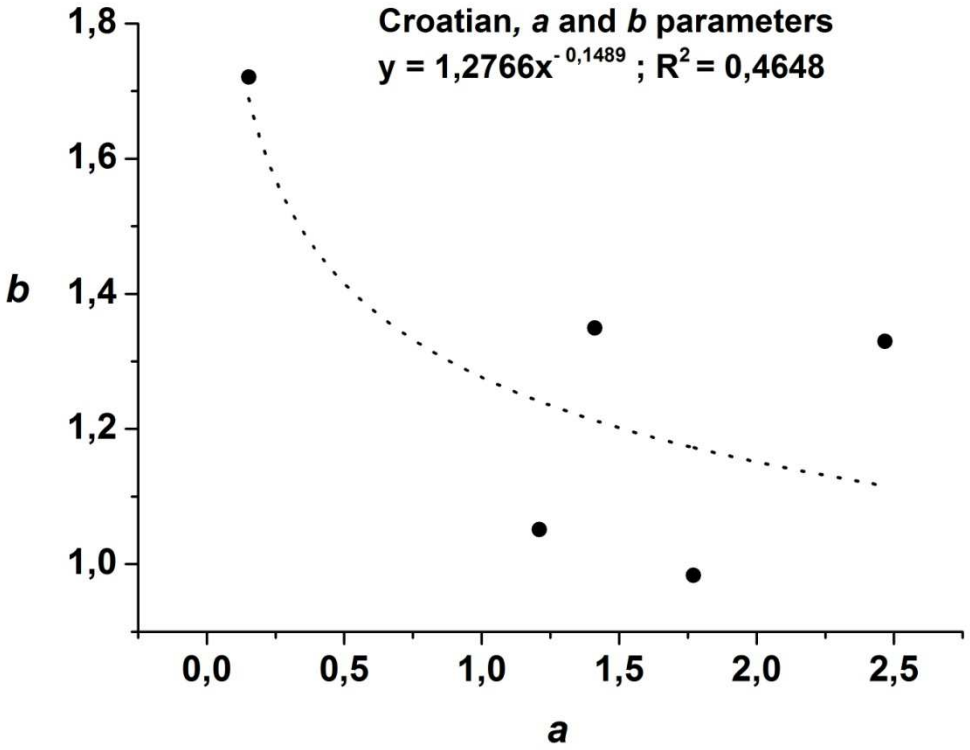

Figure 2.4. Dependence of *b* on *a* in Croatian

As can be seen, the first two Croatian texts strongly deviate from the course of the other ones. The course in the fifth text is deviating, hence the power function does not capture the trend significantly. But perhaps the study of more journalistic text would improve the general tendency.

Table 2.16
Hungarian. Sequential measurement
The power function $y = ax^b$ for the course of activity in journalistic texts

Text	***a***	***b***	$\mathbf{R^2}$
T 1	0.1890	1.5671	0.9534
T 2	0.9555	0.8570	0.9615
T 3	0.1515	1.5953	0.9571
T 4	0.2048	1.3336	0.9946
T 5	0.7697	0.9933	0.9881

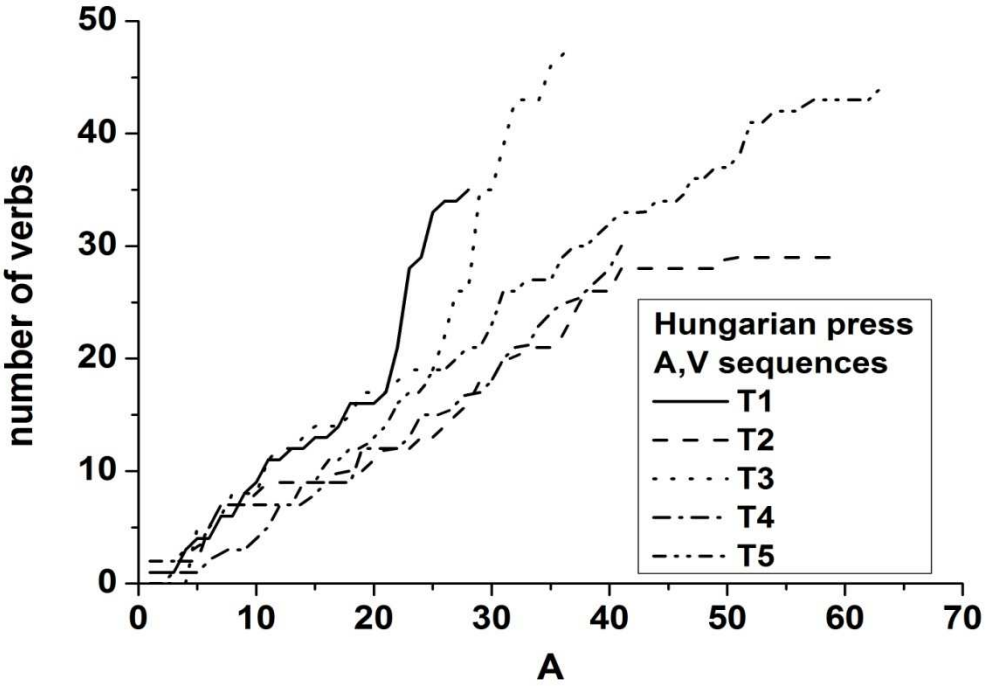

Figure 2.5. Increase of activity in Hungarian

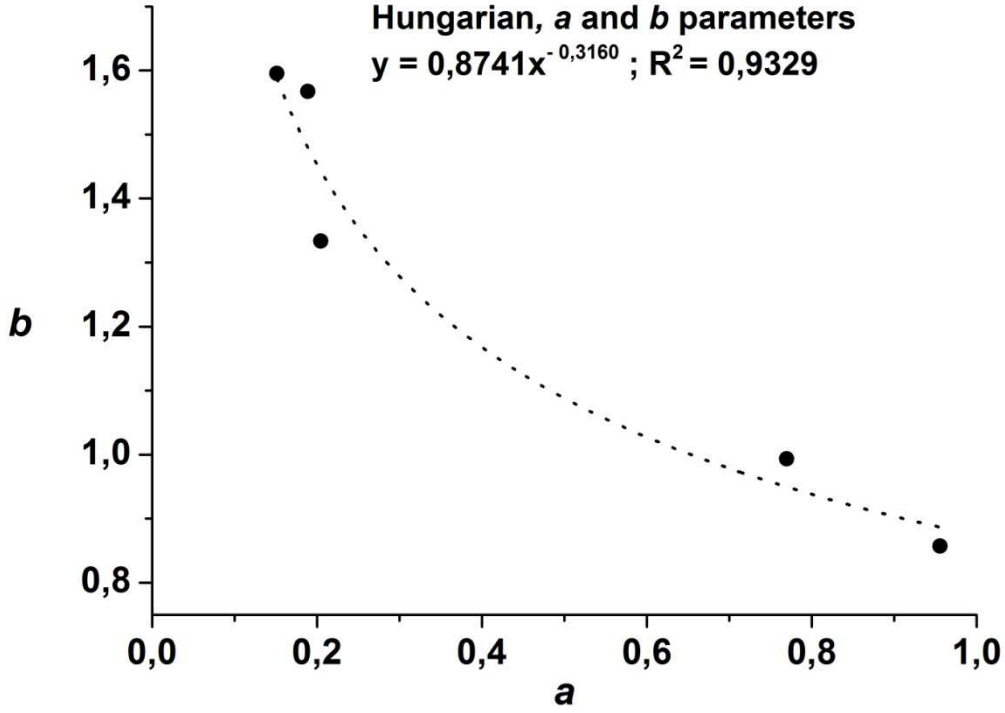

Figure 2.6. Dependence of *b* on *a* in Hungarian

Table 2.17
Chinese. Sequential measurement
The power function $y = ax^b$ for the course of activity in journalistic texts

Chinese			
Text	***a***	***b***	**R^2**
T 1	2.0700	1.1752	0.9585
T 2	4.3182	1.0432	0.9948
T 3	22.5892	0.8444	0.9602
T 4	5.6749	0.9985	0.9846
T 5	7.7702	1.0003	0.9854

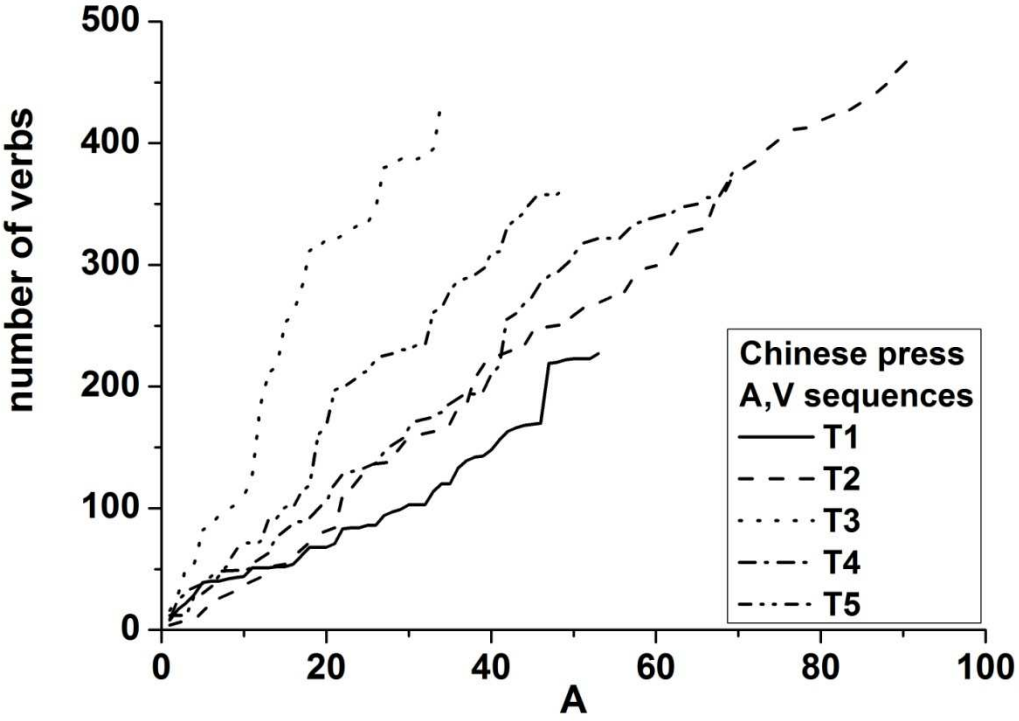

Figure 2.7. Increase of activity in Chinese

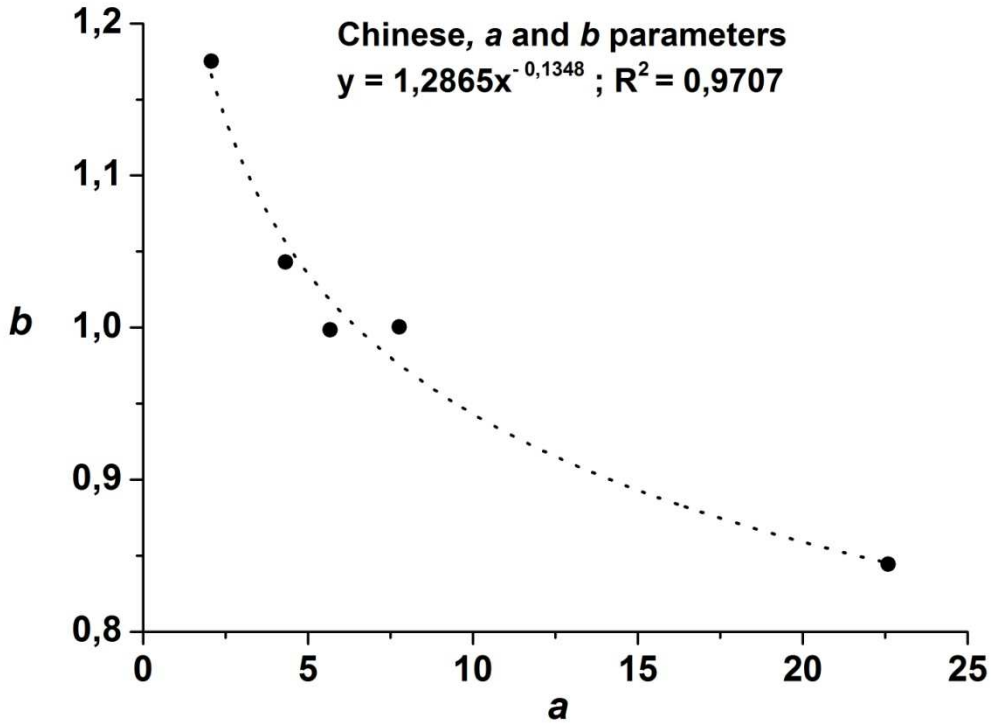

Figure 2.8. Dependence of *b* on *a* in Chinese

Table 2.18
Persian. Sequential measurement
The power function $y = ax^b$ for the course of activity in Persian journalistic texts

Persian			
Text	***a***	***b***	**R^2**
T 1	2.1433	0.8330	0.9818
T 2	0.1755	1.2787	0.9959
T 3	0.1795	1.2186	0.9920
T 4	3.4579	0.7054	0.9910
T 5	2.6828	0.7402	0.9914

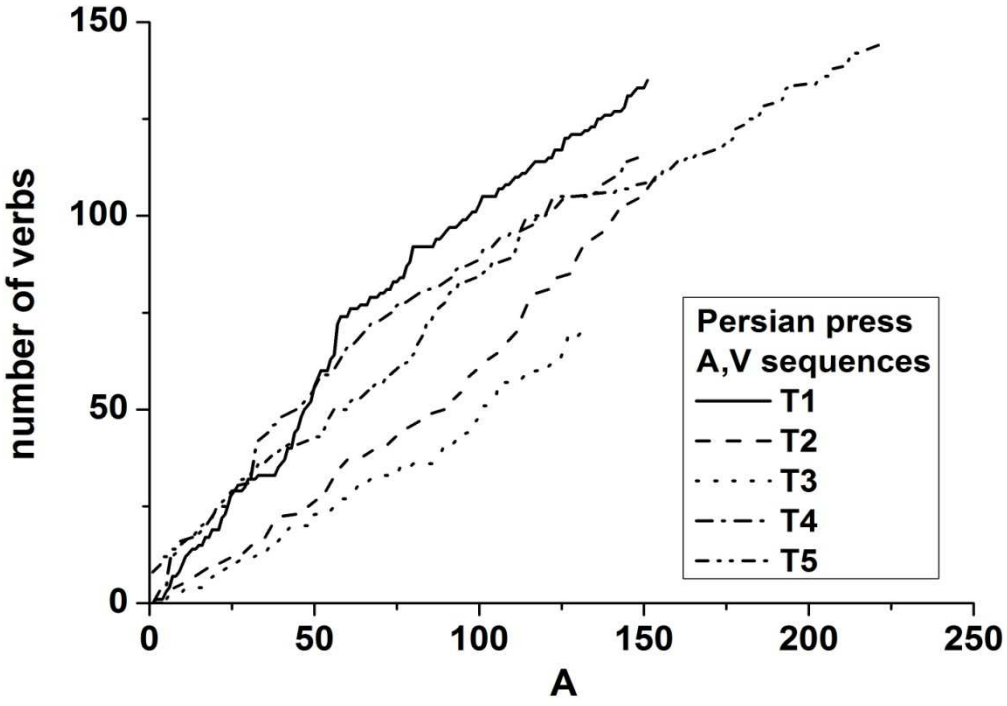

Figure 2.9. Increase of activity in Persian

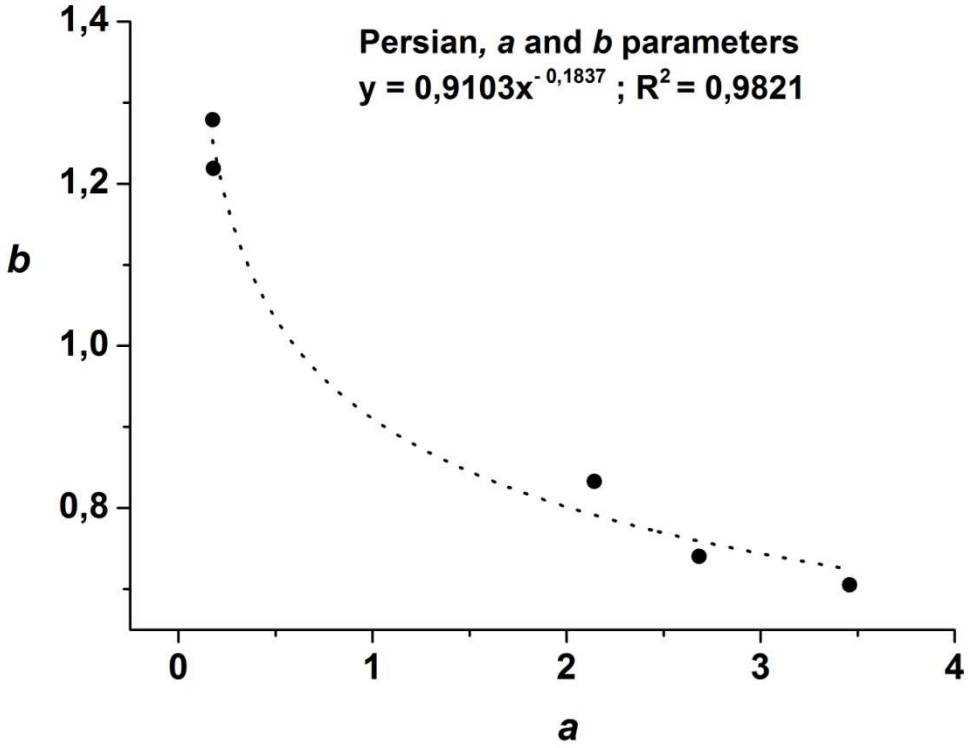

Figure 2.10. Dependence of *b* on *a* in Persian

Table 2.19
German. Sequential measurement
The power function $y = ax^b$ for the course of activity in German journalistic texts

German			
Text	***a***	***b***	**R^2**
T 1	0.3402	1.5179	0.8863
T 2	2.4649	1.1149	0.9848
T 3	2.8359	0.8226	0.9605
T 4	0.4490	1.3462	0.9867
T 5	1.9394	0.9004	0.9822

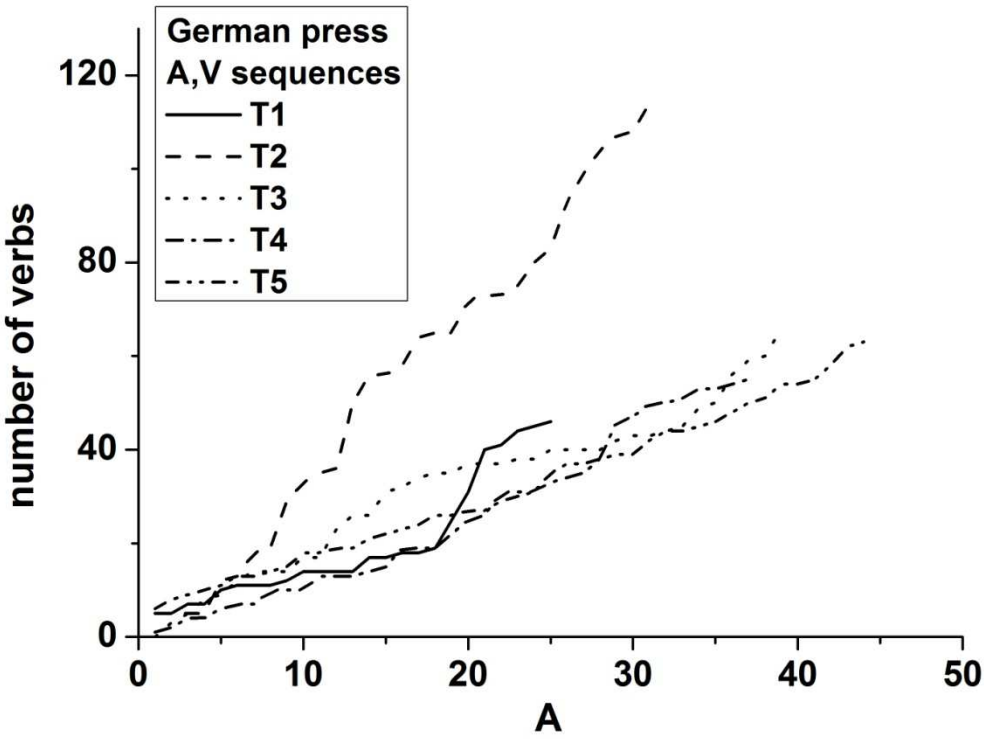

Figure 2.11. Increase of activity in German

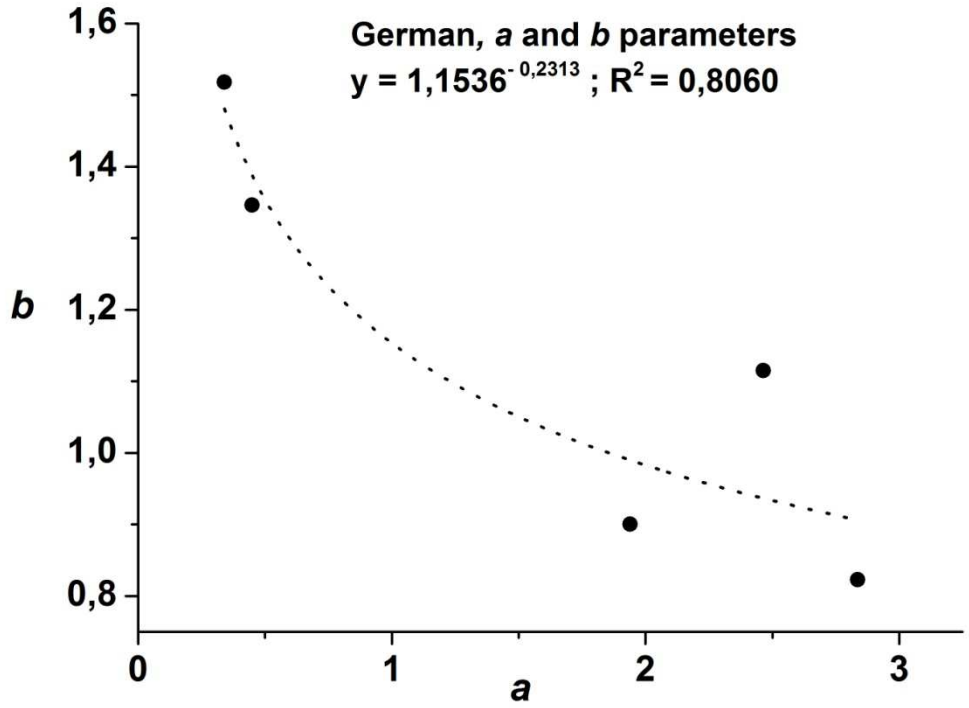

Figure 2.12. Dependence of *b* on *a* in German

In German, the deviation in one text as compared with the other ones may cause a strong reduction of the determination coefficient.

Table 2.20
Odia. Sequential measurement
The power function $y = ax^b$ for the course of activity in Odia journalistic texts

Odia			
Text	***a***	***b***	**R^2**
T 1	2.2254	0.8141	0.9869
T 2	1.3488	0.9541	0.9847
T 3	1.0561	1.0339	0.9680
T 4	1.3301	0.8691	0.9850
T 5	0.4917	1.2050	0.9822

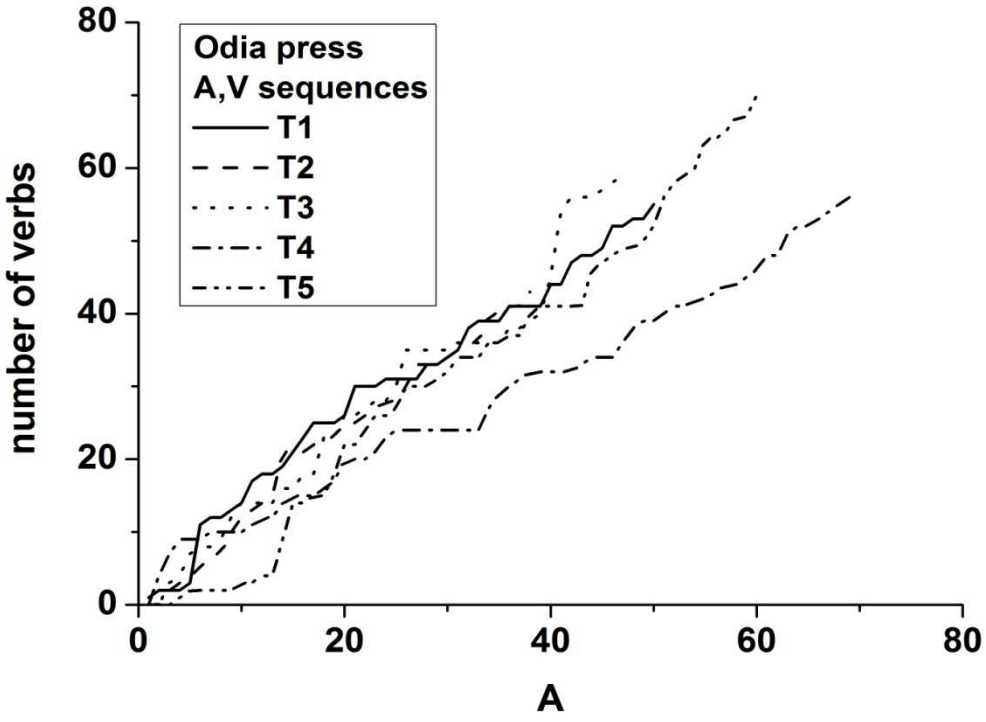

Figure 2.13. Increase of activity in Odia

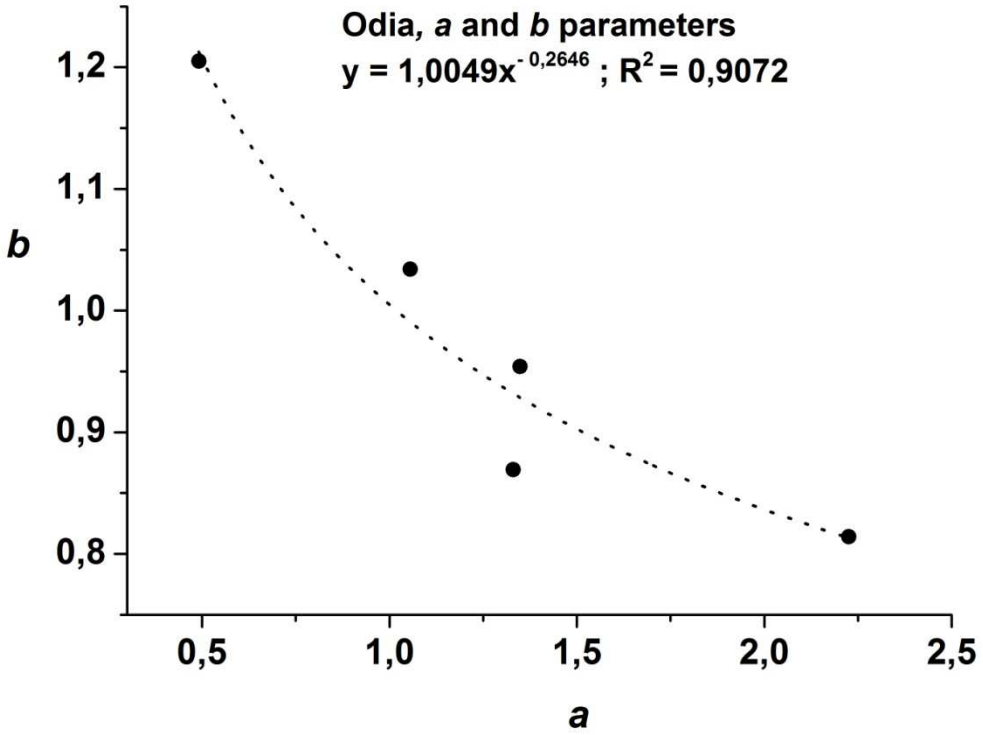

Figure 2.14. Dependence of *b* on *a* in Odia

Table 2.21
Russian. Sequential measurement
The power function $y = ax^b$ for the course of activity in Russian journalistic texts

Russian			
Text	***a***	***b***	**R^2**
T 1	2.1342	0.9021	0.9086
T 2	1.4683	0.9987	0.9732
T 3	6.3161	0.4615	0.8965
T 4	3.2392	0.6556	0.9426
T 5	1.0471	0.9270	0.9903

As to the dependence of *b* on *a*, in Russian and Turkish one can observe a slight deviation from the usual course. For a better fit one can use the Zipf-Alekseev function $y = ax^{(b + c \ln x)}$ which, perhaps, better captures the local deviation. Here *y* is our parameter *b* and *x* is our parameter *a*. As can be seen in Figures 2.16 and 2.18, this adaptation is better, but one cannot know whether further texts would speak in the favour of the power function.

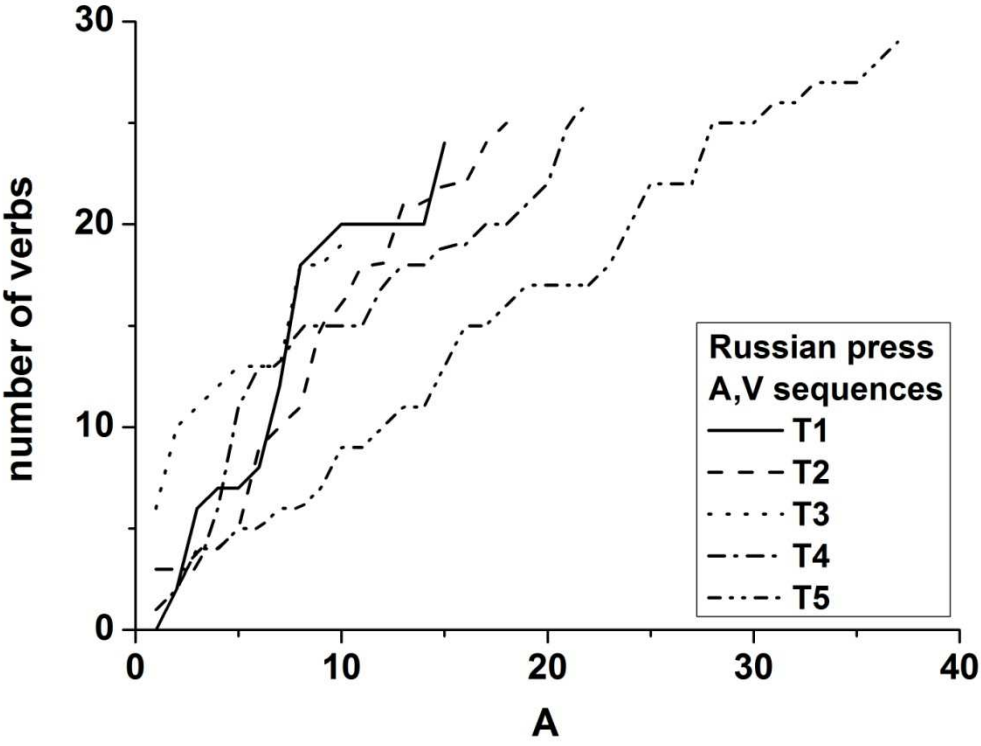

Figure 2.15. Increase of activity in Russian

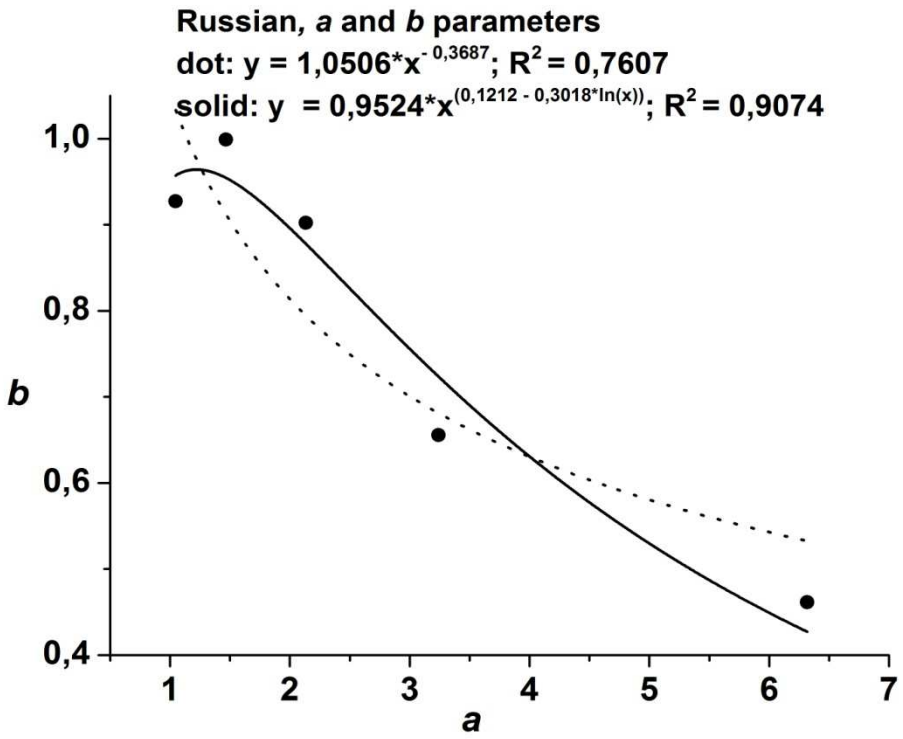

Figure 2.16. Dependence of *b* on *a* in Russian

Table 2.22
Turkish. Sequential measurement
The power function $y = ax^b$ for the course of activity in Turkish journalistic texts

Turkish			
Text	***a***	***b***	$\mathbf{R^2}$
T 1	0.0465	1.7132	0.9622
T 2	0.8425	1.0272	0.9937
T 3	0.1728	1.1369	0.9917
T 4	0.2559	1.0757	0.9961
T 5	0.0835	1.2630	0.9910

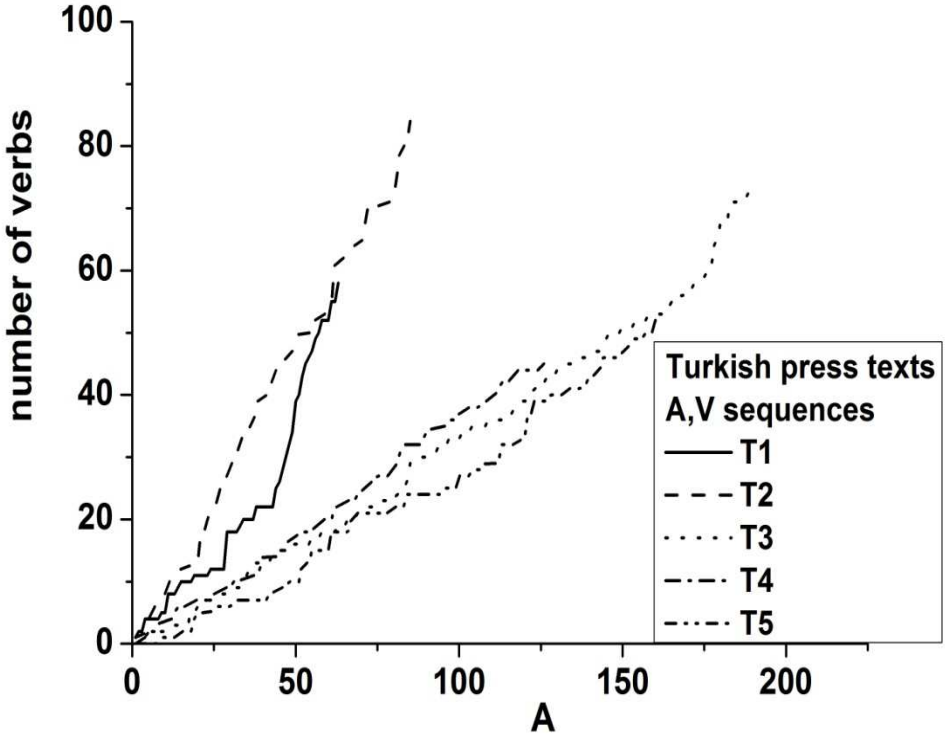

Figure 2.17. Increase of activity in Turkish

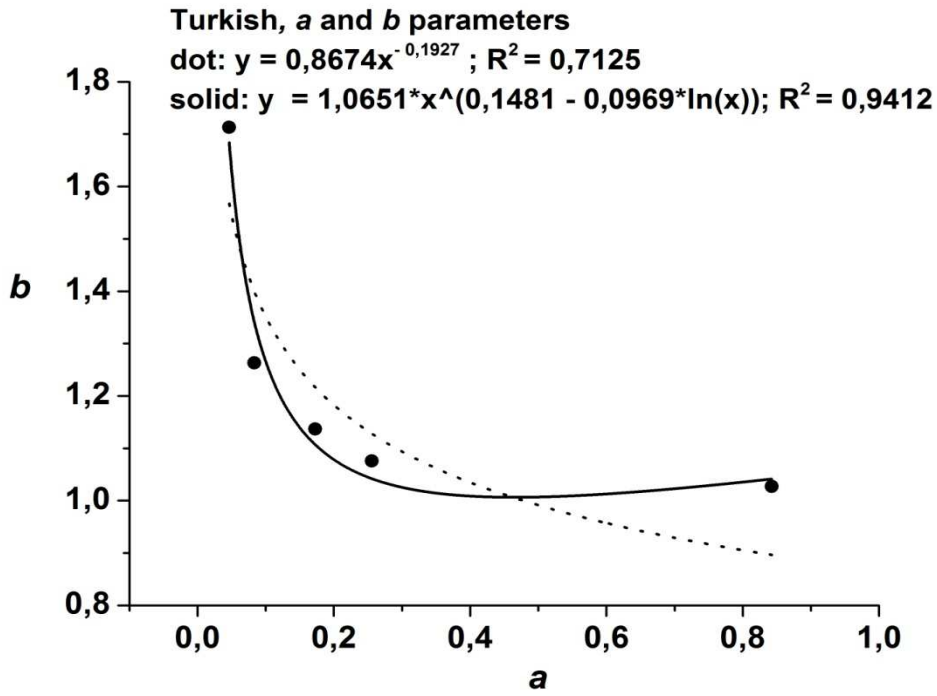

Figure 2.18. Dependence of *b* on *a* in Turkish

Nevertheless, it can be shown in all cases that the greater the parameter *a*, the smaller gets *b*. It can be supposed that, in general, the runs of *V* do not have a constant length. They are controlled by some kind of self-regulation evoked by the beginning of the text.

We can state that from the sequential point of view there is some regularity operating in the text. We do not want to call it a law because we still miss an explicative derivation and its links to other properties of texts. Besides, the number of texts and languages analyzed so far is very modest.

2.3. Runs

Regularities of sequential organization can be studied from different points of view using different methods. Since we consider merely two categories (*A* and *V*), we can test the randomness/structuring of the sequence using the theory of runs. A run is an uninterrupted sequence of the same symbol. These methods have been used in text analysis several times (cf. e.g. Altmann 1988; Altmann, Altmann 2008). Our problem is (1) to state whether the sequence of *A*'s and *V*'s contains too many/few runs, (2) to compare the run structure of two texts. That means, runs present a kind of view of text development and not a simple evaluation of frequencies. A significantly active text can have a quite "normal" sequential structure and vice versa.

Let the number of *A*'s in the text be n_A, that of *V*'s n_V and $n_A + n_V = n$. Let the number of runs of *A*'s be r_A, that of *V*'s r_V and $r_A + r_V = r$. That means, *n* is

the number of entities in the sequence and *r* is the number of runs. We shall consider all texts as large samples, i.e. we shall use always approximations.

Example. Let us consider a slightly longer text, i.e. one with many *A*'s and *V*'s, e.g. the Slovak text T 4 = [V,V,A,A,V,A,V,A,V,V,V,A,V,A,V,V,A,V, A,V,V,V,A,V,V,A,A,A,A,V,A,A,A, V,A,V,A,A,A,V,V,A,V,V,V,V,V, A,A,V,A, V,A,V,V,A,A, V,V,A,V,V,A]

There are

$n_A = 29$, $n_V = 34$, $n = 63$, $r_A = 19$, $r_V = 19$, $r = 38$,

The asymptotic normal test for the randomness of the sample is defined as (cf. Brownlee 1960: 169)

$$(2.11)\quad z = \frac{n(r-1) - 2n_A n_V}{\left[\frac{2n_A n_V (2n_A n_V - n)}{n-1}\right]^{1/2}}$$

We decide as follows: If z > 1.96, then the number of runs is significantly large; if z < −1.96, then the number of runs is significantly small. A significantly large number of runs means a rather conscious alternation of *A*'s and *V*'s; a significantly small number of runs means a kind of structuring the text, making preferences place-wise, the heaping of adjectives and verbs. In a text with *z* ε <−1.96; 1.96> there is no structuring, the number of runs is random. Here we obtain

$$z = \frac{63(38-1) - 2(29)34}{\left[\frac{2(29)34[2(29)34 - 63]}{63-1}\right]^{1/2}} = 1.46$$

that means, even if the text displays a preference for activity, the ordering of entities is random; there is no significant (too small or too large) number of runs.

The results concerning journalistic texts whose *A*—*V*-sequences were available to us are presented in Table 2.23

Table 2.23
Runs of *A* and *V*

Text	$\mathbf{n_A}$	$\mathbf{n_V}$	**n**	$\mathbf{r_A}$	$\mathbf{r_V}$	**r**	**z**
Slovak							
T 1	35	39	74	24	24	48	2.37
T 2	44	73	117	30	30	60	0.81
T 3	41	66	107	23	24	47	-0.94

T 4	29	34	63	19	19	38	1.46
T 5	47	55	192	31	31	62	2.07
Hungarian							
T 1	27	35	62	17	18	35	0.92
T 2	59	29	88	21	21	42	0.51
T 3	37	48	85	20	19	39	-0.84
T 4	41	29	70	23	22	47	2.99
T 5	63	44	107	30	39	60	0.44
Croatian							
T 1	8	42	50	8	8	18	1.93
T 2	8	29	37	7	7	14	0.23
T 3	32	52	84	18	18	37	-0.84
T 4	46	66	112	32	33	65	1.92
T 5	31	52	83	18	18	36	-0.91
Chinese							
T 1	52	225	277	39	40	79	-1.28
T 2	91	470	561	82	82	164	1.64
T 3	33	436	469	30	31	61	-0.48
T 4	70	382	452	57	57	114	-0.96
T 5	48	362	410	43	44	87	0.30
Persian							
T 1	150	135	285	79	79	158	1.77
T 2	154	111	265	72	72	144	1.77
T 3	131	70	201	52	52	104	1.83
T 4	147	115	262	71	71	142	1.50
T 5	222	145	367	92	93	185	0.94
German							
T 1	24	46	79	15	16	31	-0.41
T 2	30	114	144	24	25	49	0.13
T 3	38	66	104	26	26	52	0.59
T 4	37	55	92	25	25	30	1.04
T 5	42	61	1ß3	33	34	67	3.33
Odia							
T 1	24	46	79	15	16	31	-0.41
T 2	30	114	144	24	25	49	0.13
T 3	38	66	104	26	26	52	0.59
T 4	37	55	92	25	25	30	1.04
T 5	42	61	103	33	34	67	3.33
Russian							
T 1	15	24	39	10	9	19	-0.16
T 2	18	25	43	14	14	28	1.93
T 3	9	19	28	6	7	13	-0.10
T 4	21	26	47	12	13	25	0.23

T 5	37	29	66	21	21	42	2.14
Turkish							
T 1	62	58	129	29	30	59	-0.35
T 2	84	84	168	45	46	91	0.92
T 3	188	73	261	62	62	124	2.75
T 4	125	45	179	40	40	80	2.54
T 5	159	52	211	35	35	70	-1.74

Again, the run structure of two texts can be compared using the normal test. Several linguistic examples can be found in Grotjahn (1980). Let us define

n_{1A} n_{1V} n_1 r_1 for the first text and
n_{2A} n_{2V} n_2 r_2 for the second text.

The expectation of the number of runs for the given text is

$$(2.12)\quad E(r) = \frac{2n_A n_V + n}{n}$$

and the variance

$$(2.13)\quad Var(r) = \frac{2n_A n_V (2n_A n_V - n)}{n^2 (n-1)}.$$

The asymptotic normal test is then defined as

$$(2.14)\quad u = \frac{r_1 - r_2 - [E(r_1) - E(r_2)]}{\sqrt{Var(r_1) + Var(r_2)}}.$$

As an example let us compare the Slovak texts 1 and 2. According to Table 2.2.3 we have $r_1 = 48$, $r_2 = 60$, $E(r_1) = [2(35)39 + 74]/74 = 37.89$, $E(r_2) = [2(44)73 + 117]/117 = 55.91$, $Var(r_1) = \{2(35)39[2(35)29 - 74]\}/[74^2(73)] = 18.1386$, $Var(r_2) = \{2(44)73\{2(44)73 - 117]\}/[117^2(116)] = 25.5152$, hence

$$u = \frac{48 - 60 - (37.89 - 55.91)}{\sqrt{18.1386 + 25.5152}} = 0.91$$

All results are presented in Table 2.24.

Table 2.24
u-test for comparison between *AV*-runs of individual **Slovak** texts

Text	1	2	3	4	5
1	x	0.91	2.27	0.76	-5.13*
2	0.91	x	1.23	0.25	5.55*
3	2.27*	1.23	x	1.64	7.29*
4	0.76	0.25	1.64	x	6.52*
5	-5.13*	5.55*	7.39*	6.52*	x

As can be seen, only text 5 differs significantly from all the rest. Besides, there is a significant difference between text 1 and 3.

Table 2.25
u-test for comparison between *AV*-runs of individual **Hungarian** texts

Text	1	2	3	4	5
1	x	0.25	1.23	-1.53	-0.58
2	0.25	x	0.16	-0.69	-0.78
3	1.23	0.16	x	-3,65*	-2,.47*
4	-1.53	-0.69	-3.65*	x	0.76
5	-0.58	-0.78	-2.47*	0.76	x

In Hungarian there are two significant differences.

Table 2.26
u-test for comparison between *AV*-runs of individual **Croatian** texts

Text	1	2	3	4	5
1	x	1.14	1.54	-1.15	1.60
2	1.14	x	0.86	-1.72	0.92
3	1.54	0.86	x	-2.01*	0.04
4	-1.15	-1.72	-2.01*	x	2.06*
5	1.60	0.92	0.04	2.06*	x

In Croatian there are 2 significant differences.

Table 2.27
u-test for comparison between *AV*-runs of individual **Chinese** texts

Text	1	2	3	4	5
1	x	-2.08	-0.89	-0.15	-1.18
2	-2.08	x	1.69	1.87	1.21
3	-0.89	1.69	x	0.64	-0.51
4	-0.15	1.87	0.64	x	-0.95
5	-1.18	1.21	-0.51	-0.95	x

Table 2.28
u-test for comparison between *AV*-runs of individual **Persian** texts

Text	1	2	3	4	5
1	x	0.08	0.30	0.25	0.51
2	0.08	x	0.22	0.18	0.45
3	0.30	0.22	x	-0.02	0.28
4	0.25	0.18	-0.02	x	0.28
5	0.51	0.45	0.28	0.28	x

Table 2.29
u-test for comparison between *AV*-runs of individual **German** texts

Text	1	2	3	4	5
1	x	0.06	-0.02	0.56	-0.42
2	0.06	x	-0.06	0.43	0.40
3	-0.02	-0.06	x	0.42	-0.29
4	0.56	0.43	0.42	x	-1.52
5	-0.42	0.40	-0.29	-1.52	x

Table 2.30
u-test for comparison between *AV*-runs of individual **Odia** texts

Text	1	2	3	4	5
1	x	-0.31	-0.13	3.12	-2.46
2	0.31	x	-0.37	2.61	-2.52
3	-0.13	-0.37	x	2.74	-1.99
4	3.12	2.61	2.74	x	-4.71
5	-2.46	-2.52	-1.99	-4.71	x

In Odia, seven out of ten comparisons are significant.

Table 2.31
u-test for comparison between *AV*-runs of individual **Russian** texts

Text	1	2	3	4	5
1	x	-1.52	-0.07	-0.28	-1.82
2	-1.52	x	1.62	1.15	-0.48
3	-0.07	1.62	x	-0.24	-1.91
4	-0.28	1.15	-0.24	x	-1.49
5	-1.82	-0.48	-1.91	-1.49	x

Table 2.32
u-test for comparison between *AV*-runs of individual **Turkish** texts

Text	1	2	3	4	5
1	x	-0.46	-1.92	-2.06	1.60
2	-0.46	x	-1.29	-1.27	1.83
3	-1.92	-1.29	x	0.21	3.22
4	-2.06	-1.27	0.21	x	3.58
5	1.60	1.83	3.22	3.58	x

The above tables show the extent of uniformity in journalistic texts. If we take into account the ten comparisons in every language and count the significant differences, we obtain the order

0 Russian, German, Persian,
1 Chinese
2 Croatian, Hungarian
3 Turkish
5 Slovak
7 Odia

Hence Russian, German and Persian have a steady run structure.

No family relationship can be recognized here, but this is merely the first modest step in this type of investigation.

Our aim is rather to see how one type of specification (activity vs. descriptiveness) increases in texts. Though it can be expected that phenomena expressed in texts have some properties and behave in some way (i.e. express some activity), we conjecture that whatever the kind of the text type, there is some law in the background. One cannot learn to act according to laws (just as in physics), one simply obeys them. In this way some regularities can be observed in texts created subconsciously by the writers. Here we scrutinized merely some of them, but there are surely many others whose detection necessitates time and teams. One can approach new vistas both inductively and deductively, and, finally, everything must be inserted in a theory.

3. Nominality

If in a language nouns differ formally from verbs, then the same event can be expressed either using a verbal or an equivalent nominal phrase. In English one simply uses different synsemantics, e.g. *to laugh* vs. *the laugh;* in German one places an article before the verb to obtain a noun, e.g. *lachen* vs. *das Lachen;* practically all languages have some means to change a dynamic expression into a static one. There are styles which prefer nominality because it sounds more "formal", e.g. judicial texts. In German, this fact is known since the 14th century. There are, of course, nouns designating more or less action, but here we do not perform measurements of the degree of activity.

Nominality as a whole can be measured (a) either by comparing the number of nouns in a text with that of all other parts of speech, or (b) simply with that of verbs in order to evaluate the contrast (cf. Ziegler, Best, Altmann 2002). Both views strongly depend on grammar and writing customs.

Nominality can be scrutinized from different points of view: (i) To show the state (way of expression) of the given text in general; (ii) to study the course of nominality in individual parts of the text, e.g. chapters, strophes, stage play acts; (iii) to study the linguistic behavior of individual persons in a novel or a stage play; (iv) to evaluate the nominalizing technique in individual languages, i.e. for language comparison; (v) to evaluate the historical development of language from one type to another; (vi) to search for the relationship between nominality and other properties of language, i.e. to search for laws which may exist in this domain and incorporate nominality in text or language theory following Köhler's synergetic approach (1986, 2005). There is, for example, a hypothesis linking nominality with sentence length: the stronger the nominality the shorter the sentence length (cf. Bußman 1990: 530). Here we shall merely touch some problems, but, in general, we shall adhere to the description of nominality vs. specification. The rest is left for future research.

If we investigate nominality vs. specification, we compare the number of nouns (which name the objects) with that of predicates of the first level, namely adjectives and verbs. The first type says how the object is, the second says what it does. The words of the sentence can be scaled according to the level of their predicativity (specification level) – e.g. adverbs are predicates of the second level because they say something about (specify) the adjectives and verbs, etc. One could analyze the sentence also using the philosophical concept of predicates starting from Aristotle, but we shall consider here merely the first level. The result can be generalized.

3.1. Nominality vs. predicativity

The first problem is stating the equilibrium between nominality and predicativity. If the text is to some extent stylistically "neutral", one may expect that each noun

is specified by an adjective and it "does" something. Hence equilibrium is attained if each of these parts of speech is represented equally. Considering here $n = A + N + V$, one can test the hypothesis of equilibrium using the chi-square test

$$(3.1) \quad X^2 = \frac{(A - n/3)^2}{n/3} + \frac{(N - n/3)^2}{n/3} + \frac{(V - n/3)^2}{n/3}.$$

The result depends, in part, on the way of counting. If one omits auxiliary verbs, modal verbs, and compound verb forms, languages will be more similar than with considering exhaustedly everything, i.e. every word in the sentence. For the sake of illustration let us consider the first journalistic texts in Brazilian-Portuguese. We obtain the vector

(A,N,V) = (41, 165, 168)

(n = 374, n/3 = 124.67), and the chi-square is

$X^2 = (41 - 124.67)^2/124.67 + (165 - 124.67)^2/124.67 + (168 - 124.67)^2/124.67 =$
$= 56.15 + 13.05 + 15.06 = 84.26$

a very high value signaling strong disequilibrium. The greatest contribution to the chi-square is yielded by the small number of adjectives, hence, the text is non-descriptive, as shown also above.

Testing simply nominality vs. predicativity, the expected value of N is n/3 and that of $A+V = 2n/3$ that is

$$(3.2) \quad X^2 = \frac{(A + V - 2n/3)^2}{2n/3} + \frac{(N - n/3)^2}{n/3}.$$

In the above example we obtain

$X^2 = (41 + 168 - 2(374)/3)^2/[2(374/3)] + (165 - 374/3)^2/(374/3) = 19.57.$

This value indicates a significant deviation from the randomness. Since the number of nouns is much greater than expected, the text can be considered as significantly non-predicative (= nominative). Computing the chi-square values for all text we obtain the results presented in Table 3.1.

Table 3.1
Testing for nominativity vs. predicativity in 11 languages
(SN = significantly nominative, SP = significantly predicative, NE = neutral)

Text No	Vector (A,N,V)	Sum	X^2	Type
Brazilian-Portuguese (Ziegler 1998)				
1	(41, 165, 168)	374	19.57	SN
2	(32, 119, 61)	212	49.59	SN
3	(40, 255, 130)	425	136.00	SN
4	(208, 689, 174)	1071	463.13	SN
5	(115, 238, 127)	480	57.04	SN
6	(82, 308, 193)	583	99.73	SN
7	(114, 392, 154)	660	201.71	SN
8	(147, 338, 142)	627	119.43	SN
9	(71, 163, 91)	325	41.38	SN
10	(54, 129, 36)	219	64.44	SN
11	(132, 383, 137)	652	189.42	SN
12	(32, 98, 128)	258	2.51	NE
13	(139, 391, 168)	698	161.62	SN
14	(132, 496, 156)	784	316.08	SN
15	(181, 458, 128)	767	240.19	SN
16	(96, 298, 97)	491	165.39	SN
17	(83, 200, 85)	368	73.13	SN
18	(46, 203, 97)	346	99.96	SN
19	(84, 223, 83)	390	99.80	SN
20	(59, 146, 141)	346	12.23	SN
21	(43, 117, 80)	240	25.67	SN
Portuguese (Ziegler 2001)				
1	(45, 96, 47)	188	26.60	SN
2	(30, 86, 28)	144	45.13	SN
3	(30, 100, 45)	175	44.64	SN
4	(28, 119, 41)	188	75.96	SN
5	(39, 115, 54)	208	45.12	SN
6	(47, 126, 54)	227	50.22	SN
7	(52, 92, 48)	192	18.38	SN
8	(44, 135, 56)	235	61.49	SN
9	(45, 131, 70)	246	43.92	SN
10	(41, 134, 63)	238	56.50	SN
11	(61, 164, 45)	270	91.27	SN
12	(68, 178, 61)	307	83.92	SN
13	(46, 140, 45)	231	77.32	SN

14	(27, 116, 38)	181	77.04	SN
15	(39, 163, 35)	237	133.97	SN
16	(44, 111, 66)	221	28.38	SN
17	(27, 108, 39)	174	64.66	SN
18	(29, 145, 24)	198	141.84	SN
19	(43, 146, 53)	242	79.37	SN
20	(34, 181, 37)	252	168.02	SN
Slovak				
T1	(35, 126, 39)	200	79.21	SN
T2	(44, 124, 73)	241	35.60	SN
T3	(41, 148, 66)	255	70.04	SN
T4	(29, 104, 34)	167	62.95	SN
T5	(47, 150, 55)	252	77.79	SN
Hungarian				
T1	[27,85,35]	147	39.67	SN
T2	[59,111,29]	199	45.12	SN
T3	[37,121,48]	206	59.83	SN
T4	[41,102,29]	173	51.90	SN
T5	[63,109,43]	216	28.52	SN
Croatian				
T 1	[8,94,41]	146	72.00	SN
T 2	[8,84,29]	121	80.91	SN
T 3	[32,95,52]	180	31.22	SN
T 4	[46,205,66]	319	139.20	SN
T 5	[31,166,52]	249	124.50	SN
Chinese				
T 1	[52,343,225]	620	134.90	SN
T 2	[91,489,470]	1050	82.80	SN
T 3	[33,405,436]	874	66.52	SN
T 4	[70,497,382]	949	154.78	SN
T 5	[48,353,362]	763	57.42	SN
Persian				
T 1	[150,494,135[	779	317.21	SN
T 2	[154,362,110]	627	168.74	SN
T 3	[130,382,70]	583	272.81	SN
T 4	[147,435,115]	697	265.18	SN
T 5	[222,475,145]	841	200.69	SN
German				
T 1	[22,88,46]	158	35.56	SN
T 2	[30,184,114]	328	76.49	SN
T 3	[38,163,66]	267	92.29	SN
T 4	[37,135,55]	227	69.79	SN
T 5	[42,112,61]	215	34.05	SN

Odia				
T 1	[49,270,55]	404	254.13	SN
T 2	[37,132,43]	212	79.85	SN
T 3	[46,179,59]	284	112.68	SN
T 4	[68,165,56]	289	73.42	SN
T 5	[59,192,70]	321	101.29	SN
Russian				
T 1	[15,72,24]	111	49.66	SN
T 2	[18,100,25]	143	86.19	SN
T 3	[9,99,19]	127	113.78	SN
T 4	[21,101,26]	148	81.17	SN
T 5	[37,89,29]	155	40.46	SN
Turkish				
T 1	[62,188,58]	308	106.39	SN
T 2	[84,322,84]	490	231.20	SN
T 3	[188,281,73]	542	83.58	SN
T 4	[125,254,45]	424	134.72	SN
T 5	[159,232,52]	443	72.24	SN

There is only one text which has a neutral structure, all the other texts show a clear preference for nominality, most probably the classical topic-comment or thema-rhema structure.

It is well known that the chi-square used in this form increases linearly with increasing sample size. Different coefficients have been proposed to eliminate this disadvantage but, on the other hand, the latter decrease with increasing sample size. Thus we simply state that almost all journalistic texts in all languages have a nominal character. Perhaps, this is a characteristic feature of journalistic texts, but, again, only a comparison with other text types could help us to solve the problem. The significant chi-square value shows that either there is a scarce specification of nouns or the specification is not symmetric.

It is to be remarked that predicativity may be defined in different ways, cf. e.g. Wildgen (2002, 2005), Löbner (2003) and the interpretation of the numbers will be changed accordingly. In logic, there is a very old and a very extensive domain concerning predication (cf. e.g. Barwise, Etchemendy 2005; Mates 1997; Salmon 1983) but our interest differs from that of logic and we remain at a low, easily attainable level which can be extended in the future.

An indicator analogous to Busemann's coefficient could be defined as the ratio of predicates of the first level, i.e. $A + V$ against the sum $n = A+V+N$. It yields, after testing, the same results as the chi-square test in (15) or (16); hence, it can be omitted here.

3.2. Variability

The next question concerns the variability of texts in a given language. One may expect that each noun has at least one predicate of the first order but texts/ languages may differ from this point of view. Nevertheless, journalistic texts may have, in general, the same structure. In our view this would mean that the vector (A,N,V) has the same structure in all journalistic texts of one language. In order to test this hypothesis, we take the vectors of one language from Table 3.1 and obtain a contingency table, in Brazilian-Portuguese 21×3, in Portuguese 20×3, for other languages 5× 3 = 15. The variability/uniformity can be expressed by a simple chi-square test with 40 DF in the first case, and 38 DF in the second. The critical value at the α = 0.05 level is 55.8 in the first case, and 53.4 in the second one. For languages with 5 texts each, we obtain a chi-square with 8 DF with the critical value X^2 = 15.5. The chi-square test statistic has the form

$$(3.3) \quad X^2 = \sum_{i=1}^{3} \sum_{j=1}^{T} \frac{(n_{ij} - \frac{n_{i.} n_{.j}}{n})^2}{\frac{n_{i.} n_{.j}}{n}}$$

where n_{ij} are the individual numbers in the table, $n_{i.}$ is the sum of the row i, $n_{.j}$ is the sum of the column j, and n is the sum of all frequencies. Inserting the values for Slovak from Table 3.1 we compute X^2 = 11.23 which is not significant. Hence, Slovak journalistic texts seem to have a uniform vector. We do not measure the variability in one text but rather the uniformity of the vector in all texts of a language.

For the individual languages we obtain

Slovak	11.23
Hungarian	17.93
German	20.20
Russian	20.71
Odia	23.22
Croatian	23.94
Persian	25.90
Chinese	51.19
Turkish	71.67

As can be seen, only Slovak expresses a kind of uniformity of journalistic texts. In all the other texts there is a significant variability. Preliminarily, it cannot be said why one obtains the given results. The “cause” may be hidden in style, text type, language, theme, etc. The enormous difference between Croatian and Slovak, both very similar Slavic languages, can be tracked down step by step, analyzing a great number of texts. It must, however be, remarked that the chi-

square test depends strongly on the sample size, hence the greater the numbers in the vector, the greater values can the chi-square test attain.

The difference of the vector structure in two texts may be expressed also in another way. The first is the usual Euclidean distance defined here as

(3.4) $D_{1,2} = [(A_2 - A_1)^2 + (N_2 - N_1)^2 + (V_2 - V_1)^2]^{1/2}$.

where 1 and 2 are two texts. For example, the distance between text 1 and 2 in Brazilian-Portuguese is $D_{1,2} = [(41 - 32)^2 + (165 - 119)^2 + (168 - 61)^2]^{1/2} = 116.82$. In order to make this measure comparable, one should use rather the relative vector elements. In the first text, the sum is $n = 374$, hence [0.1096, 0.4412, 0.4492]; in the second, we have $n = 212$, hence [0.1509, 0.5613, 0.2877]. The distance is now

$D_{1,2,rel} = [(0.1096 - 0.1509)^2 + (0.4412 - 0.5613)^2 + (0.4492 - 0.2877)^2]^{1/2}$
$= 0.2054$

Another measure is the cosine between the vectors expressed in terms of radians according to

(3.5) $$\cos\alpha_{1,2} = \frac{A_1A_2 + N_1N_2 + V_1V_2}{\sqrt{A_1^2 + N_1^2 + V_1^2}\sqrt{A_2^2 + N_2^2 + V_2^2}}.$$

In this case we obtain

$\cos \alpha_{1,2} = (41*32+165*119+168*61)/\{[41^2+165^2+168^2]^{1/2}[32^2+119^2+61^2]^{1/2}\}$
$= 0.6680$

from which the radians are 0.8393.

Unfortunately, in both cases the computation of variances is rather tedious, hence comparison can be made by decision. Nevertheless, comparing all texts with all, a weighted graph may be constructed in which the Euclidean distances or the radians represent the distances between texts.

3.3. Triads

Predicative regularity is given if to each noun in the text there is exactly one adjective and one verb. This is the classical schema corresponding to topic-comment structure. Since the order is irrelevant – there are 6 possibilities – we seek those triads in which there is A, N, V in any order. Consider, for example the sequence (Croatian T 1) with n = 146:

[V,N,N,N,N,V,N,N,N,V,N,V,N,N,N,N,N,N,N,N,V,N,A,N,N,V,N,N,A,V,N,V,A,
N,N,N,N,V,N,N,V,N,N,N,N,N,N,A,N,V,N,V,V,V,N,N,V,N,V,V,V,N,V,N,V,N,

A,N,V,N,N,N,N,V,N,V,N,N,V,V,N,V,V,N,N,V,N,V,V,N,V,N,A,N,N,V,V,N,V,
N,N,N,N,N,N,N,N,V,N,N,N,N,N,N,N,N,N,N,N,N,N,N,N,N,N,N,V,V,N,V,V,A,
N,N,V,N,N,N,A,N,N,V,N,N,N,N]

The regularity can be tested as follows. One counts all not-intersecting triads in the sequence, e.g. in AVNAVN there are two regular triads, but in AVNVAV there is only one because the second one intersects the first. If an entity is part of a triad, it cannot be part of the next triad. This differs from counting e.g. phoneme distributions taking into account all bigrams.

One may define a regularity indicator in the form

$$(3.6) \quad R = \frac{T}{n/3},$$

where T is the number of regular, non-intersecting triads, and n is the length of the sequence. In the above sequence there are 5 regular triads, hence

R = 5/(146/3) = 0.10

The indicator R is again a simple proportion which can be treated binomially. All tests mentioned above can be used for comparisons. Texts will deviate from full regularity and the deviation can be considered a characteristic of text.

Table 3.2
Proportions R of regular triads in texts

Text	**Regular triads T**	**n**	**R**
Slovak			
T1	20	200	0.30
T2	33	241	0.41
T3	28	255	0.33
T4	18	167	0.32
T5	30	252	0.36
Hungarian			
T1	19	147	0.39
T2	21	199	0.32
T3	15	206	0.22
T4	15	173	0.26
T5	28	216	0.39
Croatian			
P 1	5	146	0.10
P 2	6	121	0.15

P 3	19	180	0.32
P 4	29	319	0.28
P 5	14	249	0.17
Chinese			
T 1	34	620	0.16
T 2	76	1050	0.22
T 3	30	874	0.19
T 4	51	949	0.16
T 5	29	763	0.11
Persian			
T 1	74	779	0.28
T 2	73	627	0.35
T 3	45	583	0.23
T 4	62	697	0.27
T 5	105	841	0.37
German			
T 1	16	158	0.30
T 2	24	328	0.22
T 3	22	267	0.25
T 4	23	227	0.30
T 5	32	215	0.45
Odia			
T 1	22	404	0.16
T 2	23	212	0.33
T 3	25	284	0.26
T 4	36	289	0.37
T 5	32	321	0.30
Russian			
T 1	8	111	0.21
T 2	11	143	0.23
T 3	2	127	0.05
T 4	22	148	0.45
T 5	21	155	0.41
Turkish			
T 1	25	308	0.24
T 2	28	490	0.17
T 3	60	542	0.33
T 4	31	424	0.22
T 5	35	443	0.24

The regularity of journalistic texts can be characterized either by adding the *n*'s in one language and dividing the sum by 3, then multiplying by the sum of triads, or

simply as the mean of *R*. For the languages analyzed in the above Table 3.2 one obtains (using the rounded *R*-values) the means

Slovak	0.34
Hungarian	0.32
German	0.30
Persian	0.30
Odia	0.28
Russian	0.27
Turkish	0.24
Croatian	0.20
Chinese	0.17

The most regular occurrence of predicative triads can be observed in Slovak, the most irregular are the Chinese texts. It can easily be seen that regularity does not depend either of familiar or areal situation of the languages, it is a property of the text type or of the style.

Since *R* is a simple proportion, one can compare languages in the same way as it has been done using mean *Q*.

One can restrict this view considering only one of the permutations of ANV as normative, e.g. the basis of short declarative sentences. In this case, the number of "regular" triads decreases still more and in short texts we may find no regular triad (cf. the above example). Thus, a more thorough investigation is possible only with longer texts, or considering only one of the permutations of ANV as basic.

3.4. Runs of three elements

If the text would be quite regular — here perfectly predicative (at the first level) —, we would obtain a regular sequence of triads. But no text has this property. There are always runs of different lengths which give the text a special character. The smaller is the number of runs, the greater is the grammatical or semantic monotony. The authors may have special aims (both conscious and unconscious) which can be achieved by construction of runs. Since we observe here only three different parts of speech, we shall obtain a restricted image of predicativity represented by runs.

Let *A*, *N*, *V* be the numbers of adjectives, nouns and verbs, and let r_A, r_N, r_V be the number of runs of these elements. For the German T 1 we obtain r_A = 21, r_N = 47, and r_V = 38, that is, we have r = 106 runs. In order to state whether this signals some tendency, we compute the following quantities (cf. Bortz, Lienert, Boehnke 1990):

(3.7) $m = n - r$

(3.8) $$F_2 = \sum_{j=1}^{3} n_j (n_j - 1)$$

(3.9) $$F_3 = \sum_{j=1}^{3} n_j (n_j - 1)(n_j - 2)$$

(3.10) $$E(m) = \frac{F_2}{n}$$

(3.11) $$\sigma_m = \sqrt{\frac{(n-3)F_2}{n(n-1)} + \frac{F_2^2}{n^2(n-1)} - \frac{2F_3}{n(n-1)}}$$

Hence

(3.12) $$z = \frac{m - E(m)}{\sigma_m}.$$

Computing these values for the above mentioned text we obtain

A = 22, N = 88, V = 46
n = 158
r = 106
m = 158 – 106 = 52
F_2 = 22(21) + 88(87) + 46(45) = 10188
F_3 = 22(21)20 + 88(87)86 + 46(45)44 = 758736
E(m) = 10188/158 = 64.4810

$$\sigma_m = \sqrt{\frac{(158-3)10188}{158(157)} + \frac{10188^2}{158^2(157)} - \frac{2(758736)}{158(157)}} = 5.3823.$$

Inserting these numbers in (3.12) we obtain

$$z = \frac{52 - 64.4810}{5.3823} = -2.32$$

This value is significant, hence we may say that in the given text there exists a significant tendency for setting up runs. If one obtains a positive significant result (z > 1.96), there are few runs; if one obtains a negative significant result (<−1.96), there are too many runs. The interpretation is, so to say, opposite because we considered *n* - *r*. Too many runs are signs of greater regularity than too few runs.

For the other texts whose sequences are known to us we obtain the results presented in Table 3.3.

Table 3.3
Test for runs

Text	Vector	n	r	m	F_2	F_3	E(m)	σ_m	z
Slovak									
T 1	[35,126,39]	200	126	74	18422	2007904	92.11	5.7375	-3.16*
T 2	[44,124,73]	241	177	64	22400	2234008	92.95	7.1358	-4.06*
T 3	[41,148,66}	255	165	87	26680	3248002	105.87	6.8547	-2.75*
T 4	[29,104,34}	167	112	55	12646	1128586	75.72	5.2853	-3.92*
T 5	[47,150,55]	252	163	89	27482	3465304	109.06	6.7822	-2.96*
Hungarian									
T1	[27,85,35]	147	106	41	9032	631944	61.44	5.2505	-3.89*
T2	[59,111,29]	199	130	69	16444	1352932	82.63	6.9001	-1.98*
T3	[37,121,48]	206	134	72	18108	1831730	87.90	6.1634	-2.58*
T4	[41,102,29]	173	119	54	12754	1052206	73.72	4.8088	-3.40*
T5	[63,109,43]	216	162	54	17570	1339194	81.34	7.3271	-3.72*
Croatian									
T 1	[8,94,41]	146	75	71	11008	948616	75.40	4.8986	-0.89
T 2	[8,84,29]	121	56	65	7840	593644	64.79	7.1145	0.05
T 3	[32,95,52]	180	119	61	12574	963154	69.86	6.0457	-1.48
T 4	[46,205,66]	319	199	120	48180	8764112	151.03	7.0020	-4.43*
T 5	[31,166,52]	249	123	125	30972	4624622	124.39	5.9990	0.10
Chinese									
T 1	[52,343,225]	620	339	281	170368	51240650	274.77	11.3499	0.55
T 2	[91,489,470]	1050	664	386	467252	219375206	445.00	15.3164	-3.85*
T 3	[33,405,436]	874	465	449	354336	148251366	405.42	14.2886	0.25
T 4	[70,497,382]	949	543	406	396884	177328540	418.21	14.4096	-0.85
T 5	[48,353,362]	763	447	316	257194	90659472	337.08	13.1700	-1.60
Persian									
T 1	[150,494,135]	779	443	336	283982	122228934	364.55	11.4486	-2.49*
T 2	[154,362,110]	627	400	227	166234	48340748	265.13	11.4125	-3.34*
T 3	[130,382,70]	583	316	267	167142	55634660	286.69	9.9499	-1.98*
T 4	[147,435,115]	697	411	286	223362	83227794	320.46	11.1339	-3.10*
T 5	[222,475,145]	841	531	310	294844	108809628	349.76	13.6565	-2.91*
German									
T 1	[22,88,46]	158	106	52	10278	749544	65.05	5.5446	-2.32*
T 2	[30,184,114]	328	213	115	47424	7571148	114.59	8.1518	-3.63*
T 3	[38,163,66]	267	159	108	32102	4526002	120.23	6,7985	-1.80
T 4	[37,135,55]	227	149	87	22392	2563454	98.64	6.3945	-1.82
T 5	[42,112,61]	215	152	63	17814	1583544	82.85	6.73	-2.95*
Odia									
T 1	[49,270,55]	374	186	188	77952	19622348	208.43	6.5159	-3.14*
T 2	[37,132,43]	212	131	81	20430	2322080	96.37	5.9704	-2.57*
T 3	[46,179,59]	284	178	106	37354	5834720	131.53	6.8217	-3.74*
T 4	[68,165,56]	289	201	88	34696	4577236	120.06	7.6994	-4.16*
T 5	[59,192,70]	321	194	127	44924	7296238	139.95	7.6303	-1.70
Russian									
T 1	[15,72,24]	111	68	43	5874	370014	52.9189	4.0996	-2.42*

T 2	[18,100,25]	143	80	63	10806	984036	75.5664	4.2184	-2.98*
T 3	[9,99,19]	127	53	72	10116	946936	79.6535	3.2238	-1.44
T 4	[21,101,26]	148	60	88	11170	1015542	75.4729	4.4540	-1.68
T 5	[37.89,39]	155	111	44	9976	703383	64.3613	5.6116	-3.63*
Turkish									
T 1	[62,188,58]	308	192	116	42244	6724276	137.1558	7.4371	-2.84*
T 2	[84,322,84]	490	299	191	117306	33647712	239.4000	8.6470	-5.60*
T 3	[188,281,73]	542	381	161	119092	22325272	219.7269	12.4852	-4.70*
T 4	[125,254,45]	424	274	150	81742	16279414	192.7877	9.9099	-4.32*
T 5	[159,232,52]	443	299	144	81366	12459078	183.6704	11.4849	-3.45*

Slovak, Hungarian and Persian have significantly too many runs; in the other languages the situation is rather mixed. Croatian and Chinese have only one significant result each, German and Russian have three. Positive *z*-values are extremely rare. The cause of this phenomenon is not found. It may be evoked by boundary conditions, by text type, by the situation in the given language, etc.

4. Distances

In previous chapters (see sections 2.4 and 3.6) we have made use of the concept of runs in order to detect possible regularities of the sequential text organization. We tested whether the runs (their number or length) occurring in a text under study are comparable with the runs that would be observed in a random sequence.

Another approach is to study the distances between identical elements of a text and test whether the observed numbers of distances coincide basically with the distance structure of a random sequence or if there are significant deviations from the random case. It turns out that the theory of distances, some results of which are briefly outlined as follows, can be regarded as a generalization of the theory of runs. The complete theory of distances for randomly constructed sequences can be found in Zörnig (1984, 1987, 2010). Probabilistic models to describe the distance frequencies in some linguistic applications have been presented in Zörnig (2013a, b).

Let $k_1,...,k_p$ be natural numbers such that $k_1 \geq \ldots \geq k_p$ and $k_1 + \ldots + k_p = n$. The set of sequences of length n, consisting of elements from the set $\{1,\ldots,p\}$ such that the element r occurs exactly k_r times ($r = 1,\ldots,p$) is denoted by $IF(k_1,...,k_p)$. For example, IF(5,3,2) is the set of all sequences, containing 5 times the element 1, 3 times the element 2 and 2 times the element 3, i.e. this set consists of all possible permutations (rearrangements) of the sequence (1,1,1,1,1,2,2,2,3,3), e.g. (1,2,3,1,2,3,1,1,1,2) or (1,2,1,3,3,2,1,2,1,1). The number of these permutations (with repetitions) is given by the multinomial coefficient

$$\binom{10}{5,3,2} = \frac{10!}{5! \cdot 3! \cdot 2!} = 2520.$$

Any sequence of $IF(k_1,...,k_p)$ can be interpreted as an abstract text, where the elements $1,\ldots,p$ may represent word forms, lemmas, letters or other text units. A random text in the sense of the present chapter is a random selection from $IF(k_1,...,k_p)$. Consider now a sequence $F = (a_1,...,a_n) \in IF(k_1,\ldots,k_p)$. For integers μ and ν with $1 \leq \mu < \nu \leq n$, the *distance* between the elements a_μ and a_ν is defined as $c(\mu,\nu) = \nu - \mu - 1$. The distance between two elements of a sequence is thus the number of elements between them, e.g. the distance between a_3 and a_8 of the sequence $(a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8)$ is 4, since there are the four elements a_4, a_5, a_6, a_7 between a_3 and a_8. This concept of distance depends only on the positions of the elements in the sequence and not on their values. Alternatively we could have defined the distance as the number of steps required for moving from one element to the other. In this case the distance between the elements a_3 and a_8 in the last example would be 5. However we will restrict ourselves to the definition above, already introduced in Zörnig (1984).

The number of occurrences of the distance *i* between two elements of the *r*-th kind in the sequence *F* is denoted by $d_i^{(r)}(F)$. Formally, $d_i^{(r)}(F)$ is the number of position pairs (μ,ν) such that $a_\mu = a_\nu = r$ and $c(\mu,\nu) = i$ $(1 \le \mu < \nu \le n$, r = 1,…,*p*, i = 0,…,n-2).. Similarly, let $c_i^{(r)}(F)$ denote the number of occurrences of the distance *i* between *consecutive* elements of the r^{th} kind. Formally this quantity counts the number of position pairs (μ,ν) such that $a_\mu = a_\nu = r$, $a_\rho \neq r$ for all ρ with $\mu < \rho < \nu$ and $c(\mu,\nu) = i$. Finally, $d_i(F) = \sum_{r=1}^{p} d_i^{(r)}(F)$ and $c_i(F) = \sum_{r=1}^{p} c_i^{(r)}(F)$ are called the *total* and the *consecutive frequency of the distance* i.

Example: Consider again the sequence

F = (A,A,A,A,V,V,A,A,A,V,V,A,A,A.V,V,V,V,A,V,A,V,V,A,A),

By identifying the element *V* with 1 and *A* with 2 we get $k_1 = 11$, $k_2 = 14$, p = 2, n = 25. Calculating the distances, we get, for example, $d_5 = 12$, since the distance 5 occurs between the elements of the 12 position pairs (1,7), (2,8), (3,9), (5,11), (7,13), (8,14), (10,16), (11,17), (13,19), (16,22), (17,23), (19,25) of the sequence e. g. the distance between the identical elements $a_1 = A$ and $a_7 = A$ is 5. None of these pairs corresponds to a distance between two *consecutive* identical elements. For example, between the elements a_1 and a_7 are located three other *A´*s at positions 2, 3 and 4; thus $c_5 = 0$. Similarly we get $d_3 = 9$, where the distance 3 occurs between the elements of the pairs (3,7), (4,8), (6,10), (8,12), (9,13), (11,15), (16,20), (18,22), (21, 25). Only the pairs (6,10) and (11,15) represent consecutive distances. A look at the above example of a sequence shows that $a_6 = a_{10} = V$ and there is no other *V* between a_6 and a_{10} ; in addition, $a_{11} = a_{15} = V$ and there is no other *V* between the elements a_{11} and a_{15}. The complete list of distance frequencies is given in Table 4.1.

Table 4.1
Distance structure

i	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
d_i	14	9	7	9	15	12	9	5	5	8	9	10	5	2	3	4
c_i	14	3	3	2	1	0	0	0	0	0	0	0	0	0	0	0

i	16	17	18	19	20	21	22	23	Sum
d_i	6	4	1	2	2	2	2	1	146
c_i	0	0	0	0	0	0	0	0	23

One can easily verify that in general it holds $d_0 = c_0$, and the relations

$$(4.1) \qquad \sum_{i=0}^{n-2} d_i = \frac{m-n}{2} \text{ (where } m = \sum_{r=1}^{p} k_r{}^2 \text{) and } \sum_{i=0}^{n-2} c_i = n - p$$

are satisfied. Moreover, the number of runs can be expressed in terms of 0−distances, i.e. this number is given by $n-d_0$. In this sense the theory of distances generalizes the theory of runs. In the example it holds $n-d_0 = 25-14 = 11$, which is the number of runs of the sequence.

We now cite two statements (Theorems 2.1 and 2.2 in Zörnig 2010) related to the distance structure in random texts. We assume that the construction of a random text is a selection of a sequence from the set of sequences $IF(k_1,...,k_p)$. The distance frequencies d_i and c_i are therefore random variables.

Theorem 4.1: Let the sequence F be randomly chosen from $IF(k_1,...,k_p)$. For any $i \in \{0,1,\ldots,n-2\}$ the expectation and variance of the frequency d_i are given by

(a) $E(d_i) = \dfrac{(m-n)(n-1-i)}{n(n-1)}$, $\quad (m = \sum_{r=1}^{p} k_r{}^2)$,

(b) $V(d_i) = \dfrac{n-1-i}{n(n-1)} A + 2(n-2-2i) + \dfrac{(n-3)B - A^2 - 2C}{n_{(4)}}$

$$+ (n-i-1)(n-i-2)\,\frac{A^2+2C}{n_{(4)}} - \left((n-i-1)\,\frac{A}{n(n-1)}\right)^2,$$

where $A = \sum_{r=1}^{p} k_r(k_r-1)$, $B = \sum_{r=1}^{p} k_r(k_r-1)(k_r-2)$,

$C = \sum_{r=1}^{p} k_r(k_r-1)(3-2k_r)$; $n_{(i)} = n(n-1)\ldots(n-i+1)$ denotes the decreasing factorial and T^+ is the positive part of the term T, i.e. $T^+ = T$ if $T \geq 0$ and $T^+ = 0$ if $T < 0$.

The theorem states that $E(d_i)$ and $V(d_i)$ is a linear or a quadratic function of i, respectively. The quadratic term in the variance $V(d_i)$ is usually very small.

We illustrate the calculations for the above example, where we obtain $n = 25$, $m = 14^2 + 11^2 = 317$, hence $E(d_i) = \dfrac{292}{600}(24-i)$. Moreover,

$A = 14 \cdot 13 + 11 \cdot 10 = 292,$
$B = 14 \cdot 13 \cdot 12 + 11 \cdot 10 \cdot 9 = 3174,$
$C = 14 \cdot 13 \cdot (-25) + 11 \cdot 10 \cdot (-19) = -6640.$

Setting e.g. i = 3 yields

$$V(d_3) = \frac{21}{25 \cdot 24} \cdot 292 + 2 \cdot 17 \; \frac{22 \cdot 3174 - 292^2 - 2 \cdot (-6640)}{25 \cdot 24 \cdot 23 \cdot 22}$$
$$+ 21 \cdot 20 \; \frac{292^2 + 2 \cdot (-6640)}{25 \cdot 24 \cdot 23 \cdot 22} - \left(21 \cdot \frac{292}{25 \cdot 24} \right)^2 \approx 5.113$$

Theorem 4.2: Let the sequence F be randomly chosen from $IF(k_1,...,k_p)$. For any $i \in \{0,1,\ldots,n-2\}$ the expectation of the consecutive frequency c_i is given by

$$E(c_i) = \frac{1}{n_{(i+1)}} \sum_{r=1}^{p} k_r (k_r - 1)(n - k_r)_{(i)} .$$

There also exists a complicated formula for the variance of c_i (see Zörnig 1987, p.15) which will not be introduced here.

Illustrating the calculations again for the above example, we obtain

$$E(c_i) = \frac{1}{25_{(i+1)}} [14 \cdot 13 \cdot 11_{(i)} + 11 \cdot 10 \cdot 14_{(i)}].$$

Since $n_{(i)} = \frac{n!}{(n-i)!}$ for $i \le n$ and $n_{(i)} = 0$ for $i > n$, we get

$$E(c_i) = \frac{(24-i)!}{25!} [182 \cdot \frac{11!}{(11-i)!} + 110 \cdot \frac{14!}{(14-i)!} \quad \text{for } i \le 11.$$

For example, $E(c_6) = \frac{18!}{25!} [182 \cdot \frac{11!}{5!} + 110 \cdot \frac{14!}{8!}] = 0.1232.$

For a large text length n and small values of *i* we can express $E(c_i)$ in Theorem 4.2 approximately as

$$(4.2) \qquad E(c_i) \approx \sum_{r=1}^{p} f_r (k_r - 1)(1 - f_r)^i \quad \text{for } i = 0,1,\ldots,n-2$$

where $f_k = k_r/n$ are the relative frequencies. The probability functions of the random variables d_i and c_i are not known and due to the complexity one cannot

expect to find them. But by means of simulation methods (Zörnig 2010) it has been shown that they can be considered approximately normally distributed.

We have calculated all values of $E(d_i)$, $V(d_i)$ and $E(c_i)$ for the above example. The results are summarized in Table 4.2, where observed and expected distance frequencies are compared.

Table 4.2
Observed and expected distances

i	d_i observed	$E(d_i)$	$V(d_i)$	c_i observed	$E(c_i)$
0	14	11.68	5.811	14	11.68
1	9	11.19	5.578	3	5.903
2	7	10.71	5.345	3	2.901
3	9	10.22	5.113	2	1.385
4	15	9.733	4.881	1	0.6406
5	12	9.247	4.650	0	0.2864
6	9	8.760	4.419	0	0.1232
7	5	8.273	4.189	0	0.0506
8	5	7.787	3.959	0	0.0196
9	8	7.300	3.730	0	0.0070
10	9	6.813	3.502	0	0.0023
11	10	6.327	3.274	0	0.0006
12	5	5.840	3.032	0	0.0001
13	2	5.353	2.776	0	0
14	3	4.867	2.521	0	0
15	4	4.380	2.267	0	0
16	6	3.893	2.013	0	0
17	4	3.407	1.760	0	0
18	1	2.920	1.507	0	0
19	2	2.433	1.254	0	0
20	2	1.947	1.002	0	0
21	2	1.460	0.751	0	0
22	2	0.973	0.500	0	0
23	1	0.487	0.250	0	0

The data in Table 4.2 are visualized in Figures 4.1 and 4.2 in order to compare the observed values of the distance frequencies with the expected ones.

The solid line illustrates the observed values and the straight line in the middle represents the expected values $E(d_i)$. The outer curves, given by $E(d_i) \pm 1.96\sigma$, where $\sigma = \sqrt{V(d_i)}$ is the standard deviation, represent the limits of a "confidence band" for the level of significance α = 5%. This means that the

value d_i lies outside of these limits with probability 0.05 (under the normality assumption and for a fixed index i). A visual inspection of Fig. 4.1 shows that the largest deviations from the expectation occur for d_4, d_{11}, and d_{13} which are the only observations outside of the confidence band. One could interpret this as a first indication for the fact that the distances 4 and 11 between verbs or adjectives are preferred while the distance 13 is suppressed. Of course, these argumentations have only preliminary illustrative character. The considered small example allows in no way definite conclusions.

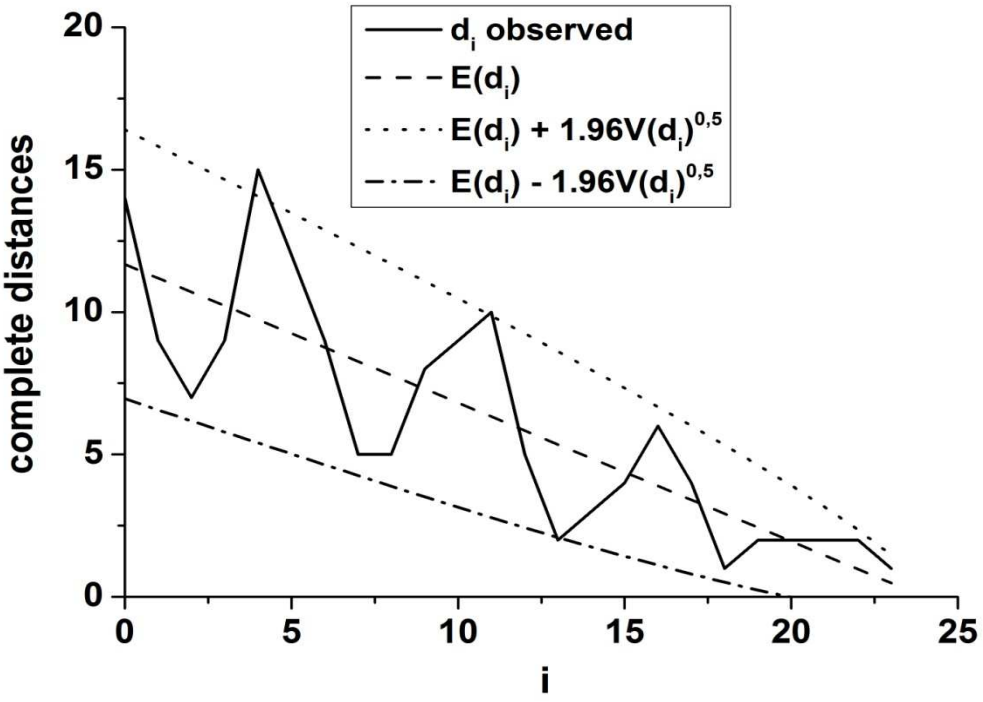

Figure 4.1. Observed and expected values for the complete distances d_i

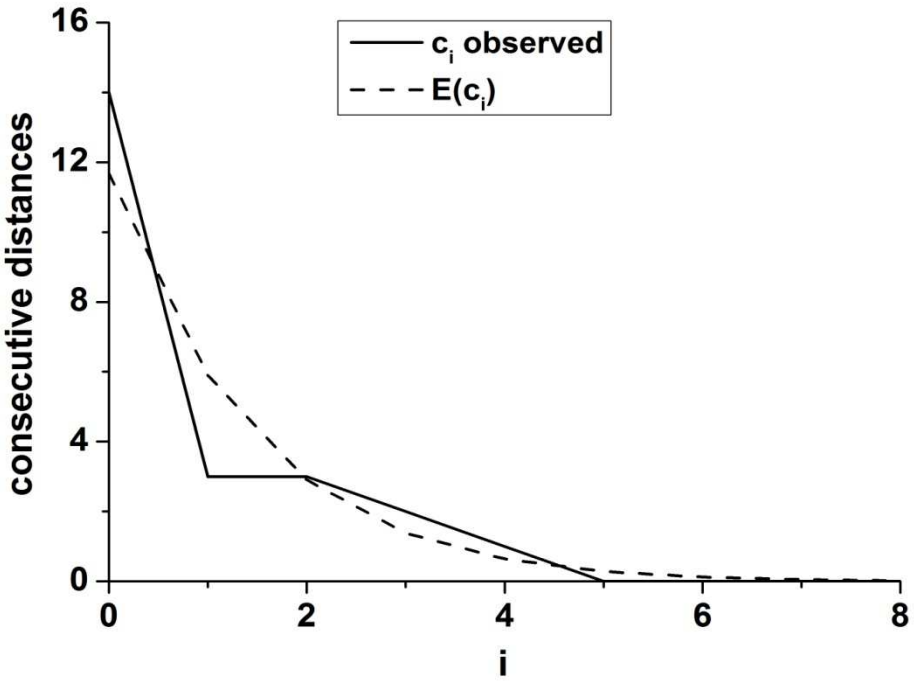

Figure 4.2. Observed and expected values for the consecutive distances c_i

Similar to the previous figure the solid line in Fig. 4.2 represents the observed values c_i and the dashed line the expected frequencies. The observations for c_0 and c_1 deviate slightly from the expectations, but there is not enough evidence for drawing conclusions. In order to determine whether a deviation is significant, simulation seems to be the best approach, since the formulas for the variances are complicated.

To avoid any possible misunderstandings, we point out that in the above considerations we considered a frequency d_i or c_i for an arbitrary but *single* index i. If we consider various values of these frequencies simultaneously (e.g. d_3, d_5 and d_{11}) we must be aware that there are complex dependencies between them in addition to the relations (4.1). Consider e.g. the set of sequences IF(2, 2) = {(1,1,2,2), (1,2,1,2), (1,2,2,1), (2,1,1,2), (2,1,2,1), (2,2,1,1)}. The occurrence of the distance 2 implies that the sequence is (1,2,2,1) or (2,1,1,2), where the distance 0 occurs. Hence, $d_2 > 0$ implies $d_0 > 0$. Studying multiple distances simultaneously there is no joint probability mass function for the random vectors $(d_0,d_1,\ldots,d_{n-2})$ or $(c_0,c_1,\ldots,c_{n-2})$ known and there is no hope to find an analytic representation for it. Possible measures for the deviation between observed and expected distances could be, for example, the chi-square statistic

$$\chi^2 = \sum_{i=0}^{n} \frac{(d_i - E(d_i))^2}{E(d_i)}$$

or the sum of squared deviations $\sum_{i=0}^{n}(d_i - E(d_i))^2$ (analogous expressions can be defined for the consecutive distances). Other measures for the goodness-of-fit can be found in Mačutek, Wimmer (2013, p.284). Due to the dependence between the random variables d_i or c_i the exact distributions of the mentioned statistics are not known and the only way to obtain exact probabilities seems to be simulation. However, in the following studies we ignore the dependence in order to obtain a practicable way and use the chi-square statistic as in the case of independent observations.

In what follows we analyze the distances in the 45 sequences of the appendix, where the elements *A*, *N* and *V* represent adjectives, nouns and verbs observed in the texts of several languages. We restrict ourselves to consecutive distances between identical elements. The results are presented in Table 4.3. For each i = 0,1,…,10 we determine the observed frequencies c_i and the expected frequency according to the random model in Theorem 4.2. In fact we have used the simpler approximate formula (4.2) whose values differ only slightly from the exact ones. In fact we have r = 3, since the considered sequences contain only the three different elements *A*, *N* and *V*, and k_1, k_2 and k_3 denote the frequencies of these elements. Here k_i determine the theoretic frequencies completely, i.e. there are no model parameters that must be "adjusted" to the observed data. As a measure for the goodness-of-fit, we applied the above mentioned chi-square statistic. For most of the 45 sequences all theoretic values $E(c_i)$ were larger than 1 for i = 0,1,…,10. In these cases we consider 12 classes: distance = 0,…,distance = 10, distance > 10. The statistic is then

$$\chi^2 = \sum_{i=0}^{11} \frac{(c_i - E(c_i))^2}{E(c_i)},$$

where c_{11} has been *renamed* as the number of distances larger than 10, i.e. $c_{11} = \sum_{i>10} c_i$, $E(c_{11}) = \sum_{i>10} E(c_i)$. The number of degrees of freedom is then DF=11 (number of classes minus 1). In the few cases in which $E(c_i) < 1$ occurred for an i close to 10 we pooled the classes of distances larger than 10 with the classes satisfying $E(c_i) < 1$ (see e.g. the sequence Croatian T 1 in part (c) of Table 4.3). For each text Table 4.3 presents the values c_i and $E(c_i)$ in the upper part, the lower part contains the number of adjectives, nouns and verbs in the sequence, denoted by *A*, *N* and *V*, the sequence length n = *A*+*N*+*V*, the number DF of degrees of freedom, the observed χ^2-value and $P = P(\chi^2 > \text{obs.})$ which denotes the probability to exceed the observed χ^2.

The fit of the observed values by means of the considered model can be considered as good if $P > 0.05$ (see e.g. Zörnig 2013a, p.120). According to this

criterion, 25 of the 45 sequences in Table 4.3 can be well fitted by the considered random model.

Table 4.3
Observed consecutive distances compared with expectations under the random hypothesis

	Slovak 1		**Slovak 2**		**Slovak 3**		**Slovak 4**		**Slovak 5**	
i	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**
0	74	92.1	64	92.9	87	105.0	55	75.7	88	107.9
1	52	40.0	70	52.3	70	52.9	50	33.6	64	51.7
2	19	19.6	46	30.8	33	28.6	21	16.7	39	27.2
3	11	11.2	19	18.9	17	16.9	7	9.6	7	16.0
4	13	7.3	13	12.2	10	10.9	8	6.3	14	10.5
5	7	5.3	6	8.2	13	7.5	4	4.5	12	7.5
6	5	4.1	5	5.7	5	5.5	3	3.4	6	5.5
7	4	3.2	7	4.1	3	4.1	7	2.7	5	4.4
8	3	2.6	1	3.0	2	3.2	4	2.2	1	3.3
9	3	2.1	2	2.2	1	2.5	2	1.7	3	2.6
10	3	1.8	1	1.7	2	2.0	0	1.4	4	2.1
>10	3	7.6	4	6.1	6	9.1	3	6.1	5	9 .4
	A = 35 N = 126 V = 39 n = 200 DF = 11 χ^2 = 16.5959 P = 0.1204		A = 44 N = 124 V = 73 n = 241 DF = 11 χ^2 = 27.6829 P = 0.0036		A = 41 N = 145 V = 65 n = 251 DF = 11 χ^2 = 16.3439 P = 0.1288		A = 29 N = 104 V = 34 n = 167 DF = 11 χ^2 = 27.4702 P = 0.0039		A = 47 N = 149 V = 54 n = 250 DF = 11 χ^2 = 26.4174 P = 0.0056	

	Hungarian 1		**Hungarian 2**		**Hungarian 3**		**Hungarian 4**		**Hungarian 5**	
i	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**
0	41	61.4	69	82.6	72	87.9	55	75.6	54	81.3
1	48	30.6	53	42.7	64	42.8	53	36.1	63	46.8
2	17	16.5	24	23.5	24	22.8	21	19.0	39	28.0
3	10	9.8	15	13.8	11	13.5	10	11.1	24	17.5
4	11	6.4	12	8.7	3	8.8	5	7.2	10	11.4
5	8	4.5	6	5.9	7	6.2	12	5.1	7	7.7
6	2	3.3	6	4.1	8	4.5	3	3.8	4	5.3
7	3	2.5	2	3.0	3	3.5	5	2.9	6	3.8
8	0	1.9	2	2.3	4	2.7	3	2.3	1	2.8
9	0	1.5	3	1.8	1	2.1	1	1.8	1	2.0
10	1	1.2	2	1.4	1	1.7	0	1.4	1	1.5

>10	3	4.5	2	6.2	5	6.6	4	5.8	3	5.0
	A = 27 N = 85 V = 35 n = 147 DF = 11 χ^2 = 27.4617 P = 0.0039		A = 59 N = 111 V = 29 n = 199 DF = 11 χ^2 = 11.2846 P = 0.4197		A = 37 N = 121 V = 48 n = 206 DF=11 χ^2 =22.3857 P = 0.0215		A = 41 N = 104 V = 30 n = 175 DF = 11 χ^2 = 28.1572 P = 0.0031		A = 63 N = 109 V = 44 n = 216 DF = 11 χ^2 = 25.9977 P = 0.0065	

	Croatian 1		**Croatian 2**		**Croatian 3**		**Croatian 4**		**Croatian 5**	
i	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**
0	71	75.4	65	64.8	61	69.9	124	153.6	125	124.4
1	37	29.8	24	23.2	45	38.4	86	63.2	45	48.4
2	11	13.3	6	9.7	29	22.2	33	30.0	25	21.8
3	8	6.9	5	5.0	9	13.6	20	16.6	12	11.9
4	4	4.1	4	3.1	7	8.8	9	10.8	8	7.7
5	2	2.7	3	2.2	6	5.9	10	7.9	9	5.7
6	2	1.9	4	1.7	5	4.2	8	6.1	3	4.4
7	0	1.4	1	1.3	6	3.0	4	4.9	2	3.6
8	3	1.1	1	1.0	3	2.2	7	4.9	1	2.9
9	0	0.8	1	0.8	2	1.7	4	3.3	6	2.4
10	0	0.6	2	0.7	0	1.3	3	2.7	2	2.0
>10	5	4.8	2	4.7	4	5.8	8	13.2	8	10.8
	A = 8 N = 97 V = 41 n = 146 DF = 9 χ^2 =7.9608 P = 0.5381		A = 8 N = 84 V = 29 n = 121 DF = 9 χ^2 = 5.6199 P = 0.7773		A = 32 N = 95 V = 52 n = 179 DF = 11 χ^2 = 11.5067 P = 0.4018		A = 46 N = 207 V = 66 n = 319 DF=11 χ^2 =21.1322 P = 0.0320		A = 31 N=166 V = 52 n = 249 DF = 11 χ^2 = 11.1298 P = 0.4325	

	Chinese 1		**Chinese 2**		**Chinese 3**		**Chinese 4**		**Chinese 5**	
i	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**
0	281	274.8	386	445.0	409	405.4	406	418.2	315	337.6
1	140	140.2	293	244.5	211	210.4	225	220.1	218	180.5
2	78	74.4	149	125.4	111	110.0	136	118.0	94	97.1
3	37	41.2	76	76.0	55	57.3	71	64.8	53	52.6
4	27	23.9	49	43.5	28	30.2	29	36.7	26	29.0
5	10	14.7	20	25.6	17	16.2	17	21.5	10	16.3

6	6	9.5	16	15.8	6	8.9	14	13.2	10	9.5
7	4	6.5	8	10.3	8	5.0	3	8.6	9	5.9
8	4	4.6	7	7.1	2	3.0	7	5.9	2	3.9
9	8	3.5	3	5.3	3	2.0	5	4.4	3	2.8
10	3	2.7	4	4.1	3	1.4	3	3.4	0	2.1
>10	19	21.1	36	34.4	18	21.6	30	31.2	4	6.1
	A = 52 N = 343 V = 225 n = 620 DF = 11 χ^2 = 10.9885 P = 0.4442		A = 91 N = 489 V = 470 n = 1050 DF = 11 χ^2 = 22.2895 P = 0.0222		A = 33 N = 405 V = 436 n = 874 DF = 11 χ^2 = 6.2913 P = 0.8532		A = 70 N = 497 V = 382 n = 949 DF = 11 χ^2 = 10.3927 P = 0.4954		A = 48 N = 354 V = 362 n = 764 DF = 11 χ^2 = 17.1861 P = 0.1025	

	Persian 1		**Persian 2**		**Persian 3**		**Persian 4**		**Persian 5**	
i	c_i	$E(c_i)$	c_i	$E(c_i)$	c_i	$E(c_i)$	c_i	$E(c_i)$	c_i	$E(c_i)$
0	336	364.5	229	266.2	267	288.0	286	320.5	310	349.8
1	177	156.7	152	132.6	131	116.1	175	141.8	187	179.8
2	80	76.4	91	71.8	55	53.6	73	70.0	137	88.4
3	42	43.5	40	42.7	34	29.5	39	40.5	52	59.5
4	38	28.6	31	27.7	16	19.0	23	26.5	43	38.4
5	19	20.9	23	19.3	19	13.8	23	19.1	28	26.5
6	15	16.1	17	14.2	9	10.6	18	14.6	23	19.1
7	12	12.8	11	10.7	12	8.5	14	11.5	12	14.2
8	13	10.4	9	8.3	10	6.8	6	9.2	16	10.8
9	14	8.4	4	6.4	5	5.6	3	7.4	10	8.4
10	8	6.9	3	5.1	4	4.6	5	6.0	3	6.5
>10	22	30.7	15	19.9	19	24.9	29	26.5	17	25.6
	A = 150 N = 494 V = 135 n = 779 DF = 11 χ^2 = 15.4759 P = 0.1617		A = 154 N = 363 V = 111 n = 628 DF = 11 χ^2 = 18.0268 P = 0.0810		A = 131 N = 383 V = 70 n = 584 DF = 11 χ^2 = 11.3612 P = 0.4135		A = 147 N = 435 V = 115 n = 697 DF = 11 χ^2 = 18.3572 P = 0.0737		A = 222 N = 474 V = 145 n = 841 DF = 11 χ^2 = 29.3477 P = 0.0020	

	German 1		**German 2**		**German 3**		**German 4**		**German 5**	
i	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**
0	52	65.1	115	144.6	108	120.2	87	98.6	63	82.9
1	49	33.7	102	73.1	63	55.1	53	47.1	48	46.3
2	20	18.6	44	38.7	30	28.0	29	24.7	41	27.2
3	11	11.0	19	21.6	22	16.0	12	14.4	20	16.8

4	6	7.0	13	12.7	9	10.3	11	9.3	20	10.9
5	3	4.7	11	8.0	7	7.2	8	6.5	7	7.4
6	3	3.3	4	5.3	8	5.4	7	4.8	6	5.2
7	2	2.4	2	3.7	0	4.1	5	3.7	2	3.7
8	3	1.8	3	2.6	6	3.3	2	2.9	2	2.7
9	0	1.4	2	2.0	2	2.6	3	2.3	1	2.1
10	2	1.1	0	1.6	2	2.1	3	1.8	1	1.6
>10	4	4.8	10	11.1	7	9.7	4	7.6	1	5.3
	A = 24 N = 88 V = 46 n = 158 DF = 11 χ^2 = 13.5141 P = 0.2611		A = 30 N = 184 V = 114 n = 328 DF = 11 χ^2 = 22.4829 P = 0.0209		A = 38 N = 163 V = 66 n = 267 DF = 11 χ^2 = 13.5491 P = 0.2590		A = 37 N = 135 V = 55 n = 227 DF = 11 χ^2 = 8.2941 P = 0.6867		A = 42 N = 112 V = 61 n = 215 DF = 11 χ^2 = 25.4803 P = 0.0077	

	Odia 1		**Odia 2**		**Odia 3**		**Odia 4**		**Odia 5**	
i	**c_i**	**E(c_i)**	**c_i**	**E(c_i)**	**c_i**	**E(c_i)**	**c_i**	**E(c_i)**	**c_i**	**E(c_i)**
0	188	208.4	81	96.4	196	131.5	88	120.1	127	140.0
1	83	66.2	41	42.8	86	57.1	73	60.8	72	66.4
2	27	25.5	37	21.3	31	28.1	50	33.4	42	34.8
3	11	13.2	14	12.2	13	16.0	27	20.0	21	20.4
4	12	9.0	12	8.0	12	10.4	14	13.1	15	13.3
5	4	7.0	5	5.8	5	7.5	7	9.1	10	9.5
6	9	5.9	4	4.4	3	5.8	8	6.7	7	7.1
7	10	5.0	1	3.5	7	4.6	2	5.0	5	5.5
8	3	4.3	5	2.8	2	3.7	4	3.9	5	4.3
9	4	3.7	3	2.2	4	3.0	3	3.0	2	3.4
10	5	3.2	1	1.8	3	2.4	2	2.3	3	2.7
>10	15	19.6	6	7.8	9	10.9	6	8.7	9	10.8
	A = 49 N = 270 V = 55 n = 374 DF = 11 χ^2 = 18.3411 P = 0.0740		A = 37 N = 132 V = 43 n = 212 DF=11 χ^2 = 20.8172 P = 0.0353		A = 46 N = 179 V = 59 n = 284 DF = 11 χ^2 = 25.7095 P = 0.0072		A = 68 N = 165 V = 56 n = 289 DF = 11 χ^2 = 26.0741 P = 0.0063		A = 59 N = 192 V = 70 n = 321 DF = 11 χ^2 = 4.5151 P = 0.9524	

	Russian 1		**Russian 2**		**Russian 3**		**Russian 4**		**Russian 5**	
i	**c_i**	**E(c_i)**	**c_i**	**E(c_i)**	**c_i**	**E(c_i)**	**c_i**	**E(c_i)**	**c_i**	**E(c_i)**
0	43	52.9	63	75.6	75	79.7	68	75.4	44	64.4
1	32	21.7	35	26.2	24	19.7	34	27.7	44	32.3
2	13	10.2	10	10.8	5	6.2	9	12.0	19	17.6

3	5	5.6	6	5.7	4	2.9	6	6.4	18	10.5
4	5	3.6	5	3.8	2	2.0	10	4.3	7	6.8
5	3	2.6	4	2.9	2	1.6	6	3.2	9	4.8
6	1	2.0	3	2.3	2	1.4	3	2.6	4	3.5
7	0	1.6	4	1.9	2	1.2	3	2.1	1	2.6
8	1	1.3	1	1.6	2	1.1	2	1.8	1	2.0
9	0	1.1	1	1.4	1	0.9	4	1.5	2	1.6
10	1	0.9	1	1.2	0	0.8	1	1.3	0	1.2
>10	4	4.4	5	6.8	5	6.6	5	6.7	3	4.7
	A = 15 N = 72 V = 24 n = 111 DF = 10 χ^2 = 10.6932 P = 0.3819		A = 18 N = 100 V = 25 n = 143 DF = 11 χ^2 =10.1136 P = 0.5202		A = 9 N = 99 V = 19 N = 127 DF = 9 χ^2 = 4.2066 P = 0.8973		A = 21 N = 101 V = 26 n = 148 DF = 11 χ^2 = 19.0247 P = 0.0607		A = 37 N = 89 V = 29 n = 155 DF = 11 χ^2 = 23.4373 P = 0.0153	

	Turkish 1		**Turkish 2**		**Turkish 3**		**Turkish 4**		**Turkish 5**	
i	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**	**c_i**	**$E(c_i)$**
0	116	137.2	191	239.4	161	219.7	150	192.8	144	183.7
1	87	63.0	153	95.9	154	120.7	130	90.7	146	99.3
2	32	32.2	41	44.3	100	68.6	55	46.3	57	55.4
3	23	18.7	19	24.7	39	40.6	17	25.9	37	32.1
4	13	12.3	15	16.3	21	25.0	12	15.9	11	19.4
5	7	8.8	12	12.1	18	16.2	11	10.6	5	12.3
6	3	6.7	12	9.6	9	10.9	9	7.5	10	8.2
7	4	5.2	7	7.7	8	7.7	5	5.6	5	5.7
8	3	4.1	6	6.4	7	5.6	8	4.2	4	4.1
9	5	3.3	10	5.3	1	4.2	7	3.3	3	3.1
10	2	2.6	6	4.3	8	3.3	3	2.6	1	2.5
>10	10	10.9	15	21.0	13	16.5	14	15.5	17	14.2
	A = 62 N = 188 V = 58 n = 308 DF = 11 χ^2 = 17.4930 P = 0.0941		A = 84 N = 322 V = 84 n = 490 DF = 11 χ^2 = 52.7990 P = $1.95 \cdot 10^{-7}$		A = 188 N = 281 V = 73 n = 542 DF = 11 χ^2 = 50.8252 P = $4.45 \cdot 10^{-7}$		A = 125 N = 254 V = 45 n = 424 DF = 11 χ^2 = 40.1653 P = $0.34 \cdot 10^{-4}$		A = 159 N = 232 V=52 n = 443 DF = 11 χ^2 = 41.2974 P = $0.21 \cdot 10^{-4}$	

The better is the fit of the model, the larger is the value P = P(χ^2 > obs.). Some texts are fitted perfectly (e.g. Croatian 2, Chinese 3 and Odia 5), but other ones differ clearly from this model (e.g. Turkish 2-5). It would be interesting to study the linguistic characteristics responsible for agreement or disagreement with the random model.

Moreover, one can observe that in most cases $c_0 < E(c_0)$ and $c_1 > E(c_1)$ hold, i.e. the distance 0 occurs more rarely than predicted by the random model, while the distance 1 occurs more frequently than predicted. In a first hypothesis, one could guess that, in principle, short distances between equal word types are preferred but grammatical rules or semantic reasons prohibit a more frequent use of the distance zero, since generally some parts of speech may not directly follow each other.

Since the fitting must be, in many cases, rejected, we suppose the presence of some boundary conditions which must be studied for every language separately. Of course, the number of data should be considerably increased. Hence, we choose a rather inductive approach and conjecture that from the psychological point of view, according to a hypothesis of B.F. Skinner (1939, 1941, 1959), the probability of the repeated occurrence of an entity in the vicinity of its previous appearance increases, but with time (here distance in text) the stimulus fades away. Hence, small distances between identical entities are more frequent than the great ones. The hypothesis holds for any type of entity. With regard to phonetic similarity of verses it was tested in the old Malay epic poetry (cf. Altmann 1968; Altmann, Köhler 2015: 160 f.) where a stochastic regularity has been found. In poetry the hypothesis can be verified in two ways: concerning individual verses and concerning individual strophes. Of course, the hypothesis of decreasing similarity with increasing distance can be tested with any kind of entities or structures. Skinner conjectured that an entity – mostly a phonetic one – activates the respective part of the brain and the stimulus diminishes slowly with time. However, if the text is not long enough, even opposite tendencies may appear. Thus, one should begin inductively. The longer is the compared entity, the longer must be the texts in order to eliminate a non-smooth course of the similarity curve.

Considering the distances from this point of view, we conjecture that the relative rate of change of frequency *y* in distance *x* between identical entities is simply negatively proportional, i.e.

$$\frac{dy}{y} = -\frac{1}{b}dx \tag{4.3}$$

yielding the simple solution

$$y = a\,exp(-x/b). \tag{4.4}$$

Fitting this function to the above data we obtain the results presented in Table 4.4.

Table 4.4
Fitting the exponential function to the distance data

	Slovak 1		**Slovak 2**		**Slovak 3**		**Slovak 4**		**Slovak 5**	
i	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**
0	74	75.53	64	75.46	87	91.78	55	59.74	88	91.13
1	52	44.68	70	52.68	70	57.62	50	37.52	64	56.56
2	19	26.44	46	36.78	33	36.18	21	23.56	39	35.1
3	11	15.64	19	25.67	17	22.71	7	14.79	7	21.78
4	13	9.25	13	17.92	10	14.26	8	9.29	14	13.52
5	7	5.47	6	12.51	13	8.95	4	5.83	12	8.39
6	5	3.24	5	8.74	5	5.62	3	3.66	6	5.21
7	4	1.92	7	6.1	3	3.53	7	2.3	5	3.23
8	3	1.13	1	4.26	2	2.21	4	1.44	1	2
9	3	0.67	2	2.97	1	1.39	2	0.91	3	1.24
10	3	0.4	1	2.07	2	0.87			4	0.77
	a = 75.5256 b = 1.9052 R^2 = 0.97		a = 75.4573 b = 2.7827 R^2 = 0.89		a = 91.7816 b = 2.1481 R^2 = 0.97		a = 59.7400 b = 2.1493 R^2 = 0.91		a = 91.1317 b = 2.0961 R^2 = 0.96	

	Hung. 1		**Hung. 2**		**Hung. 3**		**Hung. 4**		**Hung. 5**	
i	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**
0	41	47.67	69	71.61	72	78.11	54	58.86	54	65.28
1	48	32.38	53	45.11	64	47.44	51	38.12	63	46.49
2	17	22.00	24	28.41	24	28.81	21	24.69	39	33.11
3	10	14.95	15	17.89	11	17.50	10	15.99	24	23.59
4	11	10.15	12	11.27	3	10.63	5	10.36	10	16.80
5	8	6.90	6	7.10	7	6.45	12	6.71	7	11.96
6	2	4.69	6	4.47	8	3.92	3	4.34	4	8.52
7	3	3.18	2	2.82	3	2.38	4	2.81	6	6.07
8	0	2.16	2	1.77	4	1.45	4	1.82	1	4.32
9	0	1.47	3	1.12	1	0.88	2	1.18	1	3.08
10	1	1.00	2	0.70	1	0.53	0	0.76	1	2.19
	a = 47.6707 b = 2.5864 R^2 = 0.86		a = 71.6149 b = 2.1633 R^2 = 0.98		a = 78.1107 b = 2.0053 R^2 = 0. 92		a = 58.8587 b = 2.3021 R^2 = 0.91		a = 65.2779 b = 2.9469 R^2 = 0.88	

	Croatian 1		**Croatian 2**		**Croatian 3**		**Croatian 4**		**Croatian 5**	
i	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**
0	71	71.69	65	64.91	61	63.41	124	127.74	125	123.12
1	37	33.57	24	23.74	45	40.14	86	73.01	45	51.87
2	11	15.72	6	8.68	29	25.40	33	41.72	25	21.85
3	8	7.36	5	3.17	9	16.08	20	23.85	12	9.20
4	4	3.45	4	1.16	7	10.18	9	13.63	8	3.88
5	2	1.61	3	0.42	6	6.44	10	7.79	9	1.63
6	2	0.76	4	0.16	5	4.08	8	4.45	3	0.69
7	0	0.35	1	0.06	6	2.58	4	2.54	2	0.29
8	3	0.17	1	0.02	3	1.63	7	1.45	1	0.12
9	0	0.08	1	0.01	2	1.03	4	0.83	6	0.05
10	0	0.04	2	0.00	0	0.65	3	0.47	2	0.02
	a = 71,.6894 b = 1.3181 R^2 = 0.99		a = 64.9054 b = 0.9941 R^2 = 0.99		a = 63.4132 b = 2.1863 R^2 = 0.97		a = 127.7444 b = 1.7874 R^2 = 0.97		a = 123.1244 b = 1.1567 R^2 = 0.98	

	Chinese 1		**Chinese 2**		**Chinese 3**		**Chinese 4**		**Chinese 5**	
i	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**
0	281	279.19	386	405.69	409	408.90	406	406.74	315	326.54
1	140	145.17	293	248.17	211	211.54	225	226.87	218	186.00
2	78	75.48	149	151.81	111	109.43	136	126.54	94	105.94
3	37	39.25	76	92.86	55	56.61	71	70.58	53	60.35
4	27	20.41	49	56.81	28	29.29	29	39.37	26	34.37
5	10	10.61	20	34.75	17	15.15	17	21.96	10	19.58
6	6	5.52	16	21.26	6	7.84	14	12.25	10	11.15
7	4	2.87	8	13.00	8	4.05	3	6.83	9	6.35
8	4	1.49	7	7.95	2	2.10	7	3.81	2	3.62
9	8	0.78	3	4.87	3	1.09	5	2.13	3	2.06
10	3	0.40	4	2.98	3	0.56	3	1.19	0	1.17
	a = 279.1933 b = 1.5291 R^2 = 1.00		a = 405.6876 b = 2.0347 R^2 = 0.98		a = 408.9050 b = 1.5173 R^2 = 1.00		a = 406.7396 b = 1.7129 R^2 = 1.00		a = 326.5363 b = 1.7768 R^2 = 0.98	

	Persian 1		**Persian 2**		**Persian 3**		**Persian 4**		**Persian 5**	
i	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**
0	336	334.71	229	232.37	267	266.02	286	289.72	310	309.78
1	177	174.93	152	143.22	131	130.39	175	158.91	187	191.42
2	80	91.42	91	88.27	55	63.91	73	87.16	137	118.28
3	42	47.78	40	54.40	34	31.33	39	47.80	52	73.09
4	38	24.97	31	33.53	16	15.36	23	26.22	43	45.16
5	19	13.05	23	20.67	19	7.53	23	14.38	28	27.91
6	15	6.82	17	12.74	9	3.69	18	7.89	23	17.25
7	12	3.56	11	7.85	12	1.81	14	4.33	12	10.66
8	13	1.86	9	4.84	10	0.89	6	2.37	16	6.58
9	14	0.97	4	2.98	5	0.43	3	1.30	10	4.07
10	8	0.51	3	1.84	4	0.21	5	0.71	3	2.51
	a = 334.7122 b = 1.5411 R^2 = 0.99		a = 232.3705 b = 2.0662 R^2 = 0.99		a = 266.0180 b = 1.4025 R^2 = 0.99		a = 289.7198 b = 1.6650 R^2 = 0.99		a = 309.7802 b = 2.0773 R^2 = 0.99	

	German 1		**German 2**		**German 3**		**German 4**		**German 5**	
i	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**
0	52	57.35	115	125.20	108	108.05	87	87.22	63	66.28
1	49	36.17	102	77.40	63	61.12	53	51.05	48	46.65
2	20	22.82	44	47.85	30	34.57	29	29.88	41	32.83
3	11	14.39	19	29.58	22	19.55	12	17.48	20	23.10
4	6	9.08	13	18.29	9	11.06	11	10.23	20	16.26
5	3	5.73	11	11.30	7	6.25	8	5.99	7	11.44
6	3	3.61	4	6.99	8	3.54	7	3.50	6	8.05
7	2	2.28	2	4.32	0	2.00	5	2.05	2	5.67
8	3	1.44	3	2.67	6	1.13	2	1.20	2	3.99
9	0	0.91	2	1.65	2	0.64	3	0.70	1	2.81
10	2	0.57	0	1.02	2	0.36	3	0.41	1	1.98
	a = 57.3475 b = 2.1702 R^2 = 0.93		a = 125.1973 b = 2.0792 R^2 = 0.94		a = 108.0546 b = 1.7548 R^2 = 0.99		a = 87.2231 b = 1.8667 R^2 = 0.99		a = 66.2787 b = 2.8468 R^2 = 0.97	

	Odia 1		**Odia 2**		**Odia 3**		**Odia 4**		**Odia 5**	
i	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**
0	188	188.52	81	79.10	196	196.30	88	94.57	127	126.28
1	83	79.25	41	47.89	86	83.55	73	63.72	72	72.87
2	27	33.32	37	28.99	31	35.56	50	42.93	42	42.06
3	11	14.01	14	17.55	13	15.14	27	28.93	21	24.27

4	12	5.89	12	10.62	12	6.44	14	19.49	15	14.01
5	4	2.48	5	6.43	5	2.74	7	13.13	10	8.08
6	9	1.04	4	3.89	3	1.17	8	8.85	7	4.66
7	10	0.44	1	2.36	7	0.50	2	5.96	5	2.69
8	3	0.18	5	1.43	2	0.21	4	4.02	5	1.55
9	4	0.08	3	0.86	4	0.09	3	2.71	2	0.90
10	5	0.03	1	0.52	3	0.04	2	1.82	3	0.52
	a = 188.5179 b = 1.1540 R^2 = 0.99		a = 79.1021 b = 1.99248 R^2 = 0.97		a = 196.2965 b = 1.1707 R^2 = 1.00		a = 94.5715 b = 2.5325 R^2 = 0.97		a = 126.2769 b = 1.8191 R^2 = 1.00	

	Russian 1		**Russian 2**		**Russian 3**		**Russian 4**		**Russian 5**	
i	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**
0	43	45.13	63	63.74	75	75.05	68	67.92	44	48.81
1	32	25.96	35	31.06	24	23.38	34	32.04	44	33.60
2	13	14.93	10	15.14	5	7.28	9	15.12	19	23.13
3	5	8.59	6	7.38	4	2.27	6	7.13	18	15.92
4	5	4.94	5	3.60	2	0.71	10	3.37	7	10.96
5	3	2.84	4	1.75	2	0.22	6	1.59	9	7.54
6	1	1.64	3	0.85	2	0.07	3	0.75	4	5.19
7	0	0.94	4	0.42	2	0.02	3	0.35	1	3.57
8	1	0.54	1	0.20	2	0.01	2	0.17	1	2.46
9	0	0.31	1	0.10	1	0.00	4	0.08	2	1.69
10	1	0.18	1	0.05	0	0.00	1	0.04	0	1.17
	a = 45.1305 b = 1.8085 R^2 = 0.97		a = 63.7357 b = 1.3914 R^2 = 0.98		a = 75.0456 b = 0.8573 R^2 = 0.99		a = 67.9217 b = 1.3312 R^2 = 0.96		a = 48.8116 b = 2.6777 R^2 = 0.92	

	Turkish 1		**Turkish 2**		**Turkish 3**		**Turkish 4**		**Turkish 5**	
i	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**	**c_i**	**Theor**
0	116	121.07	191	203.03	161	180.99	150	162.53	144	162.73
1	87	71.34	153	113.57	154	121.65	130	98.01	146	103.24
2	32	42.04	41	63.53	100	81.77	55	59.10	57	65.50
3	23	24.77	19	35.53	39	54.96	17	35.64	37	41.55
4	13	14.60	15	19.88	21	36.94	12	21.49	11	26.36
5	7	8.60	12	11.12	18	24.83	11	12.96	5	16.72
6	3	5.07	12	6.22	9	16.69	9	7.81	10	10.61
7	4	2.99	7	3.48	8	11.22	5	4.71	5	6.73

8	3	1.76	6	1.95	7	7.54	8	2.84	4	4.27
9	5	1.04	10	1.09	1	5.07	7	1.71	3	2.71
10	2	0.61	6	0.61	8	3.41	3	1.03	1	1.72
	a = 121.0681 b = 1.8909 R^2 = 0.97		a = 203.0273 b = 1.7213 R^2 = 0.93		a = 180.9891 b = 2.5170 R^2 = 0.93		a = 162.5280 b = 1.9770 R^2 = 0.93		a = 162.7329 b = 2.1975 R^2 = 0.90	

As can be seen, the divergence between observed and theoretic values in the first two classes remains but the fitting can be accepted in each case. In fact, for linguistic reasons a composed probability model would be appropriate, but this should be postponed until additional languages have been analyzed.

5. Some further aspects

5.1. Predicativity motifs

Besides the problems scrutinized above one can imagine a large number of further problems which could display the details of this aspect of text. Let us mention and describe them stepwise.

Separating the sentences, one would obtain a series of short sequences of *A,N,V* which can be treated in the same way as Köhlerian (2015) motifs. One obtains a rank-frequency distribution, a concentration to one or more motifs, distribution of lengths, etc. However, this is adequate rather for longer texts. Consider, for example the Slovak Text 1, for which one obtains the sentence structures

N,A,N,N,V,N,N,N,

N,N,N,N,V,N,V,N,A,

V,N,N,N,

N,V,A,N,N,N,N,V,N,N,A,N,N,

V,A,N,V,N,A,N,N,

V,V,A,N,V,N,

N,A,N,N,N,A,A,N,N,V,N,N,A,N,N,N,A,N,

N,N,V,N,V,N,V,A,N,A,N,V,N,

A,V,A,N,

N,V,N,N,A,A,N,

N,V,N,A,N,N,

A,A,N,A,N,V,N,N,V,N,N,N,N,

N,V,N,N,N,N,N,

A,V,N,A,N,N,N,N,A,N,N,N,V,N,

V,A,N,V,N,A,N,N,V,

N,N,N,N,V,N,A,N,N,N,N,N,V,

N,V,A,N,N,V,

V,A,N,N,

N,N,A,N,N,V,N,V,A,N,N,N,N,

N,N,A,N,V,N,V,N,A,

N,N,V

The motifs are here qualitatively all different but there are motifs/sentences of the same length, unfortunately represented very scarcely; hence no distribution can be proposed. One obtains

Lengths	Frequency
3	1
4	3
5	0
6	3
7	2
8	2
9	3
10	0
11	0
12	0
13	5
14	1
18	1

The numbers are too small for proposing even an inductive hypothesis.

An alternative way is to count for each motif/sentence the number of *A, N, V,* to obtain various indicators: (a) A sequence of vectors containing the numbers *A,N,V* in individual sentences, and study their change; (b) An indicator of predication sentence-wise expressed quantitatively and yielding a curve which may have quite special properties; (c) If the curves are similar in one text type, one could use them for characterizing also other text types.

Let us consider sentence similarity. If we rewrite the sentence as a sequence of symbols, here *A, N, V,* then some sentence pairs are more similar than others. Here not only the number of identical symbols but also their position is relevant. One may automatically set up the hypothesis that the mean similarity decreases with increasing distance. A hypothesis of this kind has been tested with regard to phonetic similarity of verses in the old Malay epic poetry (cf. Altmann 1968; Altmann, Köhler 2015: 160 f.) where a stochastic regularity has been found. The hypothesis can be tested in poetry in two ways: concerning individual verses and concerning individual strophes. Of course, the hypothesis of decreasing similarity with increasing distance can be tested with any kind of entities or structures. It is in accordance with Skinner's principle of "formal enhancement" (cf. Skinner 1939, 1941, 1957). Skinner conjectured that an entity – mostly a phonetic one – activates the respective part of the brain and the stimulus diminishes slowly with time. Hence, text parts positioned nearer to one another may be more similar than distant ones. This hypothesis can be applied to any kind of entity – from sound to sentence – and the result may be both modeled and used for solving various textological and psycholinguistic problems. However, if the text is not long enough, even opposite tendencies may appear. Thus one should begin inductively

and the longer is the compared entity, the longer must be the texts in order to eliminate strong oscillation.

One can consider the text as a whole, i.e. as one vector. One can set up motifs of different kind, as shown in Köhler (2015), Köhler, Tuzzi (2015) and search for their properties, diversity, etc. This aspect has a good chance to trace down some laws.

5.2. Length-frequency of predication motifs

A further possibility of setting up motifs is to replace the letter *A,N,V* by their predication value, namely $N = 0$, $A = V = 1$. For the above text we obtain

01/001/00000001/01/011/000011/00001/001/0011/01/01/
00111/01/001/00011/001/001/0001/0001/01/011/01/01/0111/
001/0011/001/01/0011/01/01/001/000001/0000011/01/
00001/0001/011/01/01/001/00001/01/000001/011/00111/
00001/001/011/0000001/01/01/01/001/

We obtain the length-frequency distribution presented in Table 5.1. As is usual (cf. Popescu, Best, Altmann 2014), we fit the Zipf-Alekseev function to the data, instead of searching for a distribution (it can be shown that the Zipf-distribution is adequate). The formula is

(5.1) $$y = cx^{a+b\ln(x)}$$

Table 5.1
ANV-Motifs: **Slovak** T 1

Length	**Frequency**	**Zipf-Alekseev**
2	16	16.06
3	15	14.39
4	8	10.07
5	9	6.51
6	4	4.11
7	2	2.59
8	1	1.65
a = 2.2388, b = -1.4001, c = 6.6674, R^2 = 0.95		

For all texts and languages the motif lengths are displayed in Tables 5.2. to 5.10. The parameters *a,b,c* are those of the Zipf-Alekseev function (5.1).

Table 5.2
Length-frequencies of predication motifs: **Slovak**

Length	**T 1**		**T 2**		**T 3**		**T 4**		**T 5**	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	-	-	1	0.99	1	3.64	1	2.85	1	1.78
2	16	16.06	23	23.11	29	27.59	18	16.59	18	18.47
3	15	14.39	28	27.64	21	22.31	14	15.40	23	20.55
4	8	10.07	14	14.90	9	10.26	8	8.91	8	12.60
5	9	6.51	7	6.02	4	3.93	7	4.39	7	6.25
6	4	4.11	2	2.18	10	1.42	2	2.05	7	2.85
7	2	2.59	1	0.76	-	-	-	-	2	1.27
8	1	1.65			2	0.19	1	0.44	-	-
9									1	0.25
10									-	-
11									1	0.05
	a = 2.2388 b = -1.4001 c = 6.6674 R^2 = 0.95		a = 7.1390 b = -3.7376 c = 0.9874 R^2 = 0.997		a = 5.0992 b = -3.1385 c = 3.6367 R^2 = 0.86		a = 4.2628 b = -2.4814 c = 2.8474 R^2 = 0.94		a = 5.3384 b = -2.8322 c = 1.7798 R^2 = 0.91	

Table 5.3
Length-frequencies of predication motifs: **Croatian**

Length	**T 1**		**T 2**		**T 3**		**T 4**		**T 5**	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	1	1.85	-	-	1	0.11	-	-	1	1.01
2	11	10.11	6	5.93	12	12.37	28	28.05	9	8.58
3	9	10.56	5	5.07	22	21.21	25	24.86	12	12.85
4	9	7.13	4	4.16	10	11.59	16	15.84	12	11.70
5	3	4.12	3	3.39	5	4.12	8	9.06	10	8.76
6	-	-	3	2.77	3	1.23	6	5.01	5	6.00
7	1	1.21	4	2.29	1	0.34	3	2.77	4	3.95
8	-	-	1	1.90			1	1.54	2	2.55
9	2	0.35	1	1.60			1	0.87	-	-
10	-	-					1	0.50	1	1.06
11	-	-							-	-
12	-	-							1	0.45
20	1	0.001								
	a = 3.9238		a = 0.4147		a = 10.1980		a = 2.9837		a = 4.4137	

	b = -2.1296 c = 1,8530 R^2 = 0.90	b = -0.4453 c = 5.5055 R^2 = 0.80	b = -4.9496 c = 0.1136 R^2 = 0.98	b = -1.8314 c = 8.5491 R^2 = 0.997	b = -1.9072 c = 1.0064 R^2 = 0.98

Table 5.4
Length-frequencies of predication motifs: **German**

Length	**T 1**		**T 2**		**T 3**		**T 4**		**T 5**	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	2	3.04	-	-	1	4.80	1	3.24	-	-
2	22	21.26	45	44.62	24	19.50	20	18.32	23	22.42
3	10	11.72	22	24.62	13	18.83	16	17.69	16	18.66
4	5	3.54	19	14.33	15	12.53	11	10.80	16	11.69
5	3	0.90	8	8.80	8	7.34	6	5.63	6	6.70
6	3	0.22	4	5.65	4	4.11	1	2.78	2	3.76
7	0	0.05	3	3.76	4	2.28	6	1.35	1	2.11
8	1	0.01	-	-	2	1.27	2	0.66	1	1.20
9	1	0.004	-	-	-	-				
10	-	-	-	-	-	-				
11	-	-	1	0.97	1	0.24				
12	1	0.0001								
	a = 5.5042 b = -3.8914 c = 3.0385 R^2 = 0.94		a = -0.3924 b = -0.5993 c = 78.1123 R^2 = 0.98		a = 3.3520 b = -1.9189 c = 4.8018 R^2 = 0.84		a = 4.1296 b = -2.3529 c = 3.2419 R^2 = 0.90		a = 2.5782 b = -1.6918 c = 8.4624 R^2 = 0.94	

Table 5.5
Length-frequencies of predication motifs: **Persian**

Length	**T 1**		**T 2**		**T 3**		**T 4**		**T 5**	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
2	60	58.81	43	43.64	45	44.52	56	56.49	60	60.73
3	39	44.02	67	65.33	31	33.89	53	50.63	70	67.22
4	35	30.32	29	32.66	29	22.57	28	32.84	38	43.88
5	24	20.63	13	10.90	9	14.58	22	19.16	30	23.70
6	10	14.15	7	3.09	12	9.42	13	10.82	8	11.91
7	11	9.85	4	0.83	3	6.17	5	6.08	7	5.85
8	8	6.97	3	0.22	7	4.10	1	3.45	3	2.87

9	3	5.01	3	0.06	1	2.77	3	1.99	5	1.42
10	2	3.66	1	0.02	2	1.90	2	1.16	1	0.71
11	2	2.71	-	-	1	1.33	1	0.69	1	0.37
12	1	2.03	-	-	2	0.94	1	0.42		
13			1	0.0004	1	0.67	1	0.26		
14			-	-	1	0.49				
15			-	-						
16			1	0.00001						
	a = 0.7916		a = 9.7971		a = 1.2415		a = 2.9193		a = 4.7311	
	b = -0.8403		b = -4.9124		b = -1.0685		b = -1.7802		b = -2.5007	
	c = 508694		c = 0.5196		c = 14.7758		c =5.0738		c = 7.6ß29	
	R^2 = 0.98		R^2 = 0.99		R^2 = 0.96		R^2 = 0.99		R^2 = 0.98	

In Persian T 5, the outlier 1 at length 22 has been omitted.

Table 5.6
Length-frequencies of predication motifs: **Chinese**

Length	**T 1**		**T 2**		**T 3**		**T 4**		**T 5**	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	1	5.81	1	13.53	-	-	1	6.87	1	7.20
2	35	32.24	89	80.02	55	54.13	59	56.79	52	48.16
3	37	37.14	75	82.27	53	56.37	68	67.39	50	53.42
4	24	27.94	48	53.34	47	40.99	41	47.24	38	36.62
5	20	17.97	39	29.42	21	26.25	33	27.31	18	21.11
6	10	10.87	16	15.29	20	16.03	16	14.63	15	11.39
7	9	6.42	12	7.82	7	9.65	5	7.62	6	6.01
8	6	3.78	6	4.01	6	5.82	5	3.95	7	3.17
9	2	2.24	2	2.08	3	3.54	5	2.064	4	1.68
10	3	1.34	-	-	1	2.17	4	1.09	1	0.91
11	1	0.81	-	-	1	1.35	1	0.59	-	-
12	1	0.50	-	-	1	0.85	-	-	-	-
13			1	0.18	1	0.55	-	-	-	-
14					-	-	1	0.10	1	0.09
15					1	0.23	-	-	-	-
16					-	-	1	0.03	-	-
17					-	-			-	-
18					1	0.07			-	-
19									1	0.01
	a = 3.8099		a = 4.1388		a = 3.2221		a = 4.7047		a = 4.3094	
	b = -1.9316		b = -2.2717		b = -1.7424		b = -2.3923		b = -2.2624	
	c = 5.8146		c = 13.5302		c = 13.3988		c = 6.8670		c = 7.2028	
	R^2 = 0.97		R^2 = 0.95		R^2 = 0.98		R^2 = 0.98		R^2 = 0.97	

Table 5.7
Length-frequencies of predication motifs: **Hungarian**

Length	**T 1**		**T 2**		**T 3**		**T 4**		**T 5**	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	-	-	-	-	1	2.99	1	2.30	1	2.76
2	23	22.74	16	15.53	23	21.50	19	18.04	26	25.06
3	9	11.02	15	16.63	17	19.57	16	17.69	22	23.21
4	10	6.27	14	11.82	13	10.46	12	10.09	12	11.92
5	3	3.93	7	7.26	4	4.68	4	4.77	7	5.01
6	2	2.63	4	4.21	2	1.97	2	2.11	1	1.97
7	1	1.85	1	2.41	1	0.82	-	-	-	-
8					1	0.34	1	0.40	1	0.30
9					1	0.15				
	a = -1.3312 b = -0.2537 c = 64.6676 R^2 = 0.94		a = 3.6732 b = -1.9558 c = 3.1147 R^2 = 0.95		a = 4.7901 b = -2.8032 c = 2.9882 R^2 = 0.96		a = 4.8770 b = -2.7485 c = 2.2994 R^2 = 0.97		a = 5.3104 b =-3.0692 c = 2.7590 R^2 = 0.98	

Table 5.8
Length-frequencies of predication motifs: **Odia**

Length	**T 1**		**T 2**		**T 3**		**T 4**		**T 5**	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	1	3.64	-	-	1	3.94	1	5.07	-	-
2	21	19.47	13	13.35	30	28.29	28	23.20	19	19.84
3	22	21.86	19	18.00	23	24.58	16	23.17	27	24.45
4	13	16.10	10	11.73	11	12.47	20	15.54	13	16.85
5	13	10.18	7	5.84	8	5.29	10	9.08	11	9.37
6	4	6.06	2	2.60	4	2.11	5	5.05	4	4.78
7	5	3.54	2	1.11	2	0.84	2	2.77	5	2.37
8	4	2.06	2	0.47	4	0.34	-	-	2	1.16
9	2	1.21	-	-			-	-	1	0.58
10	1	0.71	-	-			1	0.48	-	-
11	-	-	1	-					1	0.15
12	-	-		0.02						
13	-	-								
14	1	0.10								
15	1	0.06								
	a = 3.7654		a = 6.4867		a = 4.8583		a = 3.5784		a = 5.1942	

	b = -1.9428 c = 3.6421 R^2 = 0.94	b = -3.2092 c = 0.6957 R^2 = 0.97	b = -2.9049 c = 3.9384 R^2 = 0.95	b = -1.9985 c = 5.0723 R^2 = 0.84	b = -2.6114 c = 1.9003 R^2 = 0.95

Table 5.9
Length-frequencies of predication motifs: **Russian**

Length	**T 1**		**T 2**		**T 3**		**T 4**		**T 5**	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	1	0.86	1	1.74	-	-	-	-	1	2.42
2	10	10.19	9	8.63	2	1.83	10	9.92	20	18.87
3	11	10.49	10	9.51	5	4.89	9	9.04	15	17.27
4	5	5.67	5	6.99	5	5.79	6	6.94	12	9.11
5	2	2.45	6	4.45	6	4.88	7	5.05	3	3.98
6	2	0.97	3	2.67	3	3.48	3	3.62	1	1.64
7	1	0.38	1	1.58	3	2.28	2	2.59		
8	1	0.15	1	0.55	-	1.43	-	-		
9			1	0.33	1	0.87	-	-		
10					1	0.53	1	1.00		
	a = 5.7810 b = -3.1863 c = 0.8565 R^2 = 0.97		a = 3.5137 b = -1.8834 c = 1.7419 R^2 = 0.92		a = 7.1779 b = -2.6535 c = 0.0453 R^2 = 0.86		a = 1.5543 b = -0.9952 c = 5.4485 R^2 = 0.93		a = 4.9685 b = -2.8949 c = 2.4220 R^2 = 0.94	

Table 5.10
Length-frequencies of predication motifs: **Turkish**

Length	**T 1**		**T 2**		**T 3**		**T 4**		**T 5**	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	-	-	-	-	-	-	1	3.24	-	-
2	30	29.07	46	46.22	60	59.73	42	41.28	56	55.45
3	20	24.13	46	45.13	52	53.39	42	42.64	38	40.61
4	22	16.49	25	26.65	33	29.30	23	22.72	27	23.31
5	12	10.64	14	13.23	10	13.66	8	9.61	14	12.52
6	2	6.79	7	6.18	5	6.04	7	3.73	4	6.66
7	-	-	2	2.84	5	2.64	2	1.42	1	3.58
8	-	-	2	1.32	1	1.17	2	0.54	1	1.96
9	2	1.85	2	0.62	-	-	1	0.21		
10	1	1.24			-	-				
11					1	0.11				

	a = 1.7759 b = -1.2475 c = 115.4597 R^2 = 0.91	a = 4.5244 b = -2.5579 c = 6.8637 R^2 = 0.997	a = 4.4002 b = -2.6102 c = 9.9142 R^2 = 0.99	a = 5.9386 b = -3.2701 c = 3.2391 R^2 = 0.99	a = 2.2334 b = -1.6753 c = 26.3739 R^2 = 0.99

It can be shown that the relation $b = f(a)$ is linear testifying to simple self-regulation. The resulting formulas are presented in Table 5.11.

Table 5.11
Length-frequencies of predication motifs

Language	**b = f(a)**	**R^2**
Slovak	b = – 0.4141 – 0.4784*a	0.95
Croatian	b = – 0.2607 – 0.4541*a	0.98
German	b = – 0.5271 – 0.5153*a	0.86
Persian	b = – 0.4693 – 0.4495*a	1.00
Chinese	b = – 0.2198 – 0.4707*a	0.91
Hungarian	b = – 0.7368 – 0.4126*a	0.97
Odia	b = – 0.4242 – 0.4416*a	0.84
Russian	b = – 0.6926 – 0.3544*a	0.66
Turkish	b = – 0.5265 – 0.4625*a	0.98

As can be seen, only one language, namely Russian, deviates from the usual image, all the other languages follow a unique mechanism. Here we shall not search for this single boundary condition. It can be found only after having scrutinized many other languages.

Though we have modeled the length-frequency dependence by means of a simple function, it is possible to characterize the texts and also the languages applying the <I,S> criterion (cf. Ord 1972; Popescu et al. 2009:155 ff) yielding here five values for each language. The computation formulas, to which we also add the Pearsonian excess are:

Ord's indicators:

$$I = \frac{m_2}{m_1'}$$

$$S = \frac{m_3}{m_2}$$

Pearson's excess:

$$\beta_2 = \frac{m_4}{m_2^2}$$

Table 5.12
Ord's criterion for length-frequency of motifs

Slovak	**I**	**S**	**β_2**
T1	0.6536	1.2335	2.8703
T2	0.4182	1.0410	3.6928
T3	0.7445	1.7285	3.3937
T4	0.5885	1.4480	4.0536
T5	0.8965	2.7781	5.9827
Croatian			
T1	1.4591	5.0960	8.1808
T2	0.9483	0.8911	2.0488
T3	0.4548	1.0419	3.4614
T4	0.8076	2.4094	5.0414
T5	0.9930	2.5352	5.0613
German			
T1	1.3412	4.5083	8.1326
T2	0.7294	2.9014	8.4000
T3	0.9471	2.4125	4.9194
T4	0.8418	1.7397	3.0758
T5	0.5388	1.5121	4.3398
Persian			
T1	1.5051	2.7446	4.4259
T2	1.4892	5.4839	13.8521
T3	1.9581	4.3554	6.3065
T4	1.4503	3.7508	7.4127
T5	1.1227	2.6512	5.5962
Chinese			
T1	1.1131	2.6109	4.2200
T2	0.7992	2.3567	5.9156
T3	1.3042	5.4621	11.6401
T4	1.1554	4.3409	9.0570
T5	1.2713	5.6898	14.2445
Hungarian			
T1	0.5361	1.4491	3.6212
T2	0.4836	0.8043	2.6289
T3	0.7107	2.4126	5.9055

T4	0.5268	1.5772	5.0092
T5	0.4791	1.5398	5.3128
Odia			
T1	1.4826	4.6867	7.5344
T2	0.8625	3.0357	6.4171
T3	0.8107	2.0653	3.8684
T4	0.6876	1.8790	5.3401
T5	0.8708	2.5268	4.9857
Russian			
T1	0.7355	1.9030	4.0340
T2	0.8629	1.6956	3.4512
T3	0.7768	1.5325	3.3150
T4	0.8215	2.0599	4.5744
T5	0.3676	0.6719	2.8799
Turkish			
T1	0.7147	2.7445	7.3507
T2	0.6544	2.1618	5.2992
T3	0.6216	2.6544	8.3235
T4	0.6356	2.0907	5.2352
T5	0.4812	1.3553	4.1032

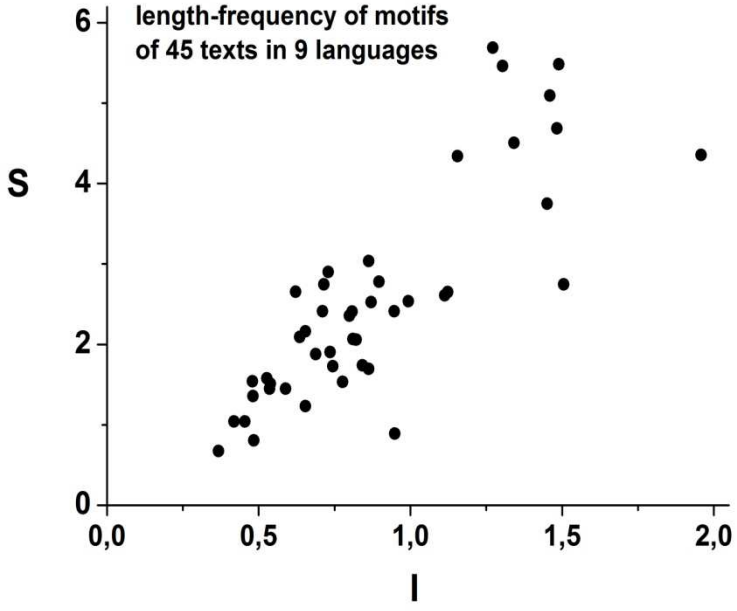

Figure 5.1. The <I,S> criterion for the length-frequency of motifs

The same can be done with <S,β_2> as displayed in Figure 5.2.

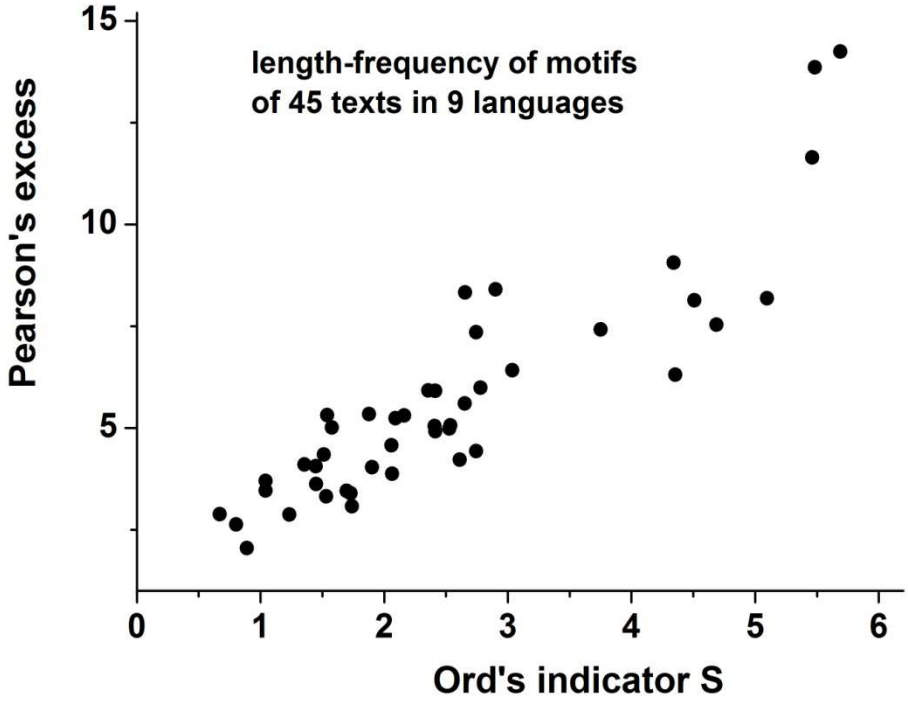

Figure 5.2. The relation of excess to Ord's indicator S.

The trends seem to occupy restricted areas, but it would be premature to formulate hypotheses because we considered only nine languages evaluating only five texts in each. Nevertheless, there are obviously special trends which will possibly hold true also in other languages and other text types. It is, of course, possible that other text types will display different trends.

5.3. Rank-frequency of predication motifs

If one ranks the predication motifs according to their frequency, one obtains for the Slovak Text 1 the results displayed in Table 5.13. The ranking has the advantage that it begins always with x = 1 and there are no zero frequencies. It is possible to use the slightly modified Zipf-Alekseev function by adding 1, i.e. $y = 1 + cx^{a+b\ ln\ x}$ in order to obtain all the expected values greater than 1. Here we examine the types of motifs, e.g. 001 differs from 011.

Table 5.13
Rank-frequency distribution of predication motifs in Slovak T 1

Rank	**Frequency**	**Zipf-Alekseev**
1	16	16.24
2	11	9.57
3	5	6.37
4	4	4.56
5	3	3.44
6	3	2.68
7	3	2.15
8	2	1.76
9	1	1.47
10	1	1.24
11	1	1.06
12	1	0.91
13	1	0.80
14	1	0.70
15	1	0.62
16	1	0.55
a = -0.6109, b = -0.2197 c = 16.2391, R^2 = 0.98		

For all our data we obtain the results in Tables 5.14 to 5.22.

Table 5.14
Rank-frequency distribution of predication motifs in **Slovak**

Rank	**T 1**		**T 2**		**T 3**		**T 4**		**T 5**	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	16	16.24	23	22.90	29	28.93	18	18.14	18	18.17
2	11	9.57	15	15.93	12	12.48	10	8.90	14	13.42
3	5	6.37	13	10.50	8	7.59	4	5.52	9	8.99
4	4	4.56	5	7.13	5	5.32	4	3.83	5	6.13
5	3	3.44	5	5.01	5	4.04	3	2.83	5	4.30
6	3	2.68	4	3.63	3	3.22	2	2.19	3	3.10
7	3	2.15	3	2.70	3	2.65	2	1.76	2	2.29
8	2	1.76	2	2.06	2	2.24	1	1.44	1	1.73
9	1	1.47	1	1.59	2	1.93	1	1.20	1	1.33
10	1	1.24	1	1.25	2	1.69	1	1.02	1	1.04
11	1	1.06	1	1.00	1	1.50	1	0.88	1	0.82
12	1	0.91	1	0.81	1	1.35	1	0.76	1	0.66

13	1	0.80	1	0.66	1	1.22	1	0.67	1	0.54
14	1	0.70	1	0.55	1	1.11	1	0.59	1	0.44
15	1	0.62			1	1.01	1	0.53	1	0.36
16	1	0.55							1	0.30
17									1	0.25
18									1	0.21
19									1	0.18
	a = -0.6109 b = -0.2197 c = 16.2391 R^2 = 0-98		a = -0.2062 b = -0.4583 c = 22.8979 R^2 = 0.98		a = -1.2042 b = -0.0122 c = 28.9294 R^2 = 0.997		a = -0.9316 b = -0.1381 c = 18.1449 R^2 = 0..99		a = -0.0906 b = -0.5000 c = 18.1664 R^2 = 0.99	

Table 5.15
Rank-frequency distribution of predication motifs in **Croatian**

Rank	**T 1**		**T 2**		**T 3**		**T 4**		**T 5**	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	11	10.99	6	5.90	12	11.99	28	27.97	9	8.82
2	6	5.74	3	3.51	12	12.34	16	15.76	8	8.62
3	3	3.96	3	2.63	10	8.81	9	10.32	7	7.04
4	3	3.05	2	2.15	5	5.92	8	7.36	6	5.61
5	3	2.49	2	1.85	4	3.98	6	5.53	6	4.48
6	3	2.12	2	1.63	2	2.71	5	4.32	4	3.61
7	2	1.85	2	1.48	2	1.88	4	3.47	2	2.95
8	2	1.64	1	1.35	1	1.33	3	2.84	2	2.43
9	1	1.48	1	1.25	1	0.95	2	2.37	1	2.02
10	1	1.35	1	1.17	1	0.70	2	2.01	1	1.70
11	1	1.24	1	1.10	1	0.52	1	1.73	1	1.44
12	1	1.15	1	1.05	1	0.39	1	1.49	1	1.23
13			1	1.00	1	0.29	1	1.31	1	1.06
14			1	0.95	1	0.23	1	1.15	1	0.92
15							1	1.02	1	0.80
16							1	0.91	1	0.70
17									1	0.62
18									1	0.55
19									1	0.48
20									1	0.43
21									1	0.39
	a = -0.9463 b = 0.0154 c = 10.9883 R^2 = 0.97		a = -0.7692 b = 0.0294 c = 5.9001 R^2 = 0.96		a = 0.5930 b = -0.7945 c = 11.9868 R^2 = 0.98		a = -0.6920 b = -0.1959 c = 27.9740 R^2 = 0.99		a = 0.2600 b = -0.4231 c = 8.8173 R^2 = 0.95	

Table 5.16
Rank-frequency distribution of predication motifs in **German**

Rank	T 1		T 2		T 3		T 4		T 5	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	22	22.09	44	43.99	24	23.67	20	19.75	23	22.75
2	8	7.04	17	16.63	8	10.22	8	9.71	9	10.57
3	3	3.90	8	9.51	8	6.50	8	6.36	7	6.68
4	2	2.66	6	6.43	5	4.79	5	4.69	7	4.80
5	2	2.02	6	4.75	5	3.82	4	3.70	4	3.70
6	2	1.64	5	3.72	5	3.20	3	3.04	3	2.99
7	1	1.38	5	3.03	2	2.76	3	2.57	3	2.49
8	1	1.20	2	2.54	2	2.44	2	2.23	2	2.13
9	1	1.06	2	2.17	2	2.20	2	1.96	1	1.85
10	1	0.96	1	1.89	2	2.00	2	1.75	1	1.63
11	1	0.88	1	1.67	2	1.84	1	1.57	1	1.45
12	1	0.81	1	1.49	2	1.71	1	1.43	1	1.31
13	1	0.76	1	1.34	1	1.60	1	1.31	1	1.19
14	1	0.71	1	1.22	1	1.51	1	1.21	1	1.09
15	1	0.67	1	1.11	1	1.43	1	1.12	1	1.00
16			1	1.02	1	1.35	1	1.04		
17					1	1.29				
	a = -1.7751 b = 0.1798 c = 22.0945 R^2 = 0.99		a = -1.4188 b = 0.0226 c = 43.9886 R^2 = 0.99		a = -1.2722 b = 0.0866 c = 23.6764 R^2 = 0.97		a = -1.0121 b = -0.0180 c = 19.7506 R^2 = 0.98		a = -1.0896 b = -0.0238 c = 22.7489 R^2 = 0.98	

Table 5.17
Rank-frequency distribution of predication motifs in **Persian**

Rank	T 1		T 2		T 3		T 4		T 5	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	60	59.05	43	43.35	45	44.34	56	55.93	60	59.55
2	22	27.36	40	39.20	18	22.14	33	32.84	36	39.45
3	19	17.50	27	25.70	17	14.26	20	21.51	34	26.98
4	17	12.77	13	16.20	14	10.28	15	15.15	16	19.37
5	11	10.00	9	10.32	7	7.91	13	11.21	13	14.45
6	9	8.20	7	6.72	6	6.35	10	8.61	11	11.12
7	7	6.94	5	4.48	6	5.25	8	6.80	9	8.76
8	7	6.00	4	3.06	4	4.44	5	5.48	8	7.05
9	6	5.28	3	2.13	4	3.82	4	4.50	6	5.76

10	6	4.71	3	1.51	4	3.34	3	3.76	5	4.77
11	4	4.25	3	1.09	2	2.95	2	3.17	4	4.00
12	3	3.87	2	0.80	2	2.63	2	2.71	3	3.39
13	3	3.55	1	0.60	2	2.36	2	2.33	3	2.90
14	3	3.28	1	0.45	2	2.14	2	2.02	2	2.50
15	3	3.05	1	0.34	2	1.95	2	1.77	2	2.17
16	3	2.84	1	0.26	1	1.78	1	1.56	2	1.90
17	2	2.66	1	0.21	1	1.64	1	1.38	1	1.67
18	2	2.50	1	0.16	1	1.52	1	1.23	1	1.48
19	2	2.36	1	0.13	1	1.41	1	1.10	1	1.31
20	1	2.24	1	0.10	1	1.31	1	0.99	1	1.17
21	1	2.12	1	0.08	1	1.22	1	0.89	1	1.05
22	1	2.02	1	0.07	1	1.14	1	0.81	1	0.94
23	1	1.93	1	0.05	1	1.07	1	0.73	1	0.85
24	1	1.84	1	0.04	1	1.01	1	0.67	1	0.77
25	1	1.76	1	0.04					1	0.70
26									1	0.63
	a = -1.1152 b = 0.0076 c = 59.0473 R^2 = 0.98		a = 0.4195 b = -0.8149 c = 43.3535 R^2 = 0.99		a = -0.9498 b = -0.0754 c = 44.3448 R^2 = 0.98		a = -0.5934 b = -0.2517 c = 59.9529 R^2 = 0.997		a = -0.3781 b = -0.3118 c = 59.5489 R^2 = 0.98	

Table 5.18
Rank-frequency distribution of predication motifs in **Chinese**

Rank	T 1		T 2		T 3		T 4		T 5	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	35	35.19	89	88.90	55	54.61	59	58.37	52	51.34
2	22	21.19	46	45.75	33	34.43	34	38.21	26	29.82
3	15	14.74	29	29.35	23	23.73	34	26.67	24	20.13
4	10	11.06	17	20.89	20	17.41	16	19.65	14	14.73
5	8	8.70	16	15.83	13	13.34	15	15.07	13	11.34
6	6	7.08	15	12.50	11	10.55	11	11.90	11	9.04
7	6	5.90	14	10.17	10	8.55	11	9.63	7	7.40
8	6	5.00	10	8.47	9	7.06	10	7.93	5	6.19
9	5	4.31	8	7.18	5	5.93	6	6.64	5	5.25
10	4	3.76	7	6.17	4	5.04	5	5.63	4	4.52
11	4	3.31	5	5.37	3	4.33	5	4.82	3	3.93
12	3	2.95	4	4.72	3	3.76	3	4.17	3	3.46
13	3	2.64	4	4.19	3	3.29	3	3.64	3	3.06
14	2	2.38	3	3.74	2	2.90	3	3.20	2	2.73
15	2	2.16	3	3.37	2	2.57	3	2.83	2	2.45
16	2	1.96	2	3.04	2	2.30	2	2.52	2	2.21

17	2	1.80	2	2.77	2	2.06	2	2.25	2	2.00
18	2	1.65	2	2.53	2	1.86	2	2.02	2	1.83
19	1	1.52	2	2.32	1	1.68	2	1.82	2	1.67
20	1	1.41	2	2.14	1	1.53	1	1.65	1	1.53
21	1	1.31	1	1.97	1	1.39	1	1.50	1	1.41
22	1	1.22	1	1.83	1	1.27	1	1.37	1	1.30
23	1	1.14	1	1.70	1	1.17	1	1.25	1	1.21
24	1	1.06	1	1.58	1	1.07	1	1.15	1	1.12
25	1	1.00	1	1.48	1	0.99	1	1.06	1	1.05
26	1	0.94	1	1.39	1	0.92	1	0.98	1	0.98
27	1	0.88	1	1.30	1	0.85	1	0.90	1	0.91
28	1	0.83	1	1.22	1	0.79	1	0.83	1	0.86
29	1	0.79	1	1.15	1	0.73	1	0.77	1	0.80
30	1	0.74			1	0.68	1	0.72	1	0.76
31					1	0.64	1	0.67	1	0.71
32					1	0.60	1	0.63		
33					1	0.56	1	0.59		
34					1	0.53				
	a = -0.6292 b = -0.1485 c = 35.1917 R^2 = 0.996		a = -0.8721 b = -0.1244 c = 88.8999 R^2 = 0.995		a = -0.5061 b = -0.2297 c = 54.6062 R^2 = 0.99		a = -0.4368 b = -0.2514 c = 58.3688 R^2 = 0.98		a = -0.6673 b = -0.1684 c = 51.3440 R^2 = 0.99	

Table 5.19
Rank-frequency distribution of predication motifs in **Hungarian**

Rank	T 1		T 2		T 3		T 4		T 5	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	23	22.83	16	16.09	23	22.86	19	19.25	26	25.88
2	5	6.78	10	9.45	11	12.13	15	13.65	13	13.77
3	5	3.86	6	6.43	9	7.49	6	7.70	9	8.54
4	4	2.77	4	4.73	6	5.07	3	4.35	7	5.80
5	4	2.22	4	3.66	3	3.64	3	2.54	4	4.17
6	2	1.90	3	2.93	2	2.72	3	1.54	3	3.14
7	1	1.70	3	2.41	2	2.10	2	0.96	2	2.43
8	1	1.56	2	2.02	1	1.67	1	0.62	2	1.93
9	1	1.46	2	1.72	1	1.35	1	0.41	1	1.57
10	1	1.39	1	1.48	1	1.11	1	0.28	1	1.29
11	1	1.33	1	1.29	1	0.92	1	0.19	1	1.08
12			1	1.14	1	0.78			1	0.91
13			1	1.01	1	0.66				

14			1	0.90	1	0.57				
15			1	0.81						
16			1	0.73						
	a = -1.9832 b = 0.3329 c = 22.8330 R^2 = 0.97		a = -0.6516 b = -0.1665 c = 16.0863 R^2 = 0.99		a = -0.7413 b = -0.2493 c = 22.8605 R^2 = 0.99		a = 0.0816 b = -0.8331 c = 19.2466 R^2 = 0.97		a = -0.7409 b = -0.2438 c = 25.8782 R^2 = 0.995	

Table 5.20
Rank-frequency distribution of predication motifs in **Odia**

Rank	T 1		T 2		T 3		T 4		T 5	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	21	21.00	13	13.01	30	30.31	28	27.83	21	21.73
2	16	16.03	11	10.91	17	14.61	13	14.30	19	15.08
3	11	11.49	8	8.09	6	8.95	11	9.13	6	0.29
4	10	8.38	6	5.98	6	6.15	6	6.47	6	7.28
5	6	6.28	4	4.50	4	4.52	5	4.88	6	5.33
6	4	4.81	4	3.46	3	3.48	4	3.84	5	4.01
7	3	3.77	3	2.70	3	2.77	3	3.11	3	3.10
8	3	3.00	2	2.15	3	2.26	3	2.58	2	2.44
9	2	2.43	2	1.73	3	1.88	2	2.18	2	1.96
10	2	1.99	1	1.41	2	1.59	2	1.87	2	1.59
11	2	1.65	1	1.16	2	1.36	1	1.62	2	1.31
12	2	1.38	1	0.97	1	1.18	1	1.42	2	1.09
13	1	1.17			1	1.03	1	1.26	1	0.91
14	1	0.99			1	0.91	1	1.12	1	0.78
15	1	0.85			1	0.81	1	1.00	1	0.66
16	1	0.74					1	0.91	1	0.57
17	1	0.64							1	0.49
18	1	0.56							1	0.43
19									1	0.38
	a = -0.1164 b = -0.3940 c = 21.0001 R^2 = 0.99		a = 0.0525 b = -0.4422 c = 13.0147 R^2 = 0.99		a = -0.9547 b = -0.1413 c = 30.3106 R^2 = 0.98		a = -0.8689 b = -0.1320 c = 27.8261 R^2 = 0.99		a = -0.2648 b = -0.3783 c = 21.7321 R^2 = 0.93	

Table 5.21
Rank-frequency distribution of predication motifs in **Russian**

Rank	T 1		T 2		T 3		T 4		T 5	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	10	10.15	9	9.16	4	3.88	10	10.19	20	19.89
2	7	6.108	8	7.36	4	4.34	8	7.01	10	10.87
3	3	4.08	5	5.28	4	3.78	4	4.90	8	6.66
4	3	2.92	3	3.80	3	3.12	3	3.56	5	4.43
5	2	2.19	3	2.80	3	2.56	3	2.69	2	3.12
6	2	1.70	2	2.10	2	2.10	2	2.08	2	2.29
7	1	1.36	2	1.62	2	1.74	2	1.65	2	1.73
8	1	1.11	1	1.26	1	1.45	1	1.34	1	1.34
9	1	0.92	1	1.00	1	1.21	1	1.10	1	1.07
10	1	0.77	1	0.81	1	1.03	1	0.91	1	0.86
11	1	0.66	1	0.66	1	0.88	1	0.77		
12	1	0.57	1	0.54			1	0.65		
13							1	0.56		
	a = -0.5657 b = -0.2399 c = 10.1450 R^2 = 0.97		a = 0.0032 b = -0.4600 c = 9.1608 R^2 = 0.98		a = 0.4775 b = -0.4581 c = 3.8839 R^2 = 0.96		a = -0.3196 b = -0.3163 c = 10.1906 R^2 = 0.97		a = -0.6590 b = -0.3062 c = 19.8919 R^2 = 0.99	

Table 5.22
Rank-frequency distribution of predication motifs in **Turkish**

Rank	T 1		T 2		T 3		T 4		T 5	
	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor	Emp	Theor
1	30	29.59	46	46.29	59	58.81	42	41.95	55	54.29
2	14	16.65	37	35.60	31	32.14	26	26.31	19	24.24
3	14	10.79	20	21.42	21	19.93	16	16.55	19	14.70
4	7	7.60	11	12.75	13	13.44	14	10.93	14	10.17
5	6	5.64	9	7.78	11	9.59	7	7.55	8	7.59
6	6	4.36	5	4.90	9	7.13	3	5.40	6	5.95
7	4	3.46	4	3.19	4	5.47	3	3.98	5	4.83
8	2	2.81	3	2.13	3	4.30	3	3.01	3	4.02
9	1	2.33	2	1.45	3	3.45	2	2.32	3	3.41
10	1	1.95	1	1.01	3	2.81	2	1.82	2	2.94
11	1	1.66	1	0.72	2	2.33	2	1.45	2	2.57
12	1	1.43	1	0.52	2	1.95	2	1.17	1	2.27

13	1	1.24	1	0.38	1	1.65	1	0.95	1	2.02
14	1	1.08	1	0.29	1	1.41	1	0.79	1	1.81
15			1	0.22	1	1.21	1	0.65	1	1.64
16			1	0.16	1	1.05	1	0.55	1	1.49
17					1	0.92	1	0.46		
18					1	0.81	1	0.39		
	a = -0.6785 b = -0.2180 c = 29.5862 R^2 = 0.97		a = 0.1725 b =-0.7955 c = 46.2912 R^2 = 0.996		a = -0.6788 b = -0.2785 c = 58.8069 R^2 = 0.997		a = -0.3760 b = -0.4285 c = 41.9478 R^2 = 0.99		a = -1.1192 b = -0.0640 c = 54.2923 R^2 = 0.98	

Since we used again the Zipf-Alekseev function, the relationship $b = f(a)$ is also a straight line given in Table 5.23 and presented graphically in Figure 5.3.

Table 5.23
Rank-frequency distribution of predication motifs

	b = f(a)	**R^2**
Slovak	b = – 0.5334 – 0.4399*a	0.98
Croatian	b = – 0.4222 – 0.4776*a	0.88
German	b = – 0.2819 – 0.2523*a	0.74
Persian	b = – 0.5672 – 0.5310*a	0.99
Chinese	b = – 0.3708 – 0.2994*a	0.82
Hungarian	b = – 0.6675 – 0.5396*a	0.92
Odia	b = – 0.4372 – 0.3243*a	0.98
Russian	b = – 0.3959 – 0.1869*a	0.70
Turkish	b = – 0.6661 – 0.5769*a	0.97

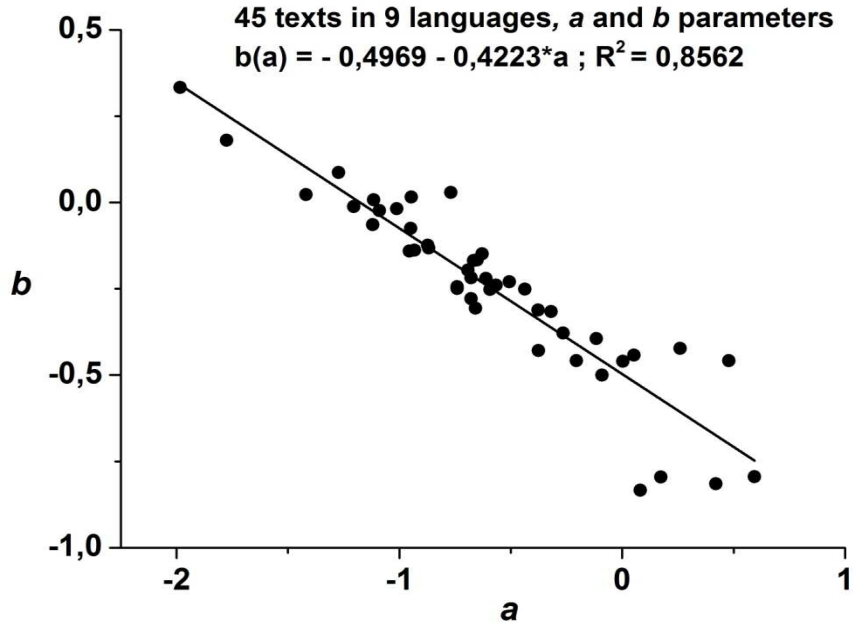

Figure 5.3. Relation between the parameters *b* and *a* for the rank-frequency function of predication motifs

The result shows that in all these analyzed languages there is a certain regularity with very small deviations. Hence, we can conjecture that behind this type of motifs there is a mechanism which could – after analyzing many other languages – be considered a law.

In order to distinguish individual languages they are presented separately in Figure 5.4.

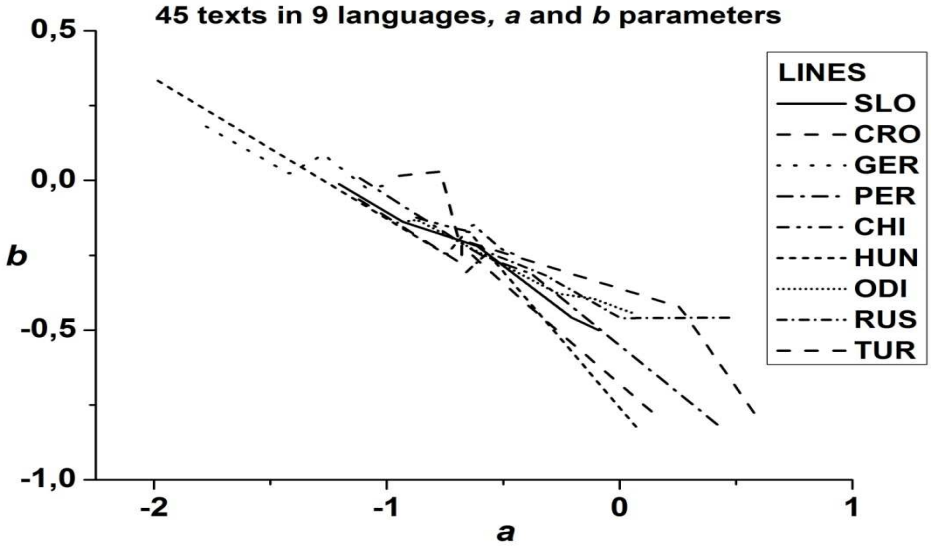

Figure 5.4. Relation between the parameters *b* and *a* for the rank-frequency function of predication motifs with individual languages presented separately

The <I,S> criterion can be applied here, too. Computing the individual values we obtain the results presented in Table 5.24 and Figure 5.5.

Table 5.24
Ord's criterion and excess for rank-frequency of individual texts

Language	**I**	**S**	**β_2**
Slovak			
T 1	3.6838	5.4294	3.9309
T 2	2.6104	4.9409	5.2612
T 3	3.2944	5.2300	4.5053
T 4	3.6508	5.5563	4.1777
T 5	4.6938	7.6764	4.7756
Croatian			
T 1	2.4495	2.8150	2.7833
T 2	2.9943	2.8501	2.3524
T 3	2.7086	5.0403	4.7232
T 4	3.1874	5.3816	4.8325
T 5	4.8428	7.0445	3.6051
German			
T 1	4.1970	6.0168	4.1643
T 2	3.4171	6.1446	5.9000
T 3	3.8774	5.3091	3.6862
T 4	3.5340	5.0943	3.8460
T 5	3.2094	5.2544	4.5950
Persian			
T 1	5.5285	8.3597	4.7521
T 2	5.5058	11.2897	7.9873
T 3	5.4784	9.1798	5.6081
T 4	5.0232	9.7456	7.1328
T 5	5.2234	10.1538	7.1589
Chinese			
T 1	7.2011	10.8732	4.8456
T 2	6.3841	11.7762	7.2866
T 3	8.1601	14.8837	7.2004
T 4	7.4363	13.7358	7.1550
T 5	7.4237	12.6462	6.1925

Hungarian			
T 1	2.3036	3.7619	4.3562
T 2	3.5360	5.1145	3.7977
T 3	2.9770	5.5105	5.3631
T 4	2.0670	3.7863	4.6120
T 5	2.2041	4.0423	4.9131
Odia			
T 1	3.7104	6.1460	4.5513
T 2	2.0723	2.9941	3.3545
T 3	3.3371	4.9634	4.0545
T 4	3.3680	5.5713	4.7486
T 5	4.2987	6.6104	4.3424
Russian			
T 1	2.6457	3.7038	3.3043
T 2	2.3519	3.4414	3.3763
T 3	1.7943	2.0021	2.5924
T 4	2.7656	3.8900	3.3558
T 5	1.7785	3.2627	4.4935
Turkish			
T 1	2.4952	4.6462	5.3118
T 2	2.7099	6.4079	8.0923
T 3	3.3614	6.8648	6.9929
T 4	3.6768	7.2797	6.5644
T 5	2.9582	5.6957	6.1252

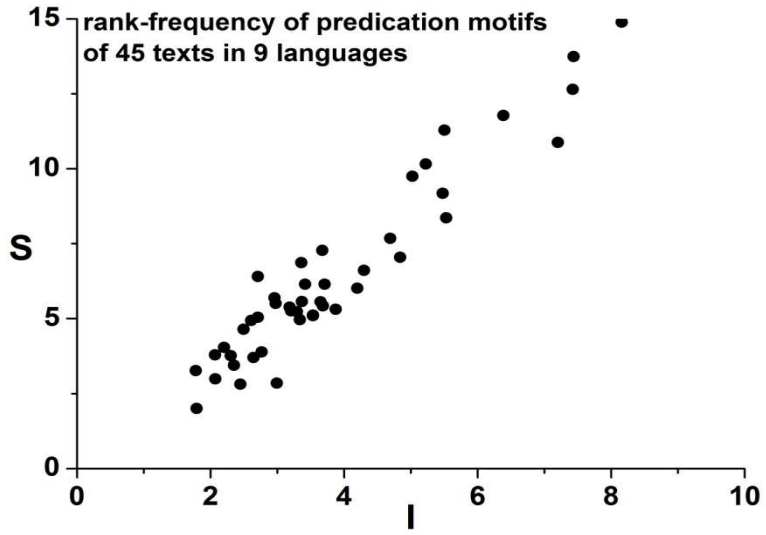

Figure 5.5. Ord's criterion for the rank-frequency relation in individual texts

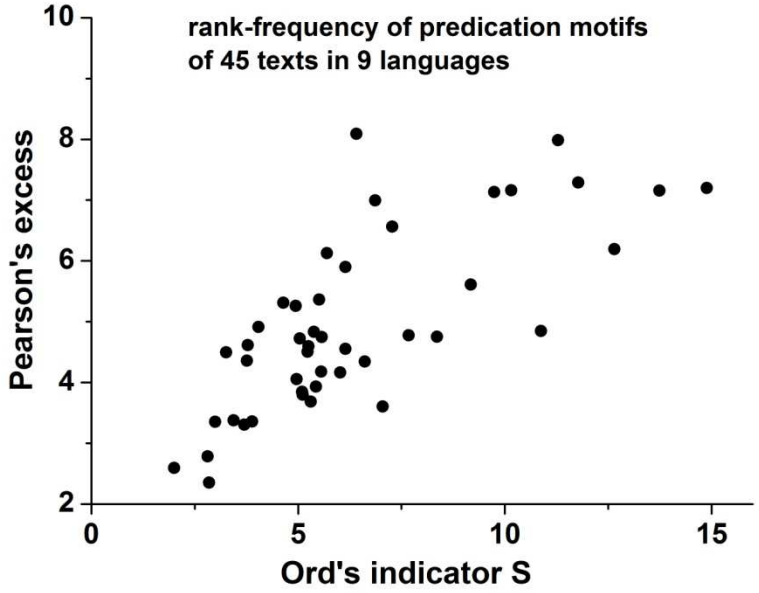

Figure 5.6. Relation between S and β_2 for the rank-frequency of motifs

The area of <S,β_2> is wider, but its exact form can be scrutinized after many languages have been analyzed. In any case we conjecture that here some background mechanisms are active.

In this way, it is possible to show that even the motifs of descriptiveness, activity, nominality are linked with other properties of text, here length and frequency. The research may concentrate on the finding of the control cycle (cf. Köhler 2005) by scrutinizing further properties. Though this would be a great step toward theorizing, however, it would be premature to consider the results as laws.

One could continue in scaling the predicativity/specification: all words that determine in some way the adjectives or verbs would obtain degree 2, etc. One would obtain the text as a sequence of numbers expressing various properties. Here we merely want to make a hint at this possibility. In the same way, one could study the valence of individual words but one would be forced to restrict the study to individual texts and the valences contained in it.

Without much effort one can use the tree-like structures or hierarchies known from various branches of linguistics, e.g. grammar, lexeme nets or definition chains (cf. Sambor, Hammerl 1991). However, in definition chains dictionary issues are analyzed, not sentences; but one could prepare all chains and apply them. The top word obtains degree 0, the other ones the degree of the level counted from above. In this way all words of a sentence obtain a degree and the sequence can be further analyzed.

Even the dependence structure of sentence can be transformed in a numerical sequence. Write the sentence as a sequence and join each dependent word with its main word using an arrow (main word → dependent word). The word

having no ingoing arrows has degree 0, the words depending directly on it have degree 1, the words depending on the words of degree 1 have degree 2, etc. In this way, the sentence obtains the form of a numerical sequence representing dependence.

A quite different approach would be the scaling of nominal, verbal and adjectival classes from different points of view, or the study of metaphoric, poetic and rhetoric aspects of the given parts of speech (cf. Beliankou, Köhler, Naumann 2013) or the whole sentences.

For typological purposes, one would be forced to analyze a great number of texts taken from various text types in every language. One would never obtain a representative image even if one would analyze all texts in a corpus. Each corpus is restricted in some sense, e.g. historically or dialectally, or does not contain texts *spoken* by children, text of everyday conversation of millions of people, etc. Hence, it is more expedient to restrict the texts to a special text type.

The same holds for the historical study of language evolution. Comparing, for example Latin texts with modern Spanish ones, one would be forced to find texts of the same text type. Here etymology does not play any role, but certain formal sequences may be created by a single author for special purposes and cause differences.

The continuation of this study will be full of boundary conditions. As long as one can do with one model, one should apply it. But if the fitting is not good, i.e. a hypothesis has been falsified, one should not hesitate to variegate the formula by adding a new parameter in the differential equation leading to the Zipf-Alekseev formula. Originally we have

$$\frac{dy}{y} = \frac{a + b\ln x}{cx} dx, \tag{5.2}$$

indicating that the relative rate of change of y is proportional to the relative rate of change of x, that is, e.g. frequency depends on length. The proportionality function in the numerator represents the state of the language expressed by the constant a which is usually modified by the requirements and forces of the speaker, here expressed in form of a simple logarithmic function (the speaker cannot make drastic changes). The constant c in the denominator represents the equilibrating force of the hearer or of the community. It is evident that boundary conditions concerning language, text type, historical epoch, etc. may be placed symbolically in the above formula. In this way, it is to be hoped that after having tested the Zipf-Alekseev formula in various data one can approach a law. The above differential equation is part of the unified theory (cf. Wimmer, Altmann 2005) and may be helpful in the first steps. If one adds further functions in the above formula, one should always substantiate them linguistically. Though we avoid polynomials, we accept them if they are well grounded.

6. Conclusions

In the present book we have shown merely some vistas of a possible future research. Though time series, autocorrelation, fractals, Fourier analysis, Hurst exponent, Lyapunov coefficients, Markov chains, etc. are well known methods, they are helpful especially at the first steps of sequential analysis: they yield some characteristics of the examined sequence, but they do not offer an insight in the linguistic background. They operate with ready-made data, show the surface mechanism, but not all of them explicate the mechanism in the background.

The linguistic background can be illuminated through a series of steps. One can begin directly in the text and define phonetically, grammatically, semantically, lexically, etc. the entities. Then a property of the entities – whether qualitative or quantitative – is defined and a hypothesis about its textual behavior is formulated. If the property is quantitative – a sign of progress – then the hypothesis can be derived from a common background, e.g. from the unified theory (cf. Wimmer, Altmann 2005), in which the necessary requirements and the forces of the speaker and hearer are interpreted in the form of mathematical functions and inserted into the basic formulas (cf. Köhler 2005). Frequently, one applies an inductive procedure, begins to search for models applying software, but in the end one must perform some necessary theoretical steps. The data which are the result of our definitions may be improved, changed, variegated, and the hypotheses must be tested (even in the changed state). In case of success one has made the first step towards a theory. In case of rejection one must again check both the data and the hypothesis and make changes wherever it is necessary. It is very important to be aware of the fact that data are not given but created. Besides, different boundary conditions may destroy an otherwise well functioning theory.

Our recommendation is as follows: quantify as many properties as possible and express the text or its parts in the form of numerical or qualitative sequences. Here we strived for describing some rarely analyzed problems and obtained at least partial results: some numerical indicators and some theoretical ones. There are sequential phenomena abiding by laws which are quite general in language. Predication/specification in the sense discussed here is only a special case and may be deepened by quantifying the levels. That means, the grammatical analysis of a sentence may and should be quantified. Instead of trees and dependency graphs one obtains the predicative/specification structure of the sentence in the form of a numerical sequence which can be processed in various ways. One can perform a grammatical analysis, a logical analysis, a topic-comment analysis which is perhaps the simplest way, etc. In all cases one can distinguish levels, e.g. main topic with degree = 0, comment of the first level = 1, specifications within comment = 2, 3, 4,…, etc. All these procedures are merely aspects of the same problem, but all should be taken into account. In going into

the depth, at some point one finds a law in the form of a formalized regularity holding true for all languages.

The linguistic literature is full of various – sometimes very detailed – classifications of the semantic aspect of individual word classes. There are, for example, semantic and grammatical classifications of verbs in 10, 20 or 100 classes. But what is to do if the same verb is strongly polysemic and falls simultaneously in different qualitative classes? One must be aware of the fact that classifications are nothing but concept formations. M. Bunge (1983: 17) makes a distinction between a taxonomic and a theoretical account. What is the criterion of truth-likeness of a classification? Which of the 500 mechanical classification methods yields "the best" result? Nevertheless, it is an important activity because it helps us to find orientation in our limitless concept-formation; it is the first step in ordering the "state of the affairs". But it is not the last. We make the next steps if we show that the given classification is linked in some way with other classifications, phenomena, concepts; if we find the *requirements* (cf. Köhler 2005) of the language community leading to exactly the given state; if we set up a derived mathematical model of the phenomenon which facilitates its treatment mechanically and exactly; and finally, if we subsume the given model under an existing theory. This way is long, or better, limitless, just as in any other science.

In dynamical systems, a classification means only an intuitive finding of attractors. But attractors are seldom isolated entities: they care for self-regulation which can be captured by control cycles. Self-organization means the abandoning of an attractor and finding a new one, a daily discovery in historical linguistics.

But all this just began to be studied because it is somehow associated with elementary mathematics which affords us with the possibility of formal treatment of our concepts.

Studying sequences, one will frequently meet entities that "do not want" to abide by any regularity. In that case one will be forced to re-define the problem at several levels. The data (their measurement) can be inadequate, but most probably there are some conditions not present in other languages. At this point one must perhaps introduce a third variable, and the formulas, especially the differential equations, will be more complex. Arriving at this point, mathematics will be necessary.

References

Altmann, G. (1968). Some phonic features of Malay shaer. *Asian and African Studies 4, 9-16.*

Altmann, G. (1987). Zur Anwendung der Quotiente in der Textanalyse. In: Altmann, G. (ed.), *Glottometrika 1: 91-106.* Bochum: Brockmeyer.

Altmann, G. (1988). *Wiederholungen in Texten*. Bochum: Brockmeyer.

Altmann, V., Altmann, G. (2008). *Anleitung zu quantitativen Textanalysen*. Lüdenscheid: RAM-Verlag.

Altmann, G., Köhler, R. (2015). *Forms and degrees of repetitions in texts. Detection and analysis.* Berlin/Boston: de Gruyter Mouton.

Antosch, F. (1969). The diagnosis of lierary style with the verb-adjective ratio. In: Doležel, L., Bailey, R.W. (eds.), *Statistics and Style: 57-65*. New York. Elsevier.

Bakker, F.J. (1965). Untersuchungen zur Entwicklung des Aktionsquotienten. *Archiv für die gesamte Psychologie 117, 78-101*.

Ballmer, T.T., Brennenstuhl, W. (1986). *Deutsche Verben. Eine sprachanalytische Untersuchung des Deutschen Wortschatzes*. Tübingen: Narr.

Banguoğlu, T. ([5]1986). *Türkçenin Grameri.* Ankara: Türk Dil Kurumu Yayınları.

Barwise, J., Etchemendy, J. (2005). *Sprache, Beweis und Logik. Band 1,2.* Paderborn: Mentis.

Beliankou, A., Köhler, R., Naumann, S. (2014). Quantitative properties of argumentation motifs. In: Obradović, I., Kelih, E., Köhler, R. (eds.), *Methods and Applications of Quantitative Linguistics: 25-43.* Belgrade: Academic Mind.

Bortz, J., Lienert, G.A., Boehnke, K. (1990). *Verteilungsfreie Methoden in der Biostatistik.* Berlin-Heidelberg-New York: Springer.

Bradley, J.V. (1968). *Distribution-free statistical tests.* Englewood Cliffs: Prentice Hall.

Brownlee, K.A. (1960). *Statistical theory and methodology in science and engineering.* New York-London: Wiley.

Bunge, M. (1983). *Treatise on Basic Philosophy. Vol. 6: Epistemology & Methodology II: Understanding the World.* Dordrecht/Boston/Lancaster: Reidel.

Busemann, A. (1925). *Die Sprache der Jugend als Ausdruck der Entwicklungsrhythmik*. Jena: Fischer.

Bußmann, H. (1990). *Lexikon der Sprachwissenschaft*, 2nd ed. Stuttgart: Kröner.

Čech, R., Popescu, I.-I., Altmann, G. (2014). Descriptivity in special texts. *Glottometrics 29, 70-80.*

Darley, F.L., Sherman, D., Siegel, G.M. (1959). Scaling of abstractness of single words. *Journal of Speech- and Hearing Research 2, 161-167.*

DeVito, J.A. (1967). Levels of abstraction in spoken and written language. *Journal of Communication 17, 354-361*.

Ersen-Rasch, M.I. (2001). *Türkische Grammatik für Anfänger und Fortgeschrittene*. Ismaning–München: Hueber

Fischer, H. (1969). Entwicklung und Beurteilung des Stils. In: Kreuzer, H., Gunzenhäuser R. (eds.), *Mathematik und Dichtung: 171-183.* München: Nymphen burger.

Grotjahn, R. (1980). The theory of runs as an instrument of research in quantitative linguistics. In: Grotjahn, R. (ed.), *Glottometrika 2: 11-43*. Bochum: Brockmeyer.

Janda, L. (ed.) (2013). *Cognitive Linguistics: The Quantitative Turn. The Essential Reader*. Berlin/Boston: de Gruyter Mouton.

Kisro-Völker, S. (1984). On the measurement of abstractness in lexicon. In: Boy, J., Köhler, R. (eds.), *Glottometrika 6: 138-151*. Bochum: Brockmeyer.

Köhler, R. (1986). *Zur linguistischen Synergetik. Struktur und Dynamik der Lexik*. Bochum: Brockmeyer.

Köhler, R. (1995). *Bibliography of Quantitative Linguistics*. Amsterdam/ Philadelphia: Benjamins.

Köhler, R. (2005). Synergetic linguistics. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.), *Quantitative Linguistics. An International Handbook:760-774*. Berlin: de Gruyter.

Köhler, R. (2015). Linguistic motifs. In: Mikros, G., Mačutek, J. (eds.), *Sequences in Language and Text: 89-108.* Berlin/Boston: de Gruyter Mouton.

Köhler, R., Tuzzi, A. (2015). Linguistic modeling of sequential phenomena: The role of laws. In: Mikros, G., Mačutek, J. (eds.), *Sequences in Language and Text: 108-123.* Berlin/Boston: de Gruyter Mouton.

Krug, M.G. (2001). Frequency, iconicity, categorization: Evidence from emerging modals. In: J. Bybee, P. Hopper (eds.), *Frequency and the emergence of linguistic structure: 309-335*. Amsterdam/Philadelphia: Benjamins.

Löbner, S. (2003). *Semantik: eine Einführung*. Berlin: de Gruyter.

Mates, B. (1997). *Elementare Logik – Prädikatenlogik der ersten Stufe*. Göttingen: Vandenhoeck & Ruprecht.

Mikros, G. K., Mačutek, J. (eds.) (2015). *Sequences in Language and Text.* Berlin/Boston: de Gruyter Mouton.

Ord, J.K. (1952). *Families of frequency distributions*. London: Griffin.

Paivio, A. (1979). *Imagery and verbal processes*. New Jersey: Erbsbaum.

Pikas, A. (1966). *Abstraction and concept formation*. Cambridge, Mass.: Harvard UP.

Popescu, I.-I. et al. (2009). *Word frequency studies*. Berlin/New York: Mouton de Gruyter.

Popescu, I.-I., Čech, R., Altmann, G. (2013). Descriptivity in Slovak lyrics. *Glottotheory 4(1), 92-104*.

Popescu, I.-I., Lupea, M., Tatar, D., Altmann, G. (2015). *Quantitatve Analysis of Poetic Texts.* Berlin: de Gruyter Mouton.

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J, (1985). *A Comprehensive Grammar of the English Language.* London: Longman.

Salmon, W.C. (1983). *Logik.* Stuttgart: Reclam.

Sambor, J., Hammerl, R. (eds.) (1991). *Definitionsfolgen und Lexemnetze*. Lüdenscheid: RAM-Verlag

Schlissmann, A. (1948/49). Sprach- und Stilanalyse mit einem vereinfachten Aktionsquotienten. *Wiener Ze4itschrift für Philosophie, Psychologie und Pädagogik 2, 42-46.*

Skinner, B.F. (1939). The alliteration in Shakespeare's sonnets: A study in literary behavior. *Psychological Record 3, 186-192.*

Skinner, B.F. (1941). A quantitative estimate of certain types of sound-patterning in poetry. *The American Journal of Psychology 54, 64-79.*

Skinner, B.F. (1957). *Verbal Behavior*. Acton, Mass.: Copley Publishing Group.

Stachowski, M. (^{2}2009). *Gramatyka języka tureckiego w zarysie*, Kraków: Księgarnia Akademicka.

Swift, L.B. (1963). *A reference grammar of modern Turkish.* Bloomington: Indiana University.

Wildgen, W. (2005). Catastrophe theoretical models in semantics. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.), *Quantitative Linguistics. An Internation al Handbook: 410-423.* Berlin: de Gruyter.

Wildgen, W. (2002). Dynamical models of predication. *Sprachtypologie und Universalienforschung 55(4), 403-420.*

Wimmer, G., Altmann, G. (2005). Unified derivation of some linguistic laws. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.), *Quuantitative Linguistics. An International Handbook:791-807.* Berlin: de Gruyter.

Ziegler, A. (1998). Word class frequencies in Brazilian-Portuguese press texts. *Journal of Quantitative Linguistics 5(3), 269-280.*

Ziegler, A. (2001). Word class frequencies in Portuguese press texts. In: Uhlířová, L., et. al. (eds.), *Text as a linguistic paradigm: Levels, constituents, constructs. Festschrift in honour of Luděk Hřebíček: 294-312*. Trier: Wissenschaftlicher Verlag.

Ziegler, A., Best, K.-H., Altmann, G. (2002). Nominalstil. *ETC – Empirical Text and Culture Research 2, 75-85.*

Zörnig, P. (1984). The distribution of distances between like elements in a sequence, part I. In: Boy, J., Köhler, R. (eds.), *Glottometrika 6, 1-15;* part II. In: U. Rothe (ed.), *Glottometrika 7, 1-14*. Brockmeyer, Bochum.

Zörnig, P. (1987). A theory of distances between like elements in a sequence. In: I. Fickermann (ed.), *Glottometrika 8, 1-22*. Brockmeyer, Bochum.

Zörnig, P. (2010). Statistical simulation and the distribution of distances between identical elements in a random sequence. *Computational Statistics & Data Analysis 54, 2317-2327.*

Quirk, R., Greenbaum, S., Leech, G., Svartvik, J, (1985). *A Comprehensive Grammar of the English Language.* London: Longman.

Salmon, W.C. (1983). *Logik.* Stuttgart: Reclam.

Sambor, J., Hammerl, R. (eds.) (1991). *Definitionsfolgen und Lexemnetze*. Lüdenscheid: RAM-Verlag

Schlissmann, A. (1948/49). Sprach- und Stilanalyse mit einem vereinfachten Aktionsquotienten. *Wiener Ze4itschrift für Philosophie, Psychologie und Pädagogik 2, 42-46.*

Skinner, B.F. (1939). The alliteration in Shakespeare's sonnets: A study in literary behavior. *Psychological Record 3, 186-192.*

Skinner, B.F. (1941). A quantitative estimate of certain types of sound-patterning in poetry. *The American Journal of Psychology 54, 64-79.*

Skinner, B.F. (1957). *Verbal Behavior*. Acton, Mass.: Copley Publishing Group.

Stachowski, M. (^{2}2009). *Gramatyka języka tureckiego w zarysie*, Kraków: Księgarnia Akademicka.

Swift, L.B. (1963). *A reference grammar of modern Turkish.* Bloomington: Indiana University.

Wildgen, W. (2005). Catastrophe theoretical models in semantics. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.), *Quantitative Linguistics. An Internation al Handbook: 410-423.* Berlin: de Gruyter.

Wildgen, W. (2002). Dynamical models of predication. *Sprachtypologie und Universalienforschung 55(4), 403-420.*

Wimmer, G., Altmann, G. (2005). Unified derivation of some linguistic laws. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.), *Quuantitative Linguistics. An International Handbook:791-807.* Berlin: de Gruyter.

Ziegler, A. (1998). Word class frequencies in Brazilian-Portuguese press texts. *Journal of Quantitative Linguistics 5(3), 269-280.*

Ziegler, A. (2001). Word class frequencies in Portuguese press texts. In: Uhlířová, L., et. al. (eds.), *Text as a linguistic paradigm: Levels, constituents, constructs. Festschrift in honour of Luděk Hřebíček: 294-312*. Trier: Wissenschaftlicher Verlag.

Ziegler, A., Best, K.-H., Altmann, G. (2002). Nominalstil. *ETC – Empirical Text and Culture Research 2, 75-85.*

Zörnig, P. (1984). The distribution of distances between like elements in a sequence, part I. In: Boy, J., Köhler, R. (eds.), *Glottometrika 6, 1-15;* part II. In: U. Rothe (ed.), *Glottometrika 7, 1-14.* Brockmeyer, Bochum.

Zörnig, P. (1987). A theory of distances between like elements in a sequence. In: I. Fickermann (ed.), *Glottometrika 8, 1-22.* Brockmeyer, Bochum.

Zörnig, P. (2010). Statistical simulation and the distribution of distances between identical elements in a random sequence. *Computational Statistics & Data Analysis 54, 2317-2327.*

Appendix

ANV Sequences

Slovak

T 1: SME 17.04.2015, Napätie v Žiline sa skončilo. Teraz to začína vriet` inde.

[N,A,N,N,V,N,N,N,N,N,N,N,V,N,V,N,A,V,N,N,N,N,V,A,N,N,N,N,V,N,N,A,N,
N,V,A,N,V,N,A,N,N,V,V,A,N,V,N,N,A,N,N,N,A,A,N,N,V,N,N,A,N,N,N,A,N,
N,N,V,N,V,N,V,A,N,A,N,V,N,A,V,A,N,N,V,N,N,A,A,N,N,V,N,A,N,N,A,A,N,
A,N,V,N,N,V,N,N,N,N,N,V,N,N,N,N,N,A,V,N,A,N,N,N,N,A,N,N,N,V,N,V,A,
N,V,N,A,N,N,V,N,N,N,N,V,N,A,N,N,N,N,N,V,N,N,N,N,V,N,A,N,N,N,N,N,V,
N,V,A,N,N,V,V,A,N,N,N,N,A,N,N,V,N,V,A,N,N,N,N,N,N,A,N,V,N,V,N,A,N,
N,V]

T 2: SME 18.4. 2015: Liberáli sa do volieb chystajú sami, na kongrese o taktike nerozhodnú

[A,N,V,A,A,N,V,A,N,N,V,N,A,N,N,N,N,N,V,N,N,N,N,V,N,N,V,N,N,V,N,N,N,
V,A,N,N,V,N,N,V,N,N,N,N,V,N,A,N,N,V,V,N,V,N,V,V,N,N,V,V,V,N,V,A,N,
V,A,V,V,N,N,N,V,N,N,V,N,V,A,N,N,V,N,N,N,N,V,N,A,N,N,A,N,N,V,A,N,N,
N,V,V,N,V,A,N,V,A,N,A,N,V,N,N,V, N,N,N,A,N,V,N,V,A,N,A,V,N,N,A,N,V,
A,N,V,N,V,N,V,V,A,N,N,A,N,A,N,V,N,V,A,N,V,A,N,A,V,N,V,N,V,A,N,N,N,
V,N,V,A,N,N,A,A,N,A,N,A,N,V,N,V,N,V,A,A,N,N,V,N,V,V,A,N,A,N,N,V,N,
N,V,A,V,V,V,N,V,V,N,V,V,A,N,V,N,A,N,N,N,A,N,V,N,N,N,V,N,A,N,N,N,N,
V,A,N,V,N,N,V,A,N,V]

T 3: SME 18.4.2015 : Langerová jemnosť neskrýva, získala štyroch Andělov

[V,V,V,N,N,N,V,A,A,N,N,V,N,A,N,A,A,N,N,N,N,N,V,N,A,A,N,N,V,N,A,N,N,
A,N,A,N,N,N,N,V,A,N,A,N,V,V,N,N,V,V,N,N,N,N,V,V,N,V,N,V,N,N,V,N,N,
N,N,V,N,V,N,N,V,A,N,A,N,N,N,V,A,N,A,N,V,V,N,N,N,V,N,A,N,N,A,N,V,N,
N,N,V,A,A,A,V,N,A,N,A,N,V,A,N,V,N,N,N,V,A,N,V,N,N,N,V,N,N,N,N,N,V,
N,V,N,V,N,N,N,V,A,N,N,N,V,N,A,N,V,N,V,A,N,N,N,N,N,A,N,V,N,N,N,V,V,
A,N,N,V,N,N,V,N,N,N,V,N,V,N,A,V,N,A,V,V,N,N,V,N,N,V,N,A,N,V,A,A,N,
A,N,V,N, N,N,N,N,V,N,V,A,N,N,N,N,V,A,N,A,N,N,V,N,N,N,V,N,N,N,N,N,V,
V,V,N,V,N,V,N,N,A,N,A,N,N,V,N,V,N,N,N,N,N,V,N,N]

T 4: SME 19.4.2015: Grécko dúfa že ho zachránia miliardy z Ruska a Číny

[N,N,V,N,V,A,N,N,A,N,N,N,V,A,N,V,N,A,N,N,N,N,V,N,V,N,N,V,N,A,N,N,V,
A,N,N,V,N,N,N,N,N,V,N,N,A,N,N,N,N,V,A,N,N,N,V,N,V,N,N,V,A,N,N,N,V,
N,N,V,A,A,N,A,A,N,N,N,V,A,N,A,N,A,N,N,N,V,A,N,N,N,N,N,N,N,V,N,A,N,
A,N,N,A,N,V,N,N,V,N,A,N,N,N,V,N,N, N,N,V,N,N,V,N,V,N,V,N,N,A,N,A,N,

N,V,N,A,N,V,A,N,N,V,N,N,N,V,A,N,A,N,N,N,V,N,V,N,A,N,N,N,V,N,N,V,A,
N,N];

T 5: SME 19.4.2015: Nespokojní poštári demonštrovali pred ministerstvom
[N,N,N,N,A,N,V,N,N,A,N,V,N,N,N,N,N,A,A,N,A,A,N,V,N,N,N,N,V,A,N,N,N,
V,N,V,A,N,V,A,N,N,V,N,N,N,N,V,N,N,A,N,N,N,N,N,N,N,V,A,N,A,N,N,V,N,
V,A,N,N,A,A,V,N,V,N,N,V,N,N,V,A,N,V,N,V,N,V,N,V,V,A,V,A,N,V,A,N,N,
V,V,A,V,N,N,N,V,V,N,A,N,N,N,A,N,N,N,N,V,A,N,A,N,N,V,V,N,V,A,N,A,N,
N,N,V,N,N,N,N,V,N,V,N,N,N,N,A,N,V,A,N,V,A,N,V,N,N,V,N,V,A,N,A,N,V,
A,N,A,N,N,V,N,A,N,N,V,N,A,N,N.V,N,N,N,A,N,N,N,N,N,N,N,N,N,N,A,N,N,
N,V,A,A,N,V,N,V,A,N,N,A,A,N,N,N,N,V,N,N,N,N,N,V,N,V,A,N,N,A,N,V,A,
N,N,N,N,N,V,N,A,N,N,N,V,N,V,A,N,N,N,N,V,N]

Hungarian

T 1: Magyar Online 20.4. 2015
[N,N,V,N,A,N,N,A,N,N,N,N,N,A,N,V,N,N,N,V,N,N,N,N,A,N,V,N,A,N,A,N,V,
N,V,A,N,A,N,V,N,V,N,N,A,N,N,V,N,N,A,N,V,N,N,V,A,N,A,N,V,A,N,A,N,V,
A,N,A,N,N,V,A,N,V,N,N,V,A,N,A,A,N,N,N,V,N,N,N,A,N,V,N,V,N,N,N,N,V,
N,V,N,A,N,V,N,N,N,N,V,V,N,N,N,N,V,N,N,N,N,V,N,N,N,N,N,V,V,A,N,N,N,V,A,N,
V,N,V,N,V,V,A,N,N,V,A,N,A,N,V]

T 2: Magyar Online 20.4. 2015: A rendszerváltás gyermekei
[N,V,A,A,N,V,A,N,N,A,N,N,N,A,N,V,V,N,N,N,V,N,N,A,N,N,V,V,A,N,N,A,A,
V,A,N,N,N,V,N,A,N,A,N,N,A,N,A,N,N,A,N,N,N,N,A,N,V,A,N,A,N,A,N,N,N,
N,V,A,N,N,N,N,N,V,A,N,A,N,N,N,N,A,N,N,V,N,A,A,V,N,A,N,N,V,N,A,N,N,
N,N,V,A,N,N,N,N,N,V,N,N,N,V,A,N,A,N,N,N,V,N,N,V,N,A,N,N,A,N,N,N,V,
N,A,N,A,N,A,N,A,N,V,N,N,A,V,V,N,N,A,N,V,V,A,N,N,A,N,N,A,V,N,V,A,A,N,A,
A,A,A,N,A,N,A,N,A,N,N,N,V,N,N,A,A,A,N,N,A,A,N,N,A,A,N,A,N,A,A,N,N,
N]

T 3: Magyar Online 20.4. 2015: Kormányellenes tüntetést tartottak Budapesten
[N,N,N,A,A,N,N,N,A,N,A,V,N,V,N,N,V,N,V,N,N,V,A,N,A,N,V,N,A,N,N,V,N,
V,A,A,A,N,N,V,V,N,N,N,V,A,N,N,N,N,V,N,A,N,N,A,N,V,A,N,V,A,A,N,A,N,
V,N,N,N,A,N,N,V,N,V,N,A,N,A,N,A,N,N,V,A,N,V,A,N,A,A,N,N,N,V,N,N,N,
N,N,N,V,V,N,N,N,A,N,N,N,N,N,N,V,N,N,N,N,N,N,V,V,V,N,N,N,A,N,A,N,N,
V,N,V,N,V,N,N,V,N,N,V,N,N,N,V,V,V,N,N,N,V,N,A,N,N,A,N,V,N,V,N,V,N,
V,A,N,N,V,N,N,V,N,V,V,N,N,N,A,A,A,N,V,N,V,N,V,N,N,N,A,N,V,N,N,N,A,
N,N,N,V,N,N,A,N]

T 4: Magyar Online 20.4. 2015: A jövö héten melegszik az idö
[N,N,A,N,A,N,V,A,A,N,A,N,V,A,V,N,A,N,N,A,A,N,N,N,N,N,N,V,A,N,N,V,N,
N,A,N,N,N,N,V.N.N,V,N,N,N,A,N,A,N,A,N,N,N,N,V,N,N,A,N,N,V,N,N,A,N,
N,A,A,N,V,N,N,N,V,N,V,N,A,N,N,A,N,A,N,A,N,N,N,V,N,N,A,N,N,V,N,V,A,
A,N,A,N,N,N,N,V,A,N,N,V,A,N,A,N,V,N,A,N,N,V,N,N,V,N,N,A,N,N,N,V,A,
N,A,N,N,V,V,N,A,N,N,V,N,N,A,V,N,N,N,N,A,A,N,V,A,N,N,N,N,V,N,N,A,V,
N,A,N,N,V,N,V,N,A,N]

T 5: Magyar Online 20.4. 2015: Vasárnapi pihenönap - MSZP: a balatoni üzletek a szezonban vasárnap is lehessenek nyitva!
[N,V,N,V,N,A,N,N,A,A,N,V,A,N,A,N,N,N,N,V,N,V,N,A,N,V,V,A,A,N,A,N,N,
A,N,A,A,N,N,A,N,N,V,V,N,A,N,A,N,N,V,V,A,N,N,A,N,V,N,A,A,N,N,N,V,N,
A,N,V,N,A,N,N,V,N,N,V,A,N,N,N,V,A,A,N,V,N,V,A,N,A,V,N,N,A,V,A,N,A,
V,V,N,A,N,V,V,V,A,N,A,N,V,A,N,A,N,N,A,N,N,N,V,V,A,V,A,N,N,A,N,V,A,
N,V,A,N,V,N,A,N,A,N,N,N,A,N,N,V,A,N,A,N,A,V,N,V,A,N,N,N,N,A,N,V,N,
A,N,N,N,A,N,V,A,N,V,V,N,V,N,N,A,A,N,N,V,N,N,N,N,A,N,A,N,N,A,N,N,V,
A,N,N,A,N,N,A,A,N,N,N,A,N,A,N,V,A,N]

Croatian

All texts from http://www.jutarnji.hr
T 1: 17. April 2015, Tko je voda napada na kovača
[V,N,N,N,N,V,N,N,N,V,N,V,N,N,N,N,N,N,N,N,V,N,A,N,N,V,N,N,A,V,N,V,A,
N,N,N,N,V,N,N,V,N,N,N,N,N,N,A,N,V,N,V,V,V,N,N,V,N,V,V,V,N,V,N,V,N,
A,N,V,N,N,N,N,V,N,V,N,N,V,V,N,V,V,N,N,V,N,V,V,N,V,N,A,N,N,V,V,N,V,
N,N,N,N,N,N,N,N,V,N,N,N,N,N,N,N,N,N,N,N,N,N,N,N,N,N,N,N,V,V,N,V,V,A,
N,N,V,N,N,N,A,N,N,V,N,N,N,N]

T 2: 17. April 2015, Ugledni časopis The Economist objavio je tekst o 'balkanskim ratnicima u inozemstvu,'
[N,A,N,N,V,N,A,N,N,V,N,N,N,N,N,V,V,N,N,N,N,N,N,V,N,N,V,N,N,N,N,A,N,
N,N,N,N,V,N,N,N,V,N,V,N,N,N,N,N,V,V,V,A,N,N,N,N,N,V,N,N,N,N,A,N,V,
N,N,N,V,N,V,A,N,N,N,N,V,V,V,N,V,A,A,N,V,V,N,V,N,N,N,N,N,N,V,N,N,N,
N,N,N,N,V,N,N,N,N,V,N,V,N,N,V,V,N,N,N,N,N,N]

T 3: 17. April 2017, Doživjeli šol na sprovodu
[V,N,N,V,N,V,N,A,N,V,V,N,V,N,N,N,N,A,N,V,V,N,A,A,N,N,A,N,V,A,N,A,N.
N,A,N,N,V,N,V,N,N,N,N,V,N,V,A,N,N,N,V,N,N,N,A,N,N,V,V,V,N,A,N,N,N,
N,N,V,V,N,V,A,A,A,N,N,V,V,N,V,N,N,V,N,V,V,N,N,V,A,N,A,N,N,A,N,N,V,
V,N,V,N,V,A,N,N,V,N,N,N,A,N,V,V,A,N,N,N,N,V,A,N,A,A,N,V,V,N,N,N,V,
N,V,N,N,V,N,V,N,N,V,V,A,V,N,V,A,N,N,A,N,N,A,V,N,N,A,N,N,A,N,V,V,V,
A,N,N,N,N,N,V,N,N,A,A,N,V,N,N,V];

T 4: 20.April 2015, Grčić i Lalovac predstavili mjere ušteda

[N,N,V,N,N,V,A,N,N,V,N,N,N,V,A,N,N,N,N,N,N,N,A,N,N,A,N,N,N,N,N,N,A,
N,V,N,N,N,N,A,N,V,N,A,N,V,V,N,V,N,N,N,V,N,A,N,V,N,V,N,N,A,N,V,N,N,
N,A,A,N,A,N,V,N,N,N,V,N,V,V,A,N,N,V,N,V,A,N,A,N,A,N,N,N,N,V,N,N,V,
N,N,V,N,V,V,N,N,N,N,V,N,N,V,N,N,N,A,N,N,V,A,N,N,N,A,N,N,N,N,N,N,N,
V,A,N,A,N,N,N,N,N,V,N,N,N,A,N,V,N,N,A,N,A,N,N,V,A,N,N,N,N,N,N,N,N,
N,V,N,N,N,N,N,V,N,N,V,A,V,N,V,N,V,A,N,N,A,N,V,N,N,V,N,V,N,N,N,N,N,
A,N,N,N,N,N,V,A,N,N,N,N,N,N,A,N,N,V,N,V,N,N,V,N,N,N,V,N,N,V,A,N,N,
N,N,N,V,N,V,A,N,N,V,N,A,N,N,N,V,V,N,N,N,A,N,N,N,N,N,V,N,N,A,N,V,A,
N,N,N,V,N,A,N,V,N,A,N,V,A,N,A,N,V,N,V,A,N,V,N,V,N,N,V,A,N,V,V,A,N,
N,N,N,A,N,N,V,N,N,N,N,N,V,N,N,N,A,N,V,N,N,N]

T 5: 24. April 2015, Što za vas znači Amerika?

[V,N,V,A,N,V,N,N,V,A,N,N,N,N,A,N,N,N,N,V,N,V,A,N,N,N,A,A,N,N,N,N,A,
N,N,N,V,N,N,N,N,V,N,A,N,V,A,N,N,V,N,N,V,N,N,A,A,N,N,V,N,N,N,N,N,N,
A,N,N,N,N,V,V,A,N,N,N,V,N,N,A,A,N,V,V,N,V,A,N,V,N,A,N,N,N,N,N,A,N,
A,N,N,N,N,N,N,N,N,N,A,N,V,N,V,N,V,N,N,N,A,N,N,N,A,N,V,N,N,N,N,V,V,
N,V,V,V,V,V,N,N,V,N,V,V,V,N,N,N,N,V,N,V,N,N,N,V,A,A,N,N,N,V,N,N,N,
V,V,N,N,N,N,N,N,N,V,N,N,N,V,N,N,N,N,V,V,N,N,V,N,N,N,N,N,V,A,N,V,V,
N,**N,V,N,N,N,A,N,N,N,N,N,N,N,N,N,**N,V,A,N,N,N,N,A,N,N,N,N,N,A,N,N,V,
A,N,N,N,N,V,N,N,N,N,A,N,N,N,N,N,N,N]

Chinese

T 1: Multiple images of local officials - a hot issue in the political science in recent years (From Beijing Daily April 20, 2015)

[N,V,N,V,N,N,N,N,N,N,N,V,N,N,V,V,V,N,V,N,N,N,N,V,N,A,N,V,V,N,V,N,N,
N,V,V,N,N,N,N,N,V,N,V,N,N,V,N,N,V,A,N,N,N,N,N,V,V,V,N,V,V,N,N,A,N,
V,V,V,N,V,N,V,N,V,V,N,A,N,N,N,V,V,V,V,N,N,N,N,V,V,N,N,N,V,N,N,V,N,
N,V,V,N,A,V,A,N,A,N,V,V,N,N,A,V,A,V,A,V,V,V,V,N,N,N,N,V,V,V,A,N,A,
A,N,N,N,N,N,N,V,N,N,A,N,A,N,N,N,N,N,N,N,V,N,N,N,V,N,N,A,N,N,N,N,N,
N,V,N,V,N,V,V,N,V,N,N,V,N,N,N,V,N,A,N,V,V,V,V,N,N,V,N,V,V,N,N,N,A,
N,N,A,N,N,A,V,V,V,N,N,N,N,N,N,N,N,A,V,N,N,N,N,V,N,V,V,N,N,N,N,N,N,
N,N,N,V,N,N,N,N,N,V,N,V,N,V,V,N,N,N,N,V,V,N,V,A,N,V,A,A,V,V,A,A,V,
V,V,V,N,V,V,V,V,A,N,V,N,N,N,V,N,V,A,V,N,N,V,A,V,N,V,V,V,A,N,A,N,N,
A,N,N,V,V,V,N,V,V,N,V,V,N,N,V,V,N,N,N,V,V,A,N,N,V,N,N,V,N,V,N,N,N,
V,N,N,N,V,V,A,N,N,A,N,V,N,N,V,N,N,V,N,N,N,N,N,N,V,N,N,V,V,N,V,N,N,
N,V,N,N,N,V,V,N,V,V,V,A,V,V,N,V,V,V,V,A,V,N,V,N,N,N,V,A,N,N,V,A,N,
V,N,N,N,V,N,V,V,V,A,N,N,N,N,V,N,N,N,V,V,V,V,N,N,N,N,N,N,V,V,V,A,N,
N,N,N,V,N,V,V,V,N,V,N,V,V,N,N,N,A,V,N,N,V,N,N,V,A,V,N,N,V,A,N,N,V,
N,A,N,N,N,N,N,N,N,V,A,N,N,V,N,N,N,N,V,V,N,N,N,V,V,N,N,N,N,V,N,V,N,
V,V,N,N,V,N,N,N,N,V,V,N,N,N,V,V,V,V,N,N,N,N,N,N,V,V,N,N,N,N,V,V,V,
V,N,N,V,N,N,N,N,N,N,N,N,V,N,N,N,V,N,N,V,N,V,N,V,N,N,V,V,V,V,N,V,N,

V,V,V,N,V,N,N,V,V,N,N,V,V,N,V,N,N,V,N,V,N,V,N,N,V,V,N,N,V,A,N,V,A,
V,N,N,N,V,A,N,N,V,N,N,N,N,N,N,A,N,A,N,A,V,N,V,N,V,N]

T 2: Can Internet+ make the wedding consumption more transparent? (From Consumer Daily April 15, 2015)

[N,V,V,N,V,V,A,N,N,V,A,N,V,V,N,V,N,V,N,A,V,N,V,N,N,N,A,N,N,V,A,N,N,
V,N,V,N,V,N,N,V,N,N,N,V,V,V,V,V,V,N,V,V,V,A,N,N,V,A,N,N,V,V,N,N,N,
V,N,V,N,V,A,N,A,N,V,N,N,V,V,N,V,V,N,V,N,A,V,N,V,V,A,V,V,V,N,A,V,V,
V,N,V,V,V,A,N,N,N,V,V,V,V,N,N,N,V,N,A,A,V,V,N,V,N,V,N,V,V,N,V,V,V,
V,N,A,N,A,N,V,V,N,V,V,V,N,V,N,V,N,N,V,V,A,V,N,N,V,N,N,V,N,V,V,N,V,
A,N,N,N,N,V,V,A,V,N,N,N,V,A,N,N,N,V,N,N,V,V,V,V,V,N,V,N,V,N,N,V,N,
N,N,V,N,V,V,N,V,V,V,N,V,V,N,V,V,V,N,V,N,V,N,N,V,V,N,V,N,V,N,V,N,V,
A,V,V,V,V,V,N,A,V,N,N,V,N,V,N,N,V,V,V,N,N,N,N,V,N,V,A,N,N,N,V,V,N,
V,N,N,V,N,N,V,V,N,V,N,V,V,N,V,A,V,N,V,N,V,A,A,N,N,N,N,N,N,V,N,V,A,
N,N,V,N,V,V,V,V,N,N,V,V,N,N,V,V,N,V,N,N,V,N,V,N,N,V,N,A,N,V,N,N,V,
V,N,N,N,V,V,A,N,V,V,N,V,V,N,A,N,A,V,N,N,N,N,N,V,N,A,N,V,V,N,N,N,V,
N,N,N,V,N,A,N,V,V,A,V,N,V,N,N,V,N,V,N,N,V,V,V,N,N,N,N,V,V,N,V,N,N,
V,N,N,V,V,V,N,A,V,V,A,N,V,N,N,V,N,N,V,N,V,N,V,N,V,N,V,N,N,V,N,N,V,
N,N,N,N,V,N,V,N,V,N,N,V,V,V,V,N,N,V,V,N,V,N,V,V,V,A,N,N,V,V,N,V,N,
V,N,N,V,V,V,V,V,N,N,N,V,A,N,N,N,N,N,N,V,V,N,V,N,V,N,V,V,V,V,N,V,N,
A,A,V,V,A,V,V,N,N,N,V,V,N,A,V,V,A,V,N,N,V,N,N,N,V,N,V,V,N,N,N,V,V,
N,V,V,N,V,N,V,N,A,V,A,V,V,N,V,V,N,A,A,V,N,N,V,N,A,V,V,N,N,N,V,N,V,
N,V,V,V,A,V,V,N,V,V,V,A,N,V,N,V,V,A,V,N,V,A,N,N,N,N,N,V,V,N,N,A,V,
N,V,N,N,N,V,N,V,N,V,N,N,A,N,A,N,V,N,V,N,V,V,N,V,N,V,N,V,N,N,N,V,N,
N,N,V,V,N,N,N,N,V,N,A,V,V,N,N,N,V,V,V,V,N,V,N,V,N,N,N,A,V,N,N,N,V,
V,A,N,N,V,N,A,V,V,V,N,N,N,N,V,V,N,V,N,V,N,V,N,N,V,A,V,N,A,V,V,V,V,
V,N,V,V,N,N,N,N,V,N,V,V,N,V,N,V,N,N,N,N,V,N,V,A,N,V,N,N,N,V,N,N,V,
V,N,A,V,N,N,V,N,A,N,N,V,A,V,N,N,V,V,V,V,V,V,V,V,V,V,N,V,V,N,N,V,
N,V,N,N,V,N,V,N,V,A,N,V,N,V,N,V,V,V,V,A,V,N,N,N,V,N,N,N,N,N,V,N,V,
N,N,N,N,V,N,N,N,N,V,V,N,V,N,V,N,V,N,V,V,V,N,V,A,V,V,N,V,V,N,V,V,V,
N,N,N,A,V,V,N,V,V,N,N,V,N,N,A,N,V,V,N,V,N,V,A,N,V,V,N,V,N,V,V,N,V,
A,V,V,V,V,N,V,V,N,N,V,N,V,N,N,N,N,A,V,V,N,V,N,V,N,N,A,N,N,V,N,V,V,
V,N,N,N,N,N,N,V,N,N,N,V,A,N,V,V,V,N,N,N,A,N,N,N,A,V,N,V,V,N,N,N,N,A,
N,V,V,N,V,V,N,A,N,V,V,N,A,N,N,V,V,N,N,N,N,N,V,V,A,V,N,N,N,N,A,V,V,
V,N,A,V,N,N,N,V,N,N,V,V,V,A,N,N,V,V,V,N,N,N,V,V,V,N,N,V,A,V,A,N,N,
N,V,N,N,N,N,V,N,V,V,N,N,V,N,V,V,A,N,V,V,N,N,N,V,V,V,V,V,N,N,V,V,N,
A,N,V,N,V,V,N,N,N,V,V,N,V,V,A,N,V,N,N,V,V,N,N,V,V,A,N]

T 3: Those people who illegally built shantytowns in the air become power-less - "limit down" verdict will be issued today, and demolitions will be started in mid May (From Beijing Daily April 17, 2015)

[V,V,V,V,N,N,V,V,V,V,N,N,V,N,N,N,N,N,V,V,V,N,V,V,N,V,N,N,N,N,N,N,N,
V,A,N,V,N,N,V,N,V,N,V,V,N,V,V,V,V,V,N,N,V,N,V,V,A,N,N,V,N,V,V,V,N,
V,V,V,N,N,V,V,N,V,V,V,V,V,V,N,N,V,N,V,V,V,A,N,V,N,V,N,N,N,V,N,V,V,
N,N,N,N,V,N,N,N,N,V,V,N,V,N,V,N,V,N,V,V,V,N,V,A,V,V,N,V,V,N,V,V,V,
N,N,N,A,V,V,N,V,V,N,N,V,N,N,A,N,V,V,N,V,N,V,A,N,V,V,N,V,N,V,V,N,V,
A,V,V,V,V,N,V,V,N,N,V,N,V,N,N,N,N,A,V,V,N,V,N,V,N,N,A,N,N,V,N,V,V,
V,N,N,N,N,N,N,V,N,N,N,V,A,N,V,V,V,N,N,N,A,N,N,N,A,V,N,V,V,N,N,N,N,A,
N,V,V,N,V,V,N,A,N,V,V,N,A,N,N,V,V,N,N,N,N,N,V,V,A,V,N,N,N,N,A,V,V,
V,N,A,V,N,N,N,V,N,N,V,V,V,A,N,N,V,V,V,N,N,N,V,V,V,N,N,V,A,V,A,N,N,
N,V,N,N,N,N,V,N,V,V,N,N,V,N,V,V,A,N,V,V,N,N,N,V,V,V,V,V,N,N,V,V,N,
A,N,V,N,V,V,N,N,N,V,V,N,V,V,A,N,V,N,N,V,V,N,N,V,V,A,N]

A,N,V,V,V,V,V,N,V,N,V,N,V,N,N,V,V,N,V,V,V,V,N,N,N,N,N,N,N,V,N,V,V,
V,V,N,V,N,N,N,N,V,N,N,N,N,N,V,N,V,N,N,N,V,N,N,N,V,V,N,V,N,A,N,N,N,
V,V,V,N,V,N,V,A,V,V,N,N,N,N,V,V,V,N,V,V,A,N,V,V,V,V,V,N,A,V,N,V,V,
N,N,N,N,N,N,N,N,N,N,V,A,N,N,N,N,V,V,N,N,N,N,N,N,N,N,V,V,N,N,N,N,N,
N,N,N,N,N,N,N,N,N,V,N,V,N,V,N,A,V,V,V,V,N,N,V,N,N,N,V,N,N,N,V,V,V,V,
V,V,N,V,V,V,V,N,A,V,N,V,V,N,N,V,V,N,V,N,N,V,N,N,N,V,V,N,V,N,V,V,N,
N,N,N,V,N,V,V,N,V,V,N,N,V,N,V,N,V,V,N,V,V,N,V,N,V,V,N,V,V,V,V,V,V,
V,N,V,V,V,N,N,N,N,V,N,V,V,N,N,V,V,V,N,V,N,N,V,N,V,V,N,N,N,V,V,N,V,
V,V,N,N,V,V,V,V,V,N,V,N,N,V,A,V,N,N,V,V,N,N,V,V,V,V,V,V,N,N,V,V,N,
V,V,V,N,V,N,N,N,V,V,N,V,N,V,V,N,N,V,N,V,V,N,V,V,V,N,N,N,V,A,V,V,N,
N,V,V,V,V,N,A,V,N,N,V,N,V,V,V,N,N,V,V,N,V,V,N,N,N,V,N,V,N,N,V,N,V,
N,N,N,N,V,V,V,V,N,V,N,V,N,N,N,V,N,N,N,V,N,V,V,V,N,N,V,N,N,N,V,N,V,
V,V,N,N,N,N,V,V,V,N,V,N,V,N,V,N,A,V,N,V,N,V,N,V,V,N,V,V,V,N,V,N,N,
N,A,V,V,N,V,V,V,V,V,N,N,N,N,N,V,V,V,N,N,V,V,N,V,V,N,N,V,N,V,N,V,N,
N,V,V,N,A,N,N,V,N,V,N,V,N,V,N,V,V,N,N,V,V,N,V,N,V,V,V,N,N,V,V,V,V,
N,V,N,V,V,N,V,N,V,V,N,V,V,V,N,N,N,N,N,V,N,N,N,N,N,N,N,N,N,N,N,N,N,
N,N,V,V,N,V,V,V,N,V,N,A,V,V,A,N,N,N,V,V,V,N,N,N,N,N,N,N,V,V,V,V,N,
N,A,A,N,N,N,V,N,V,V,A,N,V,V,V,V,N,N,N,N,V,A,N,V,V,N,V,V,A,N,N,V,A,
N,N,V,V,V,N,V,V,V,V,N,V,V,N,V,V,V,N,V,N,A,V,N,N,V,N,N,N,N,N,V,V,N,
N,V,N,N,V,N,V,V,V,N,V,V,V,N,V,V,V,V,V,V,V,V,V,N,V,N,V,V,V,N,N,N,V,N,
V,N,V,N,V,N,N,V,V,N,N,N,V,N,N,N,V,V,N,N,A,V,V,A,N,V,V,V,V,V,N,A,A,
N,A,N,V,V,V,A,N,V,N,V,N,V,N,V,V,A,N,V,V,N,V,N,N,V,N,V,N,V,V,V,V,V,
N,V,N,V,N,V,V,N,V,V,V,V,N,N,V,N,N,N,V,N,V,V,N,V,V,N,V,V,N,V,V,V,V,
V,V,N,N,V,V,N,V,V,V,V,V,V,N,N,V,

T 4: The reforms of the burial customs in Hainan: the comfort for the living people and the dignity for the dead (From China Society April 14, 2015)

[N,N,V,V,V,V,V,V,N,N,N,N,N,N,V,V,V,A,V,N,V,N,V,V,N,N,N,N,V,N,V,V,V,
N,N,V,N,V,V,N,N,N,N,V,V,N,N,V,N,V,V,N,A,N,V,N,N,V,V,N,N,V,V,A,V,V,
N,N,N,V,V,N,N,N,N,V,V,V,A,V,A,N,N,N,V,N,N,V,N,N,V,V,N,N,V,N,N,N,N,N,
V,N,A,V,N,N,V,N,N,N,N,N,N,V,N,N,N,V,N,N,N,N,N,A,N,V,A,N,N,N,N,A,A,N,
V,N,N,N,N,N,V,A,V,V,N,N,V,N,N,N,N,N,N,N,N,N,N,N,N,N,N,N,N,V,N,N,N,V,
V,N,N,V,V,V,V,N,V,N,V,N,A,A,V,N,N,V,N,V,N,V,N,N,N,V,N,V,V,N,V,V,N,
V,N,V,V,V,V,N,V,A,V,A,V,V,V,V,V,N,V,N,N,V,N,V,V,N,N,N,N,N,V,A,N,A,
V,V,N,N,V,N,A,V,N,V,N,V,V,N,V,N,V,V,V,A,N,V,V,N,N,V,V,N,V,A,N,V,V,
N,N,V,N,N,V,V,N,N,N,V,V,V,N,N,V,N,N,N,N,V,N,N,N,N,V,N,N,V,N,V,V,N,V,V,
N,A,N,V,N,N,V,N,N,V,N,N,V,V,V,N,V,N,N,N,A,V,N,V,A,N,N,V,V,A,V,N,N,
N,N,V,A,N,V,V,N,N,N,V,A,V,N,V,V,N,N,V,V,N,V,N,V,N,N,N,V,V,N,A,N,V,
V,V,V,N,N,V,V,A,N,N,N,V,N,V,N,A,V,V,N,N,N,N,N,N,N,N,V,V,N,N,N,V,V,
N,N,V,V,N,V,V,N,V,N,N,N,V,A,N,V,N,N,N,V,N,N,V,N,V,V,N,N,V,A,A,N,V,
V,V,N,V,A,N,N,N,N,V,N,V,V,N,A,V,N,N,N,V,V,V,V,V,N,V,N,N,V,N,N,N,A,
V,V,V,V,N,A,N,V,N,V,N,V,A,N,A,V,V,N,A,N,N,V,V,N,V,N,N,V,V,V,V,N,V,
N,V,V,V,V,N,V,N,V,N,N,N,N,A,N,V,N,V,N,N,V,N,V,V,N,N,N,V,N,V,N,N,N,
N,N,N,N,A,N,N,N,V,N,V,N,N,N,V,N,N,N,V,N,V,V,N,V,V,N,V,V,N,V,N,N,N,

V,V,N,V,V,N,N,N,V,N,V,V,V,V,N,N,V,N,V,V,N,V,N,N,N,V,N,V,N,N,N,N,N,
N,V,N,N,V,N,N,V,N,V,V,V,N,N,N,N,N,V,V,V,V,V,N,N,V,V,N,V,N,N,V,N,V,
N,N,V,A,A,N,N,V,N,N,N,V,N,N,N,V,N,N,V,V,V,N,N,N,V,N,N,N,V,A,N,N,V,
V,V,N,N,V,N,V,V,N,N,A,V,V,V,V,V,N,N,N,V,N,N,N,V,V,N,V,N,N,N,N,N,N,
V,V,A,V,N,V,V,N,V,N,N,V,V,A,N,V,V,V,A,V,V,N,N,V,N,N,V,A,V,N,V,V,V,
N,V,N,N,N,V,V,V,N,N,N,N,N,V,V,A,V,N,N,N,N,V,N,N,N,N,N,N,N,V,N,N,
V,N,N,N,N,N,N,N,N,N,N,N,N,N,V,N,V,N,N,N,V,V,N,N,V,V,A,N,N,A,N,N,V,
V,N,N,V,V,N,N,N,N,N,A,A,A,A,N,V,V,V,N,V,N,V,V,V,V,V,N,N,V,N,N,N,A,
V,V,N,N,N,V,A,V,N,N,N,N,V,N,N,A,V,N,V,N,V,N,V,N,N,A,A,V,V,A,N,V,N,
V,V,N,N,V,N,V,N,N,N,A,A,V,N,V,A,V,V,V,N,N,V,V,N,A,V,A,N,N,V,A,V,N,
N,V,N,N,V,V,N,V,N,N,V,N,V,N,N,V,N,N,V,V,N,V,N,N,V,N,V,N,N,V,V,N,V,
N,A,V,N,V,N,V,N,V,N,V,N,N,V,N,N,V,N,N,N,V,V,N,A,N]

T 5: The Central Bank of China has decided to drop the deposit reserve of residents for 1 percentage point, with the aim to stabilize the economic growth (From Beijing Daily April 20, 2015)

[N,V,N,V,V,V,N,V,N,N,N,N,N,V,N,N,N,V,V,N,N,N,N,N,N,V,N,N,V,V,N,V,
A,A,N,N,A,N,V,V,N,V,N,V,N,N,V,V,N,V,V,N,N,V,V,N,V,N,V,V,N,V,V,V,N,
N,A,N,N,V,N,V,A,N,V,V,V,V,V,A,V,V,V,V,V,V,V,N,N,V,V,N,N,N,V,A,V,N,
N,V,V,N,N,V,V,N,V,N,N,V,V,A,V,V,V,V,N,V,V,N,V,N,N,N,V,N,V,V,N,N,V,
N,N,V,N,N,N,N,N,V,N,N,N,A,V,N,V,N,N,N,N,N,N,N,V,N,V,N,N,N,V,A,V,N,
N,A,A,N,N,V,N,V,N,N,N,N,V,V,N,N,V,N,N,N,N,V,V,N,V,N,V,N,N,V,N,V,V,
N,N,N,V,N,N,V,N,N,N,V,N,N,V,N,V,N,N,N,V,N,N,N,V,A,V,N,N,N,A,V,N,V,
V,V,N,V,N,N,V,N,V,V,N,N,V,N,A,A,N,V,N,V,V,V,V,V,N,V,N,N,N,V,N,N,N,
V,V,V,V,V,N,A,V,V,N,V,V,N,V,N,A,V,V,V,V,N,N,V,N,N,V,N,N,N,N,N,N,N,
V,N,N,N,N,V,V,V,N,V,N,N,N,N,V,N,N,N,N,V,V,N,N,V,N,N,V,V,N,V,N,V,V,
N,V,V,V,N,V,N,N,N,N,N,V,V,N,N,V,V,N,V,N,V,N,V,V,N,N,V,N,N,N,V,N,V,
V,N,V,N,N,N,V,V,V,N,V,A,V,V,V,V,V,N,V,V,V,A,V,V,V,V,N,V,N,V,N,N,V,N,
V,N,N,V,V,N,V,V,N,N,N,V,V,N,N,V,V,V,N,V,V,N,N,N,V,N,V,V,V,V,N,N,N,
V,N,N,N,V,N,V,V,N,V,A,V,V,N,N,N,V,A,V,N,V,N,V,N,A,N,N,N,N,V,N,N,N,
V,N,V,V,N,N,V,V,V,A,V,N,N,N,V,N,V,A,V,V,N,V,N,V,V,V,V,N,V,V,N,N,A,
N,V,V,N,V,V,A,V,A,V,V,V,A,V,A,V,V,V,A,V,A,V,N,V,N,V,N,N,V,N,V,V,V,
V,N,N,V,V,N,N,N,N,V,V,V,N,N,V,N,V,V,N,V,N,N,V,N,N,N,V,V,V,V,N,N,N,
V,V,V,V,A,V,N,V,V,V,N,N,A,V,V,N,N,N,N,V,V,V,V,N,N,V,V,V,V,N,V,V,V,
N,V,A,V,V,V,N,V,N,V,V,N,V,V,A,V,V,A,V,V,V,N,A,N,V,V,V,A,N,V,N,N,N,
N,N,N,V,N,N,V,V,N,V,N,N,N,V,N,V,N,V,N,V,N,V,V,N,V,N,V,N,N,V,N,V,A,
V,A,V,V,V,N,N,N,V,V,V,N,V,N,N,N,V,V,V,V,V,V,N,V,N,V,V,N,V,V,V,V,N,
V,A,V,N,V,N,N,V,V,V,N,A,V,N,N,N,V,N,N,V,N,V,N,N,V,V,A,V,N,V,V,N,N,
N,V,V,N,V,N,N,V,N,V,N,V,N,V,N,N,V,V,N,A,V,N,V,A,V,N,N,N,A,N,A,N,V,
N,V,V,V,N]

Persian

T 1: 04.04.2015
Brzezinski warned republicans criticizing Lausanne-agreement
. هشدار برژینسکی به جمهوری‌خواهان منتقد تفاهم لوزان

[N,N,N,A,N,N,N,N,N,N,V,N,A,N,A,N,N,N,N,A,N,N,N,N,V,N,N,N,N,V,N,N,N,A,V
,N,N,N,N,N,N,N,A,N,N,N,N,N,N,N,N,N,N,V,N,N,N,V,V,N,A,N,N,N,N,N,A,N,N,
N,V,A,V,V,N,A,V,V,N,N,A,V,N,N,N,A,N,V,N,A,N,N,A,V,N,N,A,A,V,N,N,N,N,N,
N,V,N,A,N,A,N,N,N,V,V,N,N,A,N,A,N,A,V,N,N,N,V,V,N,A,V,N,A,V,V,N,N,N,N,V
,N,A,V,V,V,N,N,A,N,N,A,N,A,N,N,N,A,V,A,V,V,N,A,N,A,N,N,A,V,N,A,N,A,A,A,
N,A,A,N,V,N,V,N,N,A,V,N,A,N,N,V,A,V,V,N,N,N,V,A,A,V,N,V,V,N,N,V,N,N,A,N
,N,N,V,N,A,N,N,V,N,V,N,V,N,N,A,V,N,V,N,A,V,N,A,V,A,V,N,N,V,N,N,N,V,N,N,
N,N,V,N,A,N,V,V,N,A,N,N,N,N,V,V,N,N,N,A,A,N,A,V,N,N,N,V,N,N,V,N,A,V,N,
N,N,A,V,V,N,V,N,V,N,V,N,V,V,N,N,V,N,N,N,N,N,N,A,N,V,N,N,N,N,N,N,N,N,N,
N,V,N,N,N,A,N,N,A,A,V,N,N,N,V,N,A,N,N,N,A,N,A,N,V,N,A,A,N,N,A,V,V,N,
A,N,A,N,N,N,N,N,A,V,N,N,A,N,A,V,N,A,N,N,A,V,N,N,V,N,N,N,N,N,N,A,A,N,V
,N,N,N,N,N,N,N,N,A,N,N,N,N,N,N,N,A,N,V,N,N,N,N,N,N,N,N,N,N,N,V,N,N,N,
N,V,N,N,N,A,N,N,N,N,N,N,V,N,N,N,A,N,N,V,V,N,V,N,N,N,N,N,N,N,V,N,N,N,N
,N,A,A,N,N,N,N,A,A,N,A,N,N,A,N,A,N,V,V,A,N,N,N,A,N,N,V,N,N,N,N,A,N,V,
N,N,N,A,N,N,N,N,V,A,N,A,A,N,V,N,N,N,A,V,N,N,A,N,A,V,N,A,V,A,N,N,N,N,A
,V,N,N,V,N,N,N,A,V,N,V,N,A,A,A,V,N,A,N,N,A,N,N,N,V,N,N,N,N,V,N,N,N,N,N
,N,N,N,A,N,A,N,V,N,A,N,N,N,N,N,A,V,N,N,A,N,V,N,N,N,N,N,N,A,A,N,V,A,N,
N,A,V,N,A,N,N,N,V,N,N,N,N,A,N,V,N,N,N,N,A,N,N,N,N,N,A,N,N,N,A,N,N,A,
N,V,N,N,A,N,N,N,A,N,V,V,A,N,A,N,N,A,N,N,V,V,N,N,N,N,V,A,N,A,V,A,N,A,N,
A,N,A,N,N,N,N,N,N,N,N,V,A,N,N,N,N,A,N,N,N,V,N,A,A,N,V,N,N,N,N,N,N,V,A,
N,N,N,N,N,N,A,N,V,N,N,A,N,N,N,A,N,A,N,V,N,N,A,N,A,N,N,N,A,N,N,N,N,N,
N,N,V,N,A,N,N,N,N,N,N,N,N,N,V,N,V,N,V,N,A,A,N,V,N,N,N,N,A,V,N,N,N,N,N,
N,A,N,N,N,N,N,N,N,N,A,N,A,V]

T 2: 14.04.2015
Successful diplomacy was indebted to empathy and compassion of public and authorities
2. موفقیت دیپلماسی مرهون همدلی و همزبانی مردم و مسئولان بود

[N,N,N,A,N,A,N,N,N,N,N,A,N,N,N,N,N,N,N,N,A,N,N,N,N,V,N,N,N,N,N,N,N,N,
N,N,N,N,N,N,N,A,N,N,N,N,N,V,A,N,N,N,N,N,N,V,N,N,N,N,A,V,N,N,N,N,V,N,N,
A,N,N,A,N,N,N,A,V,A,V,N,A,N,N,N,A,N,N,A,V,N,N,N,N,N,A,N,V,N,A,A,N,N,A,
N,N,N,N,A,N,V,N,N,A,A,N,N,N,N,N,N,N,V,N,N,N,N,A,A,A,V,V,N,N,A,N,N,A,N,
A,N,A,A,N,N,A,N,N,V,N,V,A,N,N,N,N,N,N,A,V,N,N,N,N,A,N,V,N,N,N,N,N,A,N,
N,A,V,A,V,N,N,N,V,A,N,A,N,V,N,N,A,V,N,A,V,N,N,N,A,A,N,N,A,N,N,A,N,A,V,
V,N,N,N,A,N,N,A,N,N,N,A,A,N,N,N,N,A,V,N,N,N,V,N,A,N,A,N,N,A,V,N,N,N,V,
N,A,N,V,N,N,A,V,N,N,N,V,N,N,A,N,A,N,N,N,N,N,N,V,V,A,V,N,N,N,N,N,N,N,V,
N,N,A,N,A,V,N,N,A,N,A,A,N,N,A,N,A,A,N,N,N,N,N,A,V,V,A,N,N,A,N,N,A,N,N

,V,N,A,A,N,N,A,N,N,V,V,A,N,A,N,N,V,N,N,A,N,V,N,A,N,A,V,N,N,A,N,N,V,N,N,
A,N,N,A,N,N,A,N,V,V,N,A,A,N,A,N,A,N,N,N,N,V,N,A,N,A,A,N,A,V,V,N,A,N,N,
A,N,N,V,N,V,N,A,A,N,V,A,V,A,N,N,V,V,N,N,V,N,A,N,N,V,N,A,N,N,A,V,A,N,N,V
,V,N,N,A,A,N,A,V,N,A,A,N,V,N,N,N,A,N,V,A,N,N,A,N,V,V,N,A,N,V,A,N,N,A,V,
N,N,V,N,V,A,N,N,N,N,N,N,N,N,N,N,N,N,V,N,N,N,N,V,A,N,N,V,V,N,A,N,V,V,N,
A,N,A,V,N,V,N,N,N,A,N,N,A,N,A,V,N,A,A,N,V,N,V,N,N,V,N,A,N,A,A,V,N,N,A,
N,N,A,A,N,A,N,V,N,V,N,V,A,V,N,N,N,V,N,N,A,V,A,N,N,V,N,V,V,N,A,N,N,A,A,
A,A,V,N,A,N,V,N,V,A,N,N,A,N,N,V,N,N,A,V,N,V,N,V,N,A,A,V,N,N,V,N,A,A,N,
A,V,A,N,A,N,N,A,V,V,N,N,V,N,A,N,N,A,N,A,N,V,N,N,A,V,N,A,V,V,N,A,V]

T 3: 08.04.2015
Technical options of comprehensive agreement have been found
3. گزینه‌های فنی توافق جامع مشخص شده است

[N,N,N,N,N,N,N,N,N,N,N,N,A,A,N,A,N,N,N,N,N,N,N,V,N,N,N,N,N,N,N,N,N,N,
N,N,A,N,N,A,N,N,N,N,N,N,N,A,N,V,N,N,A,N,V,N,A,A,A,N,N,A,V,N,A,N,N,A,N
,N,N,A,A,N,N,N,A,N,A,A,N,N,N,N,V,A,V,N,V,N,N,A,N,N,N,N,N,A,N,N,A,N,V,N
,N,N,A,N,V,A,N,A,N,A,V,N,N,N,N,N,A,N,V,N,A,N,N,N,A,N,A,N,V,N,A,A,A,N,N
,N,N,V,N,N,A,A,N,N,N,N,N,N,N,N,N,N,A,V,N,A,V,N,N,N,N,N,V,N,A,N,A,A,V,N
,A,A,N,N,V,V,A,N,V,N,N,N,A,A,A,N,A,N,N,A,A,N,N,N,V,N,V,N,A,V,N,A,N,N,A,
N,V,A,N,A,N,A,N,A,N,V,N,N,N,A,N,N,A,V,V,N,A,A,N,A,V,N,N,N,N,N,A,N,V,N,
A,V,N,N,A,N,V,N,N,A,A,N,N,N,N,N,N,N,A,N,N,V,N,N,N,N,N,A,N,N,N,N,V,A,N,
N,N,A,N,A,N,N,A,N,N,N,N,N,N,A,N,N,A,N,A,N,A,N,V,N,N,N,V,N,N,N,A,N,A,
N,A,N,N,N,N,N,N,A,N,V,N,N,N,N,A,N,N,N,A,N,A,N,N,N,N,N,N,A,A,N,A,N,A,
V,N,N,A,N,N,V,N,N,N,N,N,N,N,A,V,V,N,N,A,N,N,N,A,N,N,N,N,A,N,V,N,N,A,V,
N,N,N,N,N,N,N,A,N,A,N,N,N,N,N,A,V,V,N,N,A,N,N,A,N,V,N,N,N,A,N,N,N,V,N,
N,N,V,N,N,A,V,N,A,N,N,N,N,V,V,N,N,N,N,V,N,N,N,N,A,N,A,V,N,N,A,N,A,N,N,
N,V,N,A,V,N,N,N,N,N,V,N,N,N,N,V,V,N,N,N,N,N,N,N,A,A,N,N,N,A,A,V,N,N,N,
A,N,A,N,A,N,N,N,N,A,N,V,N,N,N,A,N,A,N,N,A,V,N,A,N,N,N,N,A,V,N,N,N,N,N
,N,N,A,A,A,N,N,V,N,A,N,N,N,A,N,A,V,N,V,N,A,N,V,N,N,N,N,N,V,A,N,N,N,N
,N,N,N,N,N,N,V,N,N,N,N,N,N,N,N,N,V,V,A,N,N,N,A,N,A,A,V]

T 4: 02.04.2015
Delegates' defense of teachers rights
4. دفاع نمایندگان از حقوق معلمان

[N,A,N,N,N,N,N,N,N,N,N,V,N,N,N,N,N,N,N,N,N,N,N,N,V,N,A,N,V,N,N,N,N,A,
N,N,N,N,N,N,N,N,N,N,N,V,A,N,N,N,A,V,N,N,V,V,N,N,N,V,V,V,A,N,N,V,N,N,N,N,
N,N,V,V,N,V,N,N,N,N,N,N,N,N,N,A,A,N,N,N,V,N,N,N,N,V,N,N,N,N,A,N,N,N,N,
A,N,V,N,N,A,N,A,A,V,N,A,N,A,N,V,A,N,N,V,N,N,N,A,N,V,N,N,N,A,A,N,V,N,V,
V,N,N,A,V,A,N,A,N,A,V,N,A,V,N,N,N,N,V,N,N,N,V,N,N,N,N,V,N,A,A,N,N,A,V,
V,A,N,N,A,A,N,V,N,A,V,N,N,N,N,N,V,N,N,V,N,N,V,N,N,N,V,N,V,A,V,N,N,V,N,N
,N,N,N,V,N,N,A,N,N,N,N,N,N,A,V,N,V,N,N,N,N,N,N,N,A,A,N,N,N,V,N,N,N,V,N,N
,N,A,N,A,N,N,N,N,N,V,N,A,N,V,N,A,N,V,N,A,N,A,N,V,N,A,N,A,A,V,A,N,V,N,A,
V,N,A,V,N,N,N,N,A,N,V,N,N,A,N,N,N,N,V,N,N,V,N,N,A,N,N,N,V,N,N,V,N,A,

N,N,N,N,A,N,N,N,N,N,A,N,N,A,V,N,N,N,A,N,N,V,N,N,N,V,N,N,V,N,A,N,A,N,N,
V,V,A,N,N,V,N,A,N,V,A,N,A,N,N,N,N,V,N,N,N,N,N,A,V,N,V,N,A,N,N,N,N,N,N,
N,A,A,N,N,N,V,V,N,N,N,N,N,N,A,N,V,A,N,A,N,A,V,N,V,N,N,A,N,A,A,N,N,N,N,
N,A,N,N,V,V,N,A,N,A,A,N,N,V,N,N,A,N,N,V,N,A,N,N,N,N,N,V,N,A,A,A,N,N,N,
V,N,N,N,A,N,N,N,N,N,N,A,N,N,N,N,A,N,N,V,N,A,N,N,A,N,A,N,A,N,V,N,N,N,N
,N,A,N,N,A,V,V,A,N,N,V,N,N,N,A,N,V,N,A,N,N,N,A,A,V,N,N,N,N,N,A,N,A,N,A
,N,V,N,N,N,N,N,A,V,V,A,V,N,N,A,N,A,N,N,N,A,N,A,N,N,V,V,N,N,A,V,N,N,A,A,
N,N,A,N,N,N,N,N,N,N,N,N,V,N,A,A,N,A,N,A,A,V,A,N,N,A,N,V,V,N,N,N,N,V,N,N,
A,A,A,N,A,N,N,V,N,V,A,N,N,V,V,N,N,N,A,V,N,N,A,N,A,A,N,A,N,A,A,A,N,A,A,N,
N,A,N,N,A,N,V,N,A,N,A,N,A,N,A,V,A,N,V,V,A,N,A,N,N,N,N,V,N,N,N,A,N,N,N,
N,A,N,N,V,N,A,N,N,A,N,V,V,N,N,N,N,N,V,N,N,A,N,A,N,N,A,N,N,A,N,V]

T 5: 0.2.04.2015
Larijani's congratulation to Yemenis
تبریک لاریجانی به ملت یمن

[N,N,V,N,N,N,N,N,N,V,N,V,N,N,N,N,N,V,N,N,N,N,N,N,N,N,N,N,N,N,N,N,N,N,
N,V,V,V,V,A,N,N,N,N,N,N,V,N,N,A,N,A,N,V,N,N,V,V,N,A,N,A,N,A,N,N,N,A,N,
V,N,N,N,A,V,N,A,V,N,V,A,A,N,N,N,V,N,A,A,N,N,N,N,N,V,A,N,V,N,N,N,N,V,N,
N,A,N,N,N,N,A,N,V,N,A,V,A,A,N,N,N,N,N,N,N,N,V,N,N,V,A,N,V,N,A,V,N,A,V,N,
A,N,N,N,N,V,N,N,A,N,V,N,A,N,N,A,N,V,V,A,N,A,A,N,N,A,N,V,N,V,N,N,N,N,N,
N,N,N,V,A,N,A,N,N,V,N,N,V,A,N,A,A,N,V,V,A,N,N,A,V,N,N,N,N,A,A,V,A,N,N,
A,V,A,N,A,A,N,A,V,N,N,A,N,A,N,A,V,A,N,N,A,A,N,A,V,N,V,A,V,N,N,V,N,N,N,
N,N,N,N,A,V,V,A,N,V,A,N,N,A,A,N,N,A,N,A,V,N,N,V,A,N,N,N,N,N,N,N,V,N,N,
A,N,A,A,N,N,N,V,N,N,A,N,A,N,N,V,N,V,A,N,N,A,N,V,N,A,N,A,N,A,V,V,A,N,A,
N,V,N,A,N,A,A,V,V,A,N,A,N,N,A,N,V,V,N,A,A,N,V,N,N,V,N,A,N,N,N,V,N,N,N,V,V
,N,N,A,N,N,N,N,A,V,V,N,N,N,N,N,N,N,N,V,N,N,N,A,N,A,V,V,A,N,N,N,V,N,A,V,
N,V,N,A,N,N,A,N,V,A,N,V,N,N,V,N,A,N,N,N,V,A,N,A,V,V,A,N,N,A,N,A,V,N,A,
N,N,A,N,A,V,A,N,N,N,A,N,V,N,N,A,N,N,N,A,V,N,N,V,N,A,N,A,N,N,A,V,N,A,A,
N,N,A,N,A,N,V,N,N,V,N,N,N,N,A,V,N,N,V,N,N,N,N,N,N,V,A,V,N,N,N,N,N,V,V,
N,N,V,N,N,N,A,N,N,V,N,A,A,A,N,N,N,A,N,N,N,V,A,N,A,N,N,N,A,V,N,N,N,A,N,
V,A,N,A,V,N,N,N,A,V,N,N,N,A,V,N,N,A,N,N,A,N,N,A,N,A,N,N,N,A,A,A,N,A,N,
A,V,N,N,N,A,N,N,N,A,A,N,N,N,A,A,A,N,N,N,N,A,N,N,N,N,V,A,N,N,A,A,A,N,
A,V,N,A,N,N,A,A,A,V,N,A,A,N,A,V,N,N,A,A,N,N,N,N,N,V,V,N,A,N,N,N,N,N,N,
N,N,A,N,A,N,V,N,N,N,V,A,N,A,N,N,N,A,N,A,N,N,V,N,N,A,N,N,A,A,N,V,N,A,N,
N,N,N,A,N,N,A,V,N,N,A,N,N,N,N,N,N,N,N,A,N,A,N,N,N,A,V,N,A,N,N,A,V,N,N
,A,A,V,N,N,N,N,A,N,V,V,N,N,N,N,V,N,N,N,N,A,A,N,A,V,V,A,N,N,N,N,N,A,A,A,
V,N,N,N,N,A,V,N,V,N,A,V,A,N,A,A,A,N,V,A,N,N,N,N,A,N,N,A,V,V,V,A,N,N,A,A,
N,V,N,N,A,N,N,A,A,N,A,A,N,A,A,N,N,V,N,A,V,N,A,A,N,N,A,V,N,N,N,V,N,A,A,
N,A,N,N,A,V,N,V,N,A,N,A,N,V,V,A,N,A,N,A,N,N,A,V,A,N,A,V,N,N,A,A,A,N,A,
N,V]

German

T 1: Burgaufzug begeistert. ***Der Bote 25.04.2015***
[V,N,V,N,V,N,V,N,V,A,A,V,V,N,A,N,A,N,N,V,N,N,N,N,V,N,N,N,N,N,N,N,N,
V,N,N,A,N,N,N,N,N,N,N,N,N,V,A,A,N,N,A,N,V,N,A,N,V,N,V,A,A,N,A,N,A,
N,V,N,N,V,N,N,V,A,N,A,N,N,N,V,A,N,A,N,N,N,N,V,A,V,V,N,N,N,V,V,V,N,
V,A,N,N,V,N,N,V,N,V,N,V,N,V,N,N,V,A,N,V,V,V,N,N,V,N,V,N,V,N,N,V,N,
N,V,V,N,A,N,V,N,A,V,N,N,N,N,V,V,N,A,N,N,N,V,A,N,V,N]

T 2: Und plötzlich geht die Tür auf. ***Der Bote 25.04.2015***
[N,V,N,V,N,V,N,N,N,N,N,V,V,N,A,N,A,N,A,A,N,V,N,N,V,N,N,V,N,N,V,N,N,N,V,
N,N,N,V,V,N,A,N,N,V,N,N,A,N,V,N,V,N,V,N,N,V,N,V,N,A,N,N,N,V,A,N,N,N,V,V
,N,N,N,V,V,N,V,N,N,V,V,N,N,V,N,V,N,N,V,A,N,V,N,N,V,V,N,N,N,V,N,A,N,V,N,V,
N,N,A,N,V,N,A,N,N,N,N,V,V,N,N,V,N,V,N,V,V,V,N,N,V,V,N,V,N,V,N,V,V,N,N,V,
N,A,N,V,N,V,N,V,N,V,N,V,N,N,N,N,V,A,N,N,N,A,N,N,V,N,N,V,A,N,N,V,N,V,N,N
,V,V,N,N,N,V,N,V,A,N,V,N,N,N,A,N,A,N,V,V,V,V,V,N,V,V,A,N,V,N,N,A,A,N,V,N,
V,N,A,N,V,N,N,N,V,N,N,V,V,N,V,N,A,N,V,N,V,V,A,N,N,N,N,V,N,V,V,V,N,N,V,N,
N,N,V,V,N,N,N,N,N,V,V,N,N,V,N,N,A,N,V,V,N,N,N,N,N,N,V,N,V,V,V,A,N,V,N,V,
V,V,N,V,N,A,N,N,V,N,N,V,N,V,A,N,N,N,V,A,N,N,N,N,N,N,N,V,V,V,V,N,V,N,N,N,
N,N,V,N,N]

T 3. Drama des gescheiterten Hitler-Attentats. ***Der Bote 22.04.2015.***
[A,N,N,N,V,N,V,N,V,N,A,N,V,N,N,N,N,N,V,A,N,V,N,N,N,N,V,N,V,A,N,V,A,N,V,
V,N,N,N,V,N,A,N,N,V,N,N,V,A,N,A,N,A,N,N,N,V,N,N,V,N,N,N,N,N,N,V,N,N,A,
A,N,V,N,N,N,N,N,N,N,N,N,N,V,N,V,N,N,N,N,N,V,N,N,N,V,V,N,A,N,N,N,N,V,N,
V,N,N,N,V,N,N,A,N,N,N,A,N,N,N,V,V,N,N,V,V,N,N,N,N,N,N,V,N,N,A,N,N,N,V,
A,V,V,A,N,N,N,V,A,A,N,N,N,V,V,N,A,N,A,A,N,N,N,V,A,N,N,N,A,N,V,V,A,A,N,
N,N,N,N,A,A,N,N,N,N,N,N,N,V,N,N,V,N,A,N,V,N,A,N,A,N,N,V,A,N,N,N,V,A,V,
N,V,N,N,V,V,N,A,N,V,A,N,N,V,N,V,N,N,N,N,N,V,N,N,N,V,N,V,V,A,N,N,V,N,V,N,
V,N,A,N,V,N,A,V,V,N,N,V,N,N,N,V,V,N,N,N,V]

T 4. Sieben Minuten raus aus dem Alltag, rein in die Tiefe. ***Der Bote 29.04.2015***
[V,N,A,N,V,A,N,N,N,N,V,V,A,A,N,N,V,N,V,A,V,N,A,N,N,N,N,N,N,A,N,N,N,V,N,
N,N,N,N,N,V,N,N,V,A,N,A,N,N,N,N,A,V,N,V,N,V,N,A,N,N,N,A,N,N,A,N,V,N,N,
A,N,V,N,A,N,N,V,N,N,N,V,V,V,A,N,N,N,N,N,A,A,N,V,V,N,N,V,A,N,N,N,N,V,N,V
,V,N,N,A,N,N,V,N,A,N,V,N,V,N,N,V,V,V,N,A,N,A,A,N,N,V,V,N,N,N,A,N,V,A,N,
V,A,N,V,N,N,N,N,N,N,V,N,N,N,N,V,N,N,A,V,N,N,N,N,N,N,N,V,N,N,V,N,N,N,N,
V,N,V,V,N,N,V,N,V,A,V,N,A,N,V,V,N,N,N,V,A,N,A,N,V,N,A,N,N,N,V,N,V,N,N,N,
A,A,N,N,N,V,N,A,N,N,N,N,N,V,A]

T 5. Geruch ist kein Zufall. ***Der Bote 11.04.2015.***
[V,V,V,N,N,N,V,N,N,N,N,V,N,V,A,V,N,V,N,N,A,N,V,A,V,A,N,N,V,A,N,N,V,V,N,A
,N,A,N,V,A,N,V,A,N,V,V,N,N,V,A,N,A,N,V,A,N,N,A,V,N,V,N,A,N,V,A,N,N,V,N,
A,V,A,N,V,N,V,A,A,N,V,N,N,N,N,A,N,N,N,N,A,V,N,V,N,A,N,V,N,A,N,V,N,N,A,

N,V,V,V,V,A,N,V,N,N,N,V,A,N,A,N,N,V,N,A,N,N,N,N,N,N,V,A,N,A,N,V,N,N,N,
N,N,V,V,N,N,A,N,V,N,N,N,V,N,N,A,A,N,N,V,N,A,N,V,A,N,V,N,N,N,V,N,N,N,A,
V,N,N,V,A,N,N,N,V,A,N,V,V,V,N,N,A,A,N,N,V,A,N,V,N,N,V,V,N,A,N,V,V,N,N,V,
N,N,N]

Odia

T 1: *Sambad,* Page: 4. Bhubaneswar, 24 March 2015: Dalapati will take the final decision on the choice of the wood

[V,N,N,N,A,N,N,N,V,N,N,N,N,N,N,N,N,A,A,N,A,N,N,N,N,N,N,V,A,N,N,N,N,
V,N,N,N,V,N,N,N,N,N,N,V,N,N,N,N,V,N,N,N,V,N,V,N,N,N,N,V,N,N,N,V,A,
N,N,N,N,N,N,N,N,V,N,N,N,N,N,N,A,A,N,N,V,N,N,N,N,A,N,N,N,N,V,N,N,N,A,N,
N,N,N,N,V,N,V,N,N,N,N,N,N,N,V,N,N,N,A,N,N,N,N,N,N,V,N,N,A,N,N,N,N,
N,A,N,V,N,N,A,N,V,N,V,N,A,N,V,N,N,V,N,A,N,V,N,V,A,N,N,A,A,N,N,N,N,
N,V,A,N,N,N,N,N,V,N,N,V,N,N,V,N,V,A,N,N,N,N,N,N,N,N,N,N,N,A,N,
N,N,N,N,N,N,N,N,N,N,N,N,N,N,A,N,N,V,N,N,N,N,A,N,N,A,N,N,N,N,A,N,N,N,
A,N,N,V,N,N,N,N,N,N,V,N,A,N,A,V,N,N,A,N,N,N,N,V,A,N,N,N,N,N,N,N,N,
V,N,N,N,N,V,N,N,V,A,N,N,V,N,A,N,A,N,A,N,V,N,N,N,V,N,A,N,A,A,N,N,A,
N,V,N,N,N,V,N,V,A,N,N,A,N,N,N,V,N,N,N,N,V,N,N,N,N,V,N,N,N,N,N,N,N,
N,A,N,N,V,N,N,N,A,A,N,V,A,N,V,N,N,N,N,N,N,V,N,N,N,V,A,N,A,N,V,N,N,
N,N,A,N,N,A,N,N,V,N,V]

T 2: *Sambad,* Page: 4. Bhubaneswar, 24 March 2015. English Question on appointment of Commission for the Differently Abled

[N,N,A,N,A,N,N,N,V,N,N,N,V,A,N,V,N,N,A,N,V,A,N,N,V,N,A,N,N,N,V,N,V,
N,N,A,N,A,N,N,N,N,N,N,V,N,N,V,N,N,N,N,N,N,V,A,N,N,V,N,N,N,N,V,A,N,
N,V,A,N,V,A,N,N,V,A,N,N,V,V,N,N,V,N,N,N,V,N,V,N,V,A,N,N,N,N,N,A,A,
N,V,A,N,N,N,V,N,N,N,N,N,N,A,A,N,V,N,N,N,V,A,N,N,A,N,V,A,N,N,N,N,N,
N,N,N,N,N,V,A,N,V,N,A,N,N,A,N,N,V,N,V,N,N,N,A,V,N,N,N,N,V,N,V,A,N,
N,A,A,N,V,A,N,N,N,V,A,N,N,V,A,N,N,N,N,A,V,N,N,N,N,V,N,V,N,N,N,N,V,
N,A,A,N,N,V,A,N,A,N,N,V,N,V]

T 3: *Sambad,* Page: 18. Bhubaneswar, 24 March 2015. Effort be made for upholding the respect and solving the problems of farmers

[A,N,N,A,N,N,N,N,V,N,V,N,N,N,N,N,N,N,V,N,A,N,N,N,N,N,N,N,V,N,N,N,N,
N,A,N,V,N,V,N,V,A,N,N,N,V,A,A,N,N,N,V,N,N,N,N,N,N,A,N,N,V,N,V,N,V,
N,A,N,N,N,N,N,V,V,A,N,A,N,A,N,N,A,N,V,N,V,A,N,A,N,N,N,N,N,V,N,N,V,
A,N,A,N,N,N,N,V,N,N,V,V,N,N,V,N,N,N,N,N,N,V,N,A,N,N,N,A,V,N,N,N,N,
V,N,N,V,N,A,A,N,V,A,N,V,N,A,N,A,V,V,N,A,N,N,V,N,N,V,V,N,N,V,N,N,V,
N,A,A,N,A,N,N,N,N,A,N,N,A,N,N,N,V,A,N,A,N,N,A,N,A,N,A,N,N,V,N,N,N,
V,A,N,A,N,V,N,A,V,A,N,N,N,V,N,V,N,N,V,N,N,N,V,N,N,N,N,N,V,N,A,N,V,
N,V,N,N,V,N,V,N,V,V,N,N,V,N,V,N,N,N,V,N,A,N,N,N,V,N,N,V,A,A,N,N,A,
N,N,N,N,N,N,N,V,N,N,A,N,N,V,N,A,N,N,N,V]

T 4: *Sambad*, Page: 19. Bhubaneswar, 24 March 2015. Now also threatening tuberculosis
[N,A,N,V,N,N,N,N,N,N,V,N,V,N,N,V,A,N,N,N,V,V,N,V,A,V,N,N,N,N,V,A,N,
N,A,N,A,N,N,V,N,N,A,N,N,A,N,N,A,N,N,A,V,A,N,N,V,A,N,N,A,V,V,N,N,N,
A,V,A,N,N,A,N,N,A,V,A,N,N,N,V,A,N,N,N,N,V,V,N,V,N,N,N,N,N,A,N,N,A,
N,N,A,V,N,A,V,N,V,N,N,A,V,N,A,N,A,N,A,N,A,N,A,N,A,N,A,N,A,N,A,N,N,
N,V,N,N,N,V,N,N,V,A,N,N,N,V,N,N,N,V,N,N,A,V,N,A,N,N,V,A,N,N,V,N,A,
N,N,A,N,A,N,A,A,V,N,A,N,N,N,V,A,N,A,N,N,A,V,V,N,N,N,N,A,N,N,N,V,N,
N,N,V,A,V,N,A,N,A,N,V,N,A,N,V,A,N,N,A,N,N,V,A,N,N,A,V,N,N,A,V,N,A,
N,A,N,N,V,N,A,V,A,N,N,N,V,N,N,N,V,A,N,A,N,N,V,N,V,N,N,N,V,A,N,N,V,
A,N,N,A,V,N,N,N,N,N,A,V,N,N,N,N,N,N,N,N,A,V,N,A,V]

T 5: *Sambad*, Page: 14. Bhubaneswar, 24 March 2015. Child-stealing racket in the capital
[N,N,A,N,N,N,N,N,A,N,N,A,V,N,A,N,N,V,N,N,N,N,N,A,N,A,A,N,N,A,N,A,V,
A,N,A,N,V,N,A,N,N,A,V,N,N,V,N,N,N,V,V,N,N,N,N,N,N,V,A,V,N,N,N,V,N,
N,V,N,V,V,A,A,N,N,V,N,N,A,A,N,V,N,V,N,N,N,N,N,N,N,V,N,N,A,V,N,V,N,
V,N,N,N,N,N,N,N,N,N,V,A,N,N,A,N,N,N,N,N,N,V,N,N,N,N,V,N,A,N,V,N,N,
N,V,A,N,A,N,N,V,N,A,V,N,N,N,V,V,A,A,N,A,V,A,N,V,A,V,N,N,V,N,N,N,N,
A,N,N,A,N,N,A,N,V,N,N,N,N,V,A,A,N,N,N,V,A,A,V,N,V,N,N,V,N,N,N,N,A,
N,N,V,A,A,N,N,N,A,A,A,N,V,V,V,N,V,N,V,A,N,N,N,V,A,V,N,N,N,N,A,N,V,
N,N,A,N,N,A,N,N,N,N,V,N,A,N,N,N,V,V,N,A,N,N,V,N,N,V,N,V,N,N,V,A,N,
N,N,N,N,V,V,N,A,V,A,N,V,A,N,N,N,N,N,N,N,V,N,N,N,V,N,N,V,N,V,N,A,A,
N,N,N,N,V,N,N,A,V,N,N,V,N,A,A,N,N,V,N,N,V,N,N,V]

Russian

T 1: http://www.novayagazeta.ru/news/1693890.html 19. Mai 2015
[A,N,V,V,A,N,N,N,N,N,N,N,V,N,V,N,N,N,N,V,N,N,N,N,N,V,A,N,V,N,A,N,A,
N,N,V,N,N,A,N,V,V,N,N,N,V,N,V,A,N,N,N,V,N,N,V,V,N,N,V,N,N,V,N,N,V,
N,N,A,N,V,A,N,V,N,A,N,N,N,N,N,A,N,A,N,N,N,N,A,N,A,N,N,N,V,N,V,N,N,
V,N,V,N,N,N,N,N,A,N,N,N]

T 2: http://www.ng.ru/economics/2015-05-19/100_obzor190515_2.html 19. Mai 2015
[N,V,N,A,N,N,V,A,N,N,N,N,N,N,V,N,V,N,A,N,N,N,A,N,N,V,A,N,N,N,N,N,V,
N,N,V,N,N,N,N,N,V,N,V,A,N,N,V,A,N,N,N,V,A,N,N,N,N,V,N,N,N,N,V,N,N,
V,N,V,N,N,A,N,N,V,N,N,N,N,A,N,V,N,N,N,N,N,V,N,A,N,A,N,N,V,N,N,V,N,
V,A,N,N,N,N,A,N,N,N,N,N,N,N,V,A,N,N,N,A,N,N,V,N,V,N,N,N,N,N,N,N,N,
N,A,N,N,V,N,N,N,A,N,N]

T 3: http://www.ng.ru/news/503821.html 19. Mai 2015
[N,N,N,V,N,N,N,N,N,V,N,N,N,N,V,N,V,N,N,N,N,N,N,V,N,N,N,V,N,N,N,N,A,
N,N,N,N,V,N,N,V,N,N,N,V,V,N,N,N,N,N,N,N,N,N,A,N,N,N,N,V,N,N,A,N,N,

N,V,A,N,N,N,N,N,N,V,N,A,A,N,N,N,N,N,N,N,N,A,N,N,N,N,N,N,V,N,N,N,V,
N,V,N,N,N,V,N,N,V,N,N,N,N,N,A,N,N,N,N,N,A,N,N,V,N,N,N,N]

T 4: http://lenta.ru/news/2015/05/19/ukrcredban/ 19. Mai 2015

[N,N,V,V,N,N,N,V,A,N,N,N,N,A,N,A,N,V,N,N,N,V,N,V,N,N,A,N,V,N,V,N,V,
N,N,V,N,N,N,N,V,A,N,N,N,V,N,N,N,N,V,A,A,N,N,N,N,N,N,N,N,V,N,V,A,
A,N,N,A,A,N,N,V,N,N,V,N,A,N,N,V,N,A,N,N,N,N,N,N,A,N,N,N,N,N,V,N,N,
N,A,A,N,N,N,V,N,N,A,N,N,N,N,A,N,V,N,N,N,N,A,N,N,V,N,A,N,N,N,N,N,V,
N,N,N,N,V,N,N,N,V,A,N,N,V,N,N,N]

T 5: http://lenta.ru/news/2015/05/19/language/ 19 Mai 2015

[N,V,N,N,A,N,N,N,N,V,N,A,N,V,V,A,N,N,A,N,N,V,N,A,A,N,V,A,A,N,V,N,A,
N,V,N,N,V,A,N,A,N,V,A,N,N,V,N,A,A,N,V,N,N,V,A,N,V,N,N,N,V,A,A,N,V,
N,A,N,N,N,V,N,A,N,N,A,N,N,N,A,A,N,V,N,N,N,A,N,V,V,N,N,A,N,V,N,N,N,
V,N,N,N,A,N,N,N,A,N,N,A,N,N,V,N,V,N,N,N,V,N,N,A,N,A,N,N,A,N,V,A,N,
A,N,N,N,V,A,N,A,N,N,N,A,N,V,N,A,N,N,N,V,N,A,N]

Turkish

T 1: http://www.hurriyet.com.tr/ekonomi/29071597.asp
[N,N,N,V,N,N,N,A,N,N,N,N,N,N,N,V,A,N,N,N,A,N,N,V,N,N,V,N,N,N,A,A,N,
N,A,N,A,N,A,N,N,N,V,N,N,N,N,A,N,N,A,N,V,V,N,N,V,A,N,A,N,N,A,N,N,N,
V,A,N,N,N,V,N,N,N,N,A,N,A,N,A,N,N,N,A,N,V,N,A,N,N,N,A,N,N,A,A,A,N,
V,N,A,N,N,A,A,N,A,N,A,N,N,V,N,N,N,V,N,N,N,V,N,N,V,N,N,N,N,V,N,N,N,
V,N,A,N,A,N,A,N,A,N,V,N,N,N,A,N,N,V,N,A,N,N,N,A,A,A,N,N,V,V,N,A,A,
N,N,N,A,N,N,A,N,N,N,A,A,N,N,V,N,N,N,N,V,N,V,N,N,A,V,N,N,A,N,V,N,V,
A,N,V,V,N,A,N,N,V,N,N,N,V,A,N,V,N,V,N,A,N,V,N,V,N,V,N,V,V,A,N,N,V,
A,V,V,V,A,V,N,N,V,A,N,N,N,N,V,N,A,V,A,N,V,V,N,A,N,N,N,V,N,A,N,N,V,
N,N,N,N,V,N,N,A,A,N,A,N,N,N,N,N,N,N,N,N,V,N,N,N,N,N,V,N,N,V,N,A,A,
N,N,N,N,V,N,V,N,N,N,V]

T 2: http://www.milliyet.com.tr/sisi-ye-A,lmA,N,yA,-dA,-hitler/duN,yA,/detA,y/2068911/defA,ult.htm
[N,N,N,V,N,N,A,N,A,N,A,N,N,N,V,N,N,A,N,N,V,N,V,N,N,A,N,N,V,N,N,N,N,
N,A,N,N,V,N,N,N,N,A,N,N,N,N,N,V,N,N,N,A,N,N,V,N,N,N,A,N,N,A,N,N,N,
V,N,N,N,V,A,N,A,N,N,V,N,A,N,V,A,N,A,A,A,N,V,N,N,N,N,A,N,N,A,N,A,N,
V,N,N,V,N,V,N,V,N,N,V,N,N,A,N,N,A,N,N,N,V,N,V,A,N,N,N,N,V,N,N,A,N,
V,N,A,N,N,V,N,N,N,N,A,N,N,N,N,V,N,N,N,N,V,N,N,N,N,N,N,A,N,N,N,A,N,
N,N,V,N,N,N,N,V,A,N,N,V,N,A,N,V,N,N,V,N,A,N,A,N,V,N,V,V,N,A,N,N,A,
N,A,N,N,N,N,N,N,N,V,N,N,N,N,N,V,A,A,N,N,V,N,N,N,N,V,N,V,V,N,N,N,A,
N,N,A,N,A,N,V,N,A,N,N,A,N,V,N,N,V,N,V,N,N,A,N,N,V,N,N,N,A,N,N,N,V,
N,N,N,A,N,N,A,N,N,V,A,N,V,N,N,N,A,A,N,N,V,N,A,N,V,N,V,N,N,N,N,N,A,

A,N,A,N,A,N,N,N,N,N,N,N,N,A,N,N,V,N,V,N,N,N,A,N,N,V,N,A,A,N,N,N,A,
N,N,V,A,N,V,A,N,N,N,V,N,N,N,N,V,N,N,N,V,V,V,N,N,V,N,N,N,N,N,A,N,N,
N,A,N,N,N,N,V,N,N,N,A,N,A,N,N,N,V,N,A,N,V,A,N,A,N,A,N,N,A,N,V,N,V,
N,N,N,V,N,N,A,N,N,V,N,N,N,N,V,N,N,N,N,N,V,N,A,N,N,A,N,A,N,A,N,N,N,
N,N,N,N,N,V,N,A,N,N,N,A,N,A,N,A,N,N,N,V,A,N,V,V,N,V,A,N,N,N,N,N,V,
N,N,V,N,N,V,V,N,V,N,N,A,N,N,N,N,A,N,N,A,N,V,N,V,N,N,V,V]

T 3: http://www.milliyet.com.tr/2015-rA,mA,zA,N,-A,yi-N,e-zA,mA,N,-guN,dem-2069220/
[A,A,N,N,V,A,N,A,N,V,A,N,A,N,A,A,N,N,A,A,N,V,A,A,N,N,A,A,N,A,N,A,A,
N,V,N,V,N,A,N,A,N,V,A,N,A,N,N,N,N,V,A,N,A,A,N,V,N,N,N,A,A,A,N,N,A,
A,N,A,V,N,A,N,A,N,N,A,V,N,N,A,N,V,A,N,V,N,N,A,A,N,N,V,N,N,A,A,A,N,
A,N,N,A,N,N,V,V,N,N,A,N,N,A,N,A,A,N,N,N,N,A,V,A,N,N,A,N,A,N,A,N,N,
A,A,N,N,N,A,N,V,A,A,N,N,N,N,A,V,A,N,A,A,A,A,N,V,A,N,A,N,N,A,N,V,A,
N,A,A,N,V,A,A,N,V,A,A,A,N,N,N,N,A,V,A,N,A,A,N,A,N,N,A,A,N,V,A,A,N,
V,N,A,N,N,V,N,A,N,N,V,N,N,N,N,N,N,N,N,N,N,V,N,N,N,N,V,A,N,A,V,A,N,
A,N,A,A,N,V,A,A,N,A,N,V,N,A,N,N,N,N,N,A,A,N,V,A,N,A,N,N,N,A,N,A,N,
V,N,N,A,N,A,N,A,N,V,A,A,A,N,A,N,A,N,V,N,N,N,A,N,N,A,N,A,A,A,N,V,A,
N,A,A,N,V,N,A,N,V,A,N,N,A,N,A,N,N,N,N,A,N,V,A,N,V,N,N,A,N,A,N,A,N,
V,N,N,N,N,A,V,N,A,N,N,A,A,N,N,V,N,A,N,A,N,V,A,N,A,A,N,N,N,A,N,N,V,
A,N,A,A,N,N,N,A,A,N,N,N,V,A,N,A,N,A,N,N,V,A,N,N,N,V,A,N,N,V,A,N,A,
N,A,N,A,N,V,N,N,N,A,N,A,A,N,A,N,V,A,N,N,N,A,N,A,N,A,N,V,A,N,N,A,N,
A,N,A,A,N,V,A,N,N,A,N,V,A,N,A,N,A,V,A,N,N,A,N,N,N,A,N,V,A,N,N,N,N,
N,V,A,N,V,N,A,N,N,N,N,N,A,N,N,N,A,N,N,V,N,N,A,N,A,N,V,V,N,V,V,A,N,
V,A,N,V,N,N,N,V,N,V,A,N,N,N,A,A,N,N,N,N,V,V,V,A,N,N,N,A,N,N,A,N,N,
N,A,N,N,N,N,V,N,A,N,A,N,N,V]

T 4: http://www.zA,mA,N,.com.tr/kultur_kitA,p-kA,sA,bA,siN,dA,-edebiyA,t-festiV,A,li_2296797.html
[A,N,A,N,N,N,N,N,N,N,N,V,N,N,N,N,N,N,N,A,N,A,N,N,V,N,A,N,N,N,N,N,N,
A,N,A,N,V,N,N,A,N,N,V,A,N,N,A,N,A,N,A,N,N,N,V,N,A,N,V,A,N,A,N,A,N,
V,N,A,A,N,A,A,N,N,A,N,A,N,A,V,N,N,N,A,N,N,A,A,N,N,N,A,V,A,N,A,N,N,
N,A,N,N,V,N,N,A,N,N,N,N,A,A,N,V,A,N,N,A,N,A,N,N,A,N,A,N,N,V,N,N,N,
A,N,N,N,V,N,N,N,N,N,V,A,N,A,A,N,A,A,N,V,N,A,A,N,N,A,N,N,N,V,N,N,N,
N,V,A,N,A,N,A,N,N,N,V,N,A,N,A,A,N,N,A,A,N,N,N,N,V,N,N,N,A,N,N,N,N,
V,N,N,A,A,N,N,N,A,N,A,A,N,N,N,V,V,N,A,A,N,A,N,N,V,N,A,N,A,A,A,N,N,
N,N,N,V,N,N,N,A,A,N,V,N,N,V,A,N,A,N,A,N,N,A,N,N,V,A,A,N,A,A,N,A,N,
N,V,N,N,N,N,N,A,N,N,N,V,N,N,N,N,N,A,N,V,V,N,N,N,A,N,N,V,N,A,N,N,A,
A,N,N,N,N,A,N,A,N,N,A,N,V,A,N,V,N,N,A,N,N,A,N,N,V,N,A,N,N,N,A,N,A,
N,N,A,N,N,A,V,A,N,A,N,A,N,V,N,N,A,N,A,A,N,N,V,N,N,A,N,A,A,N,A,N,N,
A,N,N,N,N,N,N,V,A,N,N,A,N,A,V,A,N,N,N,A,V,N,A,N,N,N,V,A,A,N,N,A,N,
N,N,V,N,A,N,A,N,V,A,N,N,A,A,N,A,N,N,A,N,A,N,A,N,N,A,N,V]

T 5: http://www.sA,bA,h.com.tr/kultur_sA,N,A,t/2015/05/10/fA,tih-sultA,N,-mehmetiN,-kilici-sA,tildi

[A,A,N,A,N,A,A,N,N,A,N,A,N,A,N,V,A,N,A,A,A,N,N,A,V,A,N,N,N,A,N,A,A,
N,A,N,V,V,A,N,N,V,N,A,N,A,N,A,N,N,N,A,N,A,V,N,A,N,A,N,A,A,A,N,A,N,
V,A,A,N,A,N,A,N,N,A,A,N,A,N,N,A,N,N,A,N,A,N,N,A,N,V,N,N,N,N,N,N,A,
A,N,V,N,A,N,A,N,A,N,A,N,V,N,N,A,A,A,A,N,V,V,N,V,N,N,A,N,N,N,N,N,N,
A,N,A,N,N,N,V,V,A,N,A,A,N,N,N,A,N,A,N,A,N,N,N,V,V,N,N,V,A,N,N,N,A,
N,N,A,N,N,N,A,A,N,A,N,N,V,V,N,A,V,N,A,A,N,A,A,A,A,N,A,N,A,N,A,N,N,
A,N,A,N,N,V,N,N,N,A,N,N,N,A,N,A,A,A,N,V,V,N,A,A,N,A,A,A,A,N,A,N,A,
N,A,N,N,A,N,A,N,N,V,N,N,N,A,N,N,N,A,N,A,A,A,N,V,V,A,N,N,N,A,N,N,A,
A,N,V,N,A,N,A,A,N,N,A,V,A,N,A,N,N,A,A,A,N,V,V,N,N,N,V,A,A,N,A,N,A,
N,N,N,N,N,A,N,A,N,V,N,A,A,N,N,N,V,N,N,N,V,N,N,N,V,N,N,N,V,A,N,A,N,
N,V,N,N,N,V,A,N,A,A,N,A,N,N,A,N,N,N,V,A,N,A,A,A,A,N,V,N,A,A,N,A,N,
A,N,N,N,N,A,N,N,N,V,N,N,A,N,A,N,V,N,A,A,N,V,N,A,A,N,N,N,V,V,N,A,N,
N,N,A,N,N,A,A,N,N,A,N,N,A,N,V,A,N,A,N,N,V,N,A,N,V,A,N,A,A,A,N,V,N,
N,N,N,A,N,N,A,N,A,N,V,N,N,V]

Author index

Altmann, G. 2,4,5,19,33,40,66,73, 74,96,97,99-101
Altmann, V. 33,99
Antosch, F. 4,99
Aristotle 40
Bakker, F.J. 4,99
Ballmer, T.T. 3,99
Banguoğlu, T. 10,99
Barwise, J. 44,99
Beliankou, A. 96,99
Best, K.-H. 4,19,40,74,101
Boehnke, K. 49,99
Bortz, J. 49,99
Boy, J. 100
Bradley, J.V. 99
Brennenstuhl, W. 3,89
Brownlee, K.A. 34,99
Bunge, M. 98,99
Busemann, A. 4,99
Bußmann, H. 40,99
Bybee, J. 100
Čech, R. 2,4,19,99,100
Darley, F.L. 3,99
DeVito, J.A. 3,99
Erzen-Rasch, M.I. 10,99
Etchemendy, J. 49,99
Fickermann, I. 101
Fischer, H. 4,100
Greenbaum, S. 101
Grotjahn, R. 36,100
Gunzenhäuser R. 100
Hammerl, R. 95,101
Hopper, P. 100
Janda, L. 1,100
Kelih, E. 98
Kisro-Völker, S. 3.100
Köhler, R. 1,5,40,66,72-74,95,96,99-101
Kreuzer, H. 100
Krug, M.G. 2,100
Leech, G. 101
Lienert, G.A. 49,99
Löbner, S. 44,100
Lupea, M. 4,100
Mačutek, J. 3,60,100
Mates, B. 44,100
Mikros, G. 3,100
Naumann, S. 96,99
Obradović, I. 98
Ord, J.K. 100
Paivio, A. 3,100
Pikas, A. 3,100
Piotrowski, R.G. 100,101
Popescu, I.-I. 2,4,19,17,99,10
Quirk, R. 2,101
Salmon, W.C. 44,101
Sambor, J. 95,101
Schlissmann, A. 4,101
Sherman, D. 3,99
Siegel, G.M. 3,99
Skinner, B.F. 66,73,101
Stachowski, K. 10,101
Svartvik, J. 101
Swift, L.B. 10,101
Tatar, D. 4,100
Tuzzi, A. 74,100
Uhlířová, L. 101
Wildgen, W. 44,101
Wimmer, G. 60,96,97,101
Ziegler, A. 4,8,19,40,101
Zörnig, P. 53,55-57,60,101

Subject index

Abstractness 3
activity 12
angle 13
beta function 19
binomial distribution 4,5,47
Busemann-indicator 4,14,44
chi-square test 41,44,45,59,60
classification 1,4,14,19,98
concreteness 3
distance 53-71
dynamic system 1
equilibrium 4,40,41
Euclidean distance 46
Excess 80,81,83,95
Generic 3
hierarchy 2,
hypergeometric series 4
hypothesis 1,
imagery 3
Köhlerian motif 72
Language
- Austroasiatic 10
- Brasilian-Portuguese 6-8,12-14, 18,41,42,45,46
- Chinese 7,8,12-15,17,18,24,25, 35,37,39,43,45,48, 49,51,52,62, 63,65,68,77,80,81,87,88,91,93
- Croatian 7,8,12-14,17,18,21-23, 35, 37,39,43,45-47,49, 51,52,60,62, 65,68, 75,80,81,85,91,93
- Dravidian 9,10
- English 2,40
- German 12-18,27-29,35,38-40,43, 45,48,49,51,52,63,64,69,76,80,81, 86, 91,93
- Greek 10
- Hindi 9
- Hungarian 1,2,7,8,12-15,17,19, 23,24,35,37,39,43,54,47,49, 51,52,61,67,78, 80,81,88,89, 91,94
- Indian 9,10
- Indonesian 1
- Indo-Aryan 9
- Indo-European 2,10.
- Khasi 10
- Latin 10,96
- Malay 66,73
- Munda 9,10
- Odia 2,8,9,10,12-14,16-18, 29, 30,35,38,39,44,45,48,49,51, 64,65,69,70,78, 80,82,89,91, 94
- Persian 8,12-14,16-18,26,27,35, 38,39,43,45,48,49,51,63,69, 76,77,80,81, 86,87,91,93
- Portuguese 7,8,12-14,18,42,43,45
- Proto-Munda 10
- Russian 1,8,12-14,16-18,30,31, 35,38,39,44,45,48,49,51,52, 64,65,70,79,80, 82,90,91,94
- Slovak 2,7,8,12-15,17-21,34,36, 37,39,43,45,47,49,51,52,61, 67,72,74,75,80,81,84,85,91,93
- Spanish 96
- Telugu 9
- Turkic 9,10
- Turkish 2,8-10,12-14,17,18,30, 32,33,36,39,44,45,48,49,52, 65,70,71,79,80, 82,90,91,94

law 12,33,39,40,96,97
length 47,72
length-frequency 74
monotony 49
Morse function 19
Motif 73,74,83,95
nominality 1,40-52
normal test 5,14,34,36
Ord's criterion 80-83,93-95
power function 19,20,30
predicativity 2,40,41,44,49,74,83, 84, 95,97

property 2,95,97,
radian 13,46
random sequence 53
rank-frequency 72,73,83-92,95
regularity 33,46,47,49,50,53,98
requirement 97,98
runs 33-39,49-53
scaling 2,96
self-regulation 33
sentence 40,73
sequence, passim
similarity 14,66,73
specification 2,3,95,97
synergetic linguistics 1,40
thema-rhema 2,44
topic-comment 2,44,46
triad 46-49
typology 2
variability 45,46
Zipf distribution 74
Zipf-Alekseev function 30,74,83,91,96

The RAM-Publishing House edits since 2001 also the journal ***Glottometrics* –** up to now 31 issues – containing articles treating similar themes. The abstracts can be found in http://www.ram-verlag.eu/journals-e-journals/glottometrics/.

The contents of the last issue (31, 2015) is as follows:

Hanna Gnatchuk
Sound symbolism: Myths and reality 1 - 30

Emmerich Kelih, Gabriel Altmann
A continuous model for polysemy 31 - 37

Hanna Gnatchuk
Anglicisms in the Austrian Newspaper *KLEINE ZEITUNG* 38 - 49

Best, Karl-Heinz
Malay borrowings in English 50 – 53

Yu Fang, Haitao Liu
Comparison of vocabulary richness in two translated *Hongloumeng* 54 - 75

Wei Huang
Quantitative studies in Chinese language 76 – 83

Ruina Chen
Bibliography of quantitative linguistics of Chinese Researchers in International Academic Journals 84 - 88

Bibliography

Peter Grzybek, Emmerich Kelih
Glottometrics 1-30: Bibliography 89 - 102

Herausgeber – Editors of Glottometrics