Unified Modeling of Diversification in Language

by

Gabriel Altmann

2018

RAM-Verlag

Studies in Quantitative Linguistics

Editors

Fengxiang Fan (fanfengxiang@yahoo.com)
Emmerich Kelih (emmerich.kelih@univie.ac.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. Mačutek, 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 pp.
19. K.-H. Best, *Studien zur Geschichte der Quantitativen Linguistik.* 2015. III + 158 pp.

20. P. Zörnig, K. Stachowski, I.-I. Popescu, T. Mosavi Miangah, P. Mohanty, E. Kelih, R. Chen, G. Altmann, *Descriptiveness, Activity and Nominality in Formalized Text Sequences.* 2015. IV+120 pp.
21. G. Altmann, *Problems in Quantitative Linguistics Vol. 5*. 2015. III+146 pp.
22. P. Zörnig, K. Stachowski, I.-I. Popescu, T. Mosavi Miangah, R. Chen, G. Altmann, *Positional Occurrences in Texts: Weighted Consensus Strings.* 2015. II+178 pp.
23. E. Kelih, R. Knight, J. Mačutek, A.Wilson (eds.), *Issues in Quantitative Linguistics Vol. 4*. 2016. III + 231 pp.
24. J. Léon, S. Loiseau (eds.), *History of quantitative linguistics in France*. 2016. II + 232 pp.
25. K.-H. Best, O. Rottmann, *Quantitative Linguistics, an Invitation*. 2017, III+171 pp.
26. M. Lupea, M. Rukk, I.-I. Popescu, G. Altmann, *Some Properties of Rhyme,* 2017, IV + 134 pp.

ISBN: 978-3-942303-63-7

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

Preface

Quantification is a most natural development in all sciences. At the beginning, it is not very ripe but it is better to do something than to adhere to a tradition. One can distinguish several steps in quantification: (1) The use of quantitative concepts, beginning with decisions, and formalizations; then trying to ascribe to the phenomena some numbers or degrees; (2) measuring the phenomenon, especially properties. It must be noted that the properties – whatever they are – are our conceptual creations representing the reality "for us"; (3) expressing mathematically the fact found, and performing some kind of modeling; (4) testing the model on data – an enterprise that will never be finished; (5) joining the phenomenon with other phenomena, creating and deriving of hypotheses, and constructing a theory; and ultimately (6) the unification of the modeling and its simplification as far as possible.

We find all of these levels in current quantitative linguistics. In the present book we try to make a step in the last stage, namely to perform a simplification concerning classes of linguistic objects. We have collected some of the available published literature – it is enormous – and let the reader continue fitting the simple function we propose.

The approach does not mean that all the enedeavours by Zipf, Mandelbrot, etc. are not correct. On the contrary, we merely bring to them a simple differential equation which is easier to interpret than the more complex ones, but all of them belong to the same unified theory. Neither does it mean that we have discovered more truth. Truth is not concealed behind formulae. They merely help us to capture the reality and to draw mechanically consequences using the rules of mathematics. But none of our models is part of the reality.

We hope that those who have many sets of data at their disposal will test the given proposal. If the data are bell shaped, the simple formula does not work. But one can add something to the differential equation and interpret it synergetically. We are persuaded that all relations in language are very simple. We do not know them but we try to capture them *viribus unitis*.

Gabriel Altmann

Contents

1. Introduction

Diversification is a mechanical and necessary epistemological process. Living creatures must get orientation and are forced to distinguish objects. Several objects are ordered into the same class because of some similarity, and a classification emerges. Beings having a language do it explicitly in all domains but, unfortunately, in some sciences it has been conjectured that a scientific classification is the last stage of the cognition. Slowly, but incessantly, the view of classifications and orderings changes. In linguistics, it was G.K. Zipf who began to look at linguistic data from a different point of view.

Striving for mathematical models of linguistic phenomena is a trend that cannot be reversed any more. One states in all domains of language that the respective phenomena may be at least ranked because they occur with different frequencies in texts. But in many investigations, there are cases in which the phenomenon can be quantified in some other way. For ranking, one mostly used some variant of Zipf's (Zeta), Zipf-Mandelbrot's or Zipf-Alekseev`s distributions but if they did not give positive results, one set up a whole family of distributions and tested them using some software. This is a quite legitimate way of working because each text may contain some subsidiary conditions brought in by the requirements of the author. However, the number of models became too great; hence one began to search for unique solutions which were at the same time simpler. One stated that e.g. any kinds of length of linguistic units can be well captured by the Zipf-Alekseev function (cf. Popescu, Best, Altmann 2014). It is true that at the beginning one tried to find the "truth"; later on one stated that mathematical models are merely exact means enabling us to continue the epistemological processing in a formal way.

The use of the exponential function in this work does not mean that there is nothing else. It is merely a striving for the simplest model possible. Simplicity is not a property of reality but a property of our models. Whatever the complexity of our models, they are always simplifications. Here we shall try to adopt our simplified view to the diversification. Earlier one tried to find models for discrete variables in the form of discrete distributions; for continuous variables in the form of continuous distributions. Today, we know that this was rather an adhering to classical thinking and trying to find a function, a series or a distribution expressing the course of the data. One was always inclined to perform the ranking using natural numbers; today one asks why one did not perform a ranking in, say, the form 1.0, 1.1, 1.2.., etc. In Nature, there are neither natural numbers nor classes, and hence mathematical modelling is rather a mental play. One strives for good tests, for simplicity and for unification in the framework of a theory.

In linguistics we may differentiate between two types of diversification. The first one occurs within the given language and is caused by self-organization.

It is performed by the speaker who tries to form the speech in a way that is optimal for him, i.e. he wants to save effort of any kind. In this way, everything diversifies beginning from phonetics up to semantics. Allophones, allographs, allomorphs, allosemes, synonyms, polysemy, category classes, new words, etc. are steadily created; some of them survive, other ones will be eliminated. The elimination is performed by the self-regulation exerted by the hearer who wants to reduce his own effort. These procedures culminate in the birth of idiolects, sociolects, dialects, new languages, etc. Needless to say, even the script may diversify but this is performed by specialists who must care for the uniformity of rules of writing. Unfortunately, this is the more difficult the more people speak the given language. There have been several trials to change the writing of English words but since there are a number of diverging dialects, writing in English develops towards a hieroglyphic form in which the parts of words are represented by Latin letters. If one compares {gh} or {o} in words like *ghost*, *enough*, *through*, one may see that they have a phonetically diversified form. One may conjecture that intra-language diversification has a law-like course, as has been shown many times. We shall not consider here letter frequencies because there are/were also languages using signs. And we shall not analyze word frequencies because the amount of data is overwhelming and the number of investigations, too. We rather choose special cases and show a perspective for further research.

The second kind of diversification happens in an inter-language transfer. By evolution, not only different living organisms emerged but the developing Human himself saw reality in a very variegated way. Since we must express our way of seeing by words and words have their meanings, it may happen that even a special word in one language is semantically diversified in some other language. As a matter of fact, the semantics of a word diversifies both intra-linguistically and inter-linguistically. Words of one language have seldom only one meaning in another language, even if the languages are cognate. In a translation of a text from one language into another we find different expressions translating the respective unique one. Nevertheless, even here we conjecture the existence of some regularity, at least in the frequencies with which the given word is translated into different expressions in the other language.

We may ask: (a) is the diversification of entities of the same level realized in the same way? (b) do whether linguistic levels differ in their ways of diversification, e.g. grammar and semantics? (c) for each level, are there different diversification procedures when translating, e.g. between German and Dutch, between Beijing- and Shanghai-Chinese? (d) what are the laws of diversification for levels, intra- and inter-linguistics relations? (e) what is the background theory of diversification? (f) what are the boundary or subsidiary conditions, e.g. in daily and in professional conversation? and (g) are the linguistic diversifications analogous to those in other domains of nature? All these questions wait for answers.

Diversification is a phenomenon present in all domains of nature, not only in the living one. The existence of sets of elements classified by us is the best evidence. But the greatest difference is in the way we treat the diversity. In natural sciences, one tries to express everything mathematically and operates with the results formally. In social sciences, the situation is very different. Some of them do not even take it into account, other ones describe it with words, still other ones try to create classifications using one of the 500 quantitative methods, other ones characterize it with indicators (cf. Popescu, Altmann 2008), and still other ones, namely synergetic linguists, try to show its causes, its relation to other phenomena, its development, and its distribution,and try to construct systems of regularities and derive the phenomena of diversity using mathematical models (cf. Köhler 1986, 2005, Rothe 1991). Unfortunately, this is not always possible; many times one must postpone a solution. Rothe (1991b) found about eleven domains subdivided in further ones in which diversification occurs. We shall not consider all but try to use the counting results of the respective investigations.

When considering diversification, we always treat the classes in which the independent variable is mostly (but not necessarily) the rank of the given frequency. For this case, we are accustomed to use the Zipf(zeta) function, Zipf-Mandelbrot's distribution and a number of other ones (cf. Altmann 2005) which are able to capture the boundary conditions but, if possible, belong to the same family of functions. Mostly one tries to find the "best" fitting function not caring about its complexity. In the present work we shall try to find rather a simple unique model and leave it only if the frequencies are very deviating. The criterion of sufficiency is given by the determination coefficient $R^2 > 0.8$. In all cases, we use a function, not a normalized distribution.

The amount of data is enormous, mostly concerning phonemes, letters and word types or tokens. Here we shall focus on other domains, quote the sources and frequently give only the parameters of the functions and the determination coefficient, in order to save space. We will analyze rhythm, especially hexameter representing a special ordering of words, word classes, semantic diversification of individual words or categories and affixes, grammatical diversification, adnominals, borrowings, compounds, sentence types, Belza-chains, hrebs, musical data, etc.

We start from the assumption that the relative rate of change of ordered frequencies is constant and negative, i.e.

(1) $$\frac{y'}{y-1} = -b,$$

yielding the exponential function

(2) $$y = 1 + ae^{-bx}.$$

The 1 is added because the frequencies are never smaller than 1. If they are zero, one may simply omit them in the computation. The function has already been proposed by Fan, Altmann (2008), used in Popescu, Altmann, Köhler (2009) and in Tuzzi, Popescu, Altmann (2010) for the diversification of parts-of-speech in Italian presidential New Year addresses (see below). Evidently, the function is appropriate only for monotonically decreasing data but not only for ranked ones. Properties may be scaled differently. Any scaled data can be considered. Needless to say, the function is very universal but if one can simplify an approach one should do it.

There are a number of other proposals in which one focused on boundary conditions and applied mostly probability distributions which can be used any time if their fit is good and holds for a whole family of data. In some cases presented below, the differential equation (1) does not express a special intention of the writer which may be considered a subsidiary condition. In such cases, we may express it by adding a simple component representing the effort of the writer and obtain

(3) $$\frac{y'}{y-1} = -b + \frac{k}{x},$$

whose solution, after reparametrization yields:

(4) $$y = 1 + ax^c e^{-bx}$$

which is identical with the "Menzerathian" function extended by 1. Eliminating *b* from the equation (3) we obtain the Zipf-Mandelbrot function.

We shall try to show that this simplification which is at the same time a unification is sufficient for the fitting of many linguistic data. In the past, quantitative linguists strived for finding families of functions or distributions which were able to capture the trend of the data. There is a very rich literature. Today, the striving changes. Popescu, Best and Altmann (2014) showed that any kind of length data in language can be fitted by the Zipf-Alekseev function. In the present volume, we shall try to show that ranking and ordering may be captured, too, by a simple function. It does not mean that there are no "better" functions, or that one can put all the others aside. It merely means that we strive for simplicity and unification. The author himself applied a number of models in different previous publications. The present book does not imply that we are nearer to the truth if we use a simpler model; we merely simplify our view of the linguistic world. And we do not want to present a complete survey of some kind of data – it would be a very difficult task – we merely want to show the width of this phenomenon.

The number of possible classifications and orderings of linguistic phenomena is, so to say, infinite because we shall not cease to perform research, and classifications and orderings are its results. Our views develop and each step brings new data. If one seeks a model, one should begin with the simplest one which is well substantiated. The above models belong to the unified theory proposed by Wimmer and Altmann (2005).

2. Rhythm

If the rhythm of a text or a musical piece is constant, no investigation is necessary. But if different rhythmic units are possible, the rhythm may display some special tendency. Here we shall analyze only a restricted number of cases but each text could be analyzed in this way.

2.1. Hexameter

We begin with an entity which is greater than the word, namely with the hexameter. We shall analyze only Greek, Latin and German hexameters as they were presented by Drobisch (1866, 1872, 1875, 1868a,b) and Best (2008a,b). Omitting the last two feet in each verse which are always identical, we have 16 combinations of S(pondee) and D(actyl), i.e. all permutations with repetition of S and D in 4 positions (SSSS, SSSD, SSDS, SDSS, DSSS, SSDD, SDSD, SDDS, DSSD, DSDS, DDSS, DDDS, DDSD, DSDD, SDDD, DDDD). Since there is no logical ordering, we perform the ordering according to the frequency, that is, in each text the 16 types can be ordered differently. The data are presented in Table 2.1; the parameters and the determination coefficients are in the last three columns. We omit the computed values which can easily be restored using the computed parameters.

Table 2.1
Ordered frequency of hexameter feet (Drobisch; Best)

Text	**Observed values**	**a**	**b**	**R^2**
Goethe, *Reinecke Fuchs*	204,181,96,93,86,71,65,46,43, 42,38, 22,14,7,6,5	244.2568	0.2125	0.9559
Goethe, *Hermann und Dorothea*	200,149,142,104,98,73,63,41,40, 35, 34,31,25, 25,15,11	233.2567	0.1898	0.9869
Goethe, *Elegien*	96,92,72,51,45,37,32,28,19,11,8, 6,6,6,5,4	127.6750	0.2164	0.9818
Leibniz, *Epicedium*	59,57,33,31,30,24,20,16,13,12, 10,10,10,6,5,2	71.1188	0.1921	0.9655
Klopstock, *Messias*	129,128,125,102,76,66,65,60,60, 52,46,45,26,13,6,3	162.1247	0.1336	0.9323
Voss, *Luise*	188,173,164,161,154,135,131, 95,70,51,44,35,32,9	238.8932	0.1283	0.8899
Voss, *Odyssey*	125,123,114,98,96,91,86,67,65, 59,53, 39,31,16,4,1	156.5463	0.1161	0.9028
Vergil, *Georgica*	84,62,55,49,39,31,31,29,29,19, 19,18,16,12,11,9	88.4289	0.1515	0.9749

Vergil, *other sample*	78,75,57,52,44,38,37,33,30,27, 22,20,18,13,12,4	89.7732	0.1361	0.9799
Vergil, *Aeneis*	423,338,325,297,229,191,170, 167,167,136,117,106,102,66, 60,58	464.3002	0.1297	0.9823
Vergil, *Bucolica*	107,90,79,64,63,59,57,43,43,40, 38 29,27,26,23,21	111.4486	0.1105	0.9807
Horace, *Satires*	285,228,205,203,174,137,134, 115,108,102,93,92,79,63,48,46	300.6084	0.1136	0.9810
Horace, *an-other sample*	62,53,49,48,44,38,38,36,35,32, 30,22,22,20,20,11	64.6916	0.0841	0.9608
Horace, *Epistulae*	237,198,189,168,147,132,125, 124 109,109,97,94,89,59,54,36	247.5026	0.0940	0.9660
Lucrece, *De rerum natura*	88,72,63,56,51,39,37,36,26,21, 17,17,12,10,8,7	101.1521	0.1564	0.9894
Manilius, *Astronomica*	93,67,60,57,48,34,33,30,28,22, 22,19,16,11,11,9	99.3008	0.1555	0.9788
Persius, *Satires*	118,96,68,62,48,39,35,35,32,30, 27,19,14,12,9,5	132.2875	0.1824	0.9716
Juvenal, *Satires*	85,67,64,51,40,38,35,29,26,25, 22, 20,19,14,13,12	91.5714	0.1415	0.9836
Lucanus, *Pharsalia*	98,83,59,58,39,33,32,29,28,23, 19,15,13,12,11,8	111.9183	0.1797	0.9777
Quintus Ennius *Fragments*	64,39,39,35,25,24,24,24,23,21, 20,20,19,15,12,10	55.8726	0.1117	0.8525
Catull, *2 poems*	124,65,55,51,43,25,15,15,8,7,6, 5,4,3,3,1	154.7066	0.3183	0.9624
Ovid, *Meta-morphoses*	78,76,63,60,57,54,45,33,26,21, 13,11,6,6,6,5	100.9047	0.1520	0.9333
Silius Italicus, *Punica*	75,63,54,53,47,47,34,32,28,26, 25,24,18,16,11,7	81.8771	0.1205	0.9759
Valerius Flaccus, *Argonautica*	131,75,64,63,54,52,49,30,24,24, 22,21,12,11,5,3	133.7155	0.1854	0.9298
Statius, *Thebais*	83,76,57,53,43,40,39,34,32,24, 17,16,14,14,10,8	94.8034	0.1457	0.9826
Claudian, *Raptus Pro-serpinae*	102,83,75,67,51,38,34,30,24,21, 14,8,5,3,3,2	127.3123	0.1972	0.9823
Homer, *Iliad*	350,320,296,196,155,149,145, 78,76,73,56,28,22,21,19,8	456.4486	0.1966	0.9739
Homer,	410,323,277,185,176,161,149,	492.2354	0.2031	0.9857

Odyssey	92,82,71,65,35,34,20,18,5			
Theokrit, *1st Idyll*	31,21,13,11,10,9,7,5,5,5,4,2,1	37.2356	0.2916	0.9606
Theognis, *Elegic poems*	117,99,78,66,38,36,34,28,26, 25, 23,21,7,4,3,2	144.0192	0.2118	0.9733

As can be seen, for data containing a small number of different classes (here 16) the conjecture used in (1) is sufficient. The model of the exponential function is sufficient to capture the distribution of hexameter feet-types. Needless to say, the ranking may be different with different texts, i.e. the individual ranks are occupied by different patterns. Of course, if one compares the feet in individual works with those in other ones, one must compare the frequencies of the same feet. In the majority of cases the chi-square test may be performed (the classes should not be smaller than 5, otherwise one should pool cells with equal feet-types).

If one would ascribe to the spondees and the dactyls some numerical values, one would obtain a quite new type of ordering but up to now, no one has tried to go this way.

Another problem treated in Altmann, Köhler (2015: 137-141) is the study of distances between equal rhythmic patterns of hexameter which follows the well-known Skinner hypothesis. One can study the distances between each pattern separately if the poem is long enough, but one can count all equal distances. A problem of this kind has been analyzed in Strauss et al. (1984) who studied the distances between the rhythmic patterns DSSS in "Poems in Classical Prosody, Epistle II: To a Socialist in London" by R. Bridges and found the frequencies as presented in Table 2.2. We begin to count from x = 1 because "no distance" means "1 step is necessary to the next DSSS pattern".

Table 2.2
Distances between occurrence of the pattern DSSS in a poem by R. Bridges (Altmann, Köhler 2015)

Distance in steps	Frequency	Exp
1	17	16.64
2	13	11.37
3	4	7.88
4	4	5.56
5	6	4.02
6	3	3.01
7	6	2.33
8	2	1.88
9	2	1.58
10	1	1.39
11	2	1.26

12	2	1.17
14	1	1.07
34	1	1.00
a = 23.5944, b = 0.4109, R^2 = 0.8682		

We assume that a longer poem would have brought still better results. As a matter of fact, many hexameters should be analyzed in order to obtain a general trend. An older bibliography concerning hexameter is presented by U. Job (1981).

2.2. Feet

If one analyzes the representation of individual feet in texts where there is no rhythmic prescription, one may order them decreasingly, too. This can be done both with poetic and prose works. P. Grzybek and E. Kelih (2005) considered Tomaševskij's (1917) analyses of four Russian texts and published the percentages of individual feet. Since there is no prescription, we change the order in some places (i.e. the ranks do not represent the same feets in all texts) and obtain the results presented in Table 2.3. Fortunately, the exponential function may be used both for discrete and continuous frequency values.

Table 2.3
Representation of feet in some Russian works
(Tomaševskij 1917, Grzybek, Kelih 2005)

Text	Pushkin *Eugen Onegin*		Lomonosov *Oden*		Lermontov *Demon*		Pushkin *Pikovaja dama*	
Rank	%	Exp	%	Exp	%	Exp	%	Exp
1	26.8	29.31	29.4	29.94	28.6	31.12	17.7	17.79
2	26,5	20.55	23.3	20.88	27.3	20.87	15.4	15.20
3	11.7	14.50	11.5	14.65	11.1	14.11	14.5	13.02
4	10.9	10.33	11.3	10.38	9.1	9.654	9.4	11.17
5	6.3	7.44	7.3	7.44	6.5	6.71	8.6	9.60
6	4,5	5.45	5.6	5.42	4.0	4.77	8.3	8.28
7	3.6	4.07	4.9	4.04	3.9	3.49	7.5	7.15
8	3.3	3.12	2.8	3.09	3.6	2.64	7.2	6.21
	a = 40.9931 b = 0.3701 R^2 = 0.9213		a = 42.1424 b = 0.3757 R^2 = 0.9717		a = 45.6480 b = 0.4158 R^2 = 0.9213		a = 19.8476 b = 0.1673 R^2 = 0.9371	

The study of rhythm is an enormous domain with hundreds of publications. The above examples are sufficient to encourage researchers to study various aspects of rhythm (cf. https://en.wikipedia.org/wiki/Rhythm) not only in music but also in texts of whatever kind. Also a historical study is possible, as is the comparison of text types and the introduction of all necessary functions containing some subsidiary conditions.

Here, an extensive domain opens. Not only poems have a certain rhythm represented by placing of special feet at a special place but in every language there is some kind of accents which are not equal on each syllable. One can study the diversification in this domain concerning also prose texts. We may mention at least some of the problems concerning feet: (1) In any text one finds a special number of individual feet like trochee, jambus, anapest, dactyl, spondee, etc. Ordering them one obtains a decreasing sequence which can be modeled. (2) Are there differences between individual writers, text types, or languages? Can one use this property for typology? Are all languages equal? (3) Every writer develops and changes his writing customs; can one find an age-dependent difference in the numbers of different feet; or can one find a difference between his prose and poetic works? For direct comparison one need not order the data but compare the equal feet. (4) Every language develops. Can one find differences in the representation of feet? One can compare for example press texts over one whole century. Etc. We shall not try to establish this domain here because we are interested here only in diversification.

3. Parts of Speech (POS)

The number of parts-of-speech is smaller than that of hexameter feet. Usually one adheres to the antique classification but there are surely individual classes that are not present in all languages. Here we shall examine only those that have been used in Popescu, Altmann (2008) and published by Schweers, Zhu (1991) for Latin, German and Chinese, by Sambor (1989) for Polish, by Best (1994) for German, by Ziegler (2001) for Portuguese and Brazilian Portuguese. Skalmowski (1964) counted the Arabic verbal roots that can build forms in different classes (cf. also Altmann 1991). For Russian, Köhler (2012: 51) chose randomly a text from the Russian corpus. Wang (2016) analyzed the Modern Chinese Dictionary and found the ascription to POS. Best (1994, 2001) analyzed some stories from the books by Bichsel (1996), Bobrowski (1987), de Bruyn (1999), Kunert (1990) and Pestalozzi (1992), namely:

Pestalozzi: Hühner, Adler und Mäuse,
Bichsel 1; Der Mann der nichts mehr wissen wollte:
Bichsel 2: Und sie dürfen sagen, was sie wollen:
Bichsel 3: Lesebuchgeschichte
Bobrowski 1: Betrachtung eines Bildes
Bobrowski 2: Mäusefest
Bobrowski 3: Rainfarm
De Bruyn 1: Stallschreiberstraße 45
De Bruyn 2: Vergiß mein nicht
Kunert 1, Der Taucher
Kunert 2, Warum schreiben

The number of different POS in the above given languages is either 8 or 9 but this depends on the researcher and the linguistic school to which (s)he belongs. In any case, POS cannot be directly "measured" (there is no quantification up to now), but they can be ordered. Other analyses considering only selected POS (e.g. Levickij, Hikov 2004) were omitted. In many languages, the appurtenance of a word to a POS class is given by affixes in the text and in some languages a POS may belong to several classes simultaneously

The numerical data from 10 languages are presented in the Tables below

Table 3.1
Frequency of parts-of-speech

Language	**Data**	**a**	**b**	**R^2**
Polish (Sambor 1989)	144188, 79995, 71988, 56812, 33605, 31833, 21428, 18757, 8076, 650	187993.4040	0.3280	0.9677
Portuguese (Ziegler 2001)	2586,1607,949,819,776,680, 478,440, 352	3212.4383	0.3065	0.9296

Brazilian Portuguese (Ziegler 2001)	2930,2265,1743,1708,1602, 1040,936, 394	3516.3826	0.2004	0.9410
Chinese (Schweers, Zhu 1991)	247,228,140,133,107,81,55,27	331.3620	0.2474	0.9589
Chinese (Wang 2016)	29602,17773,5557,1157,460, 217, 203,163,100,85,69,56,	64464.3460	0.7462	0.9801
Arabic (Skalmowski 1964)	988,714,586,411,254,154,74, 18,10	1446.5702	0.3495	0.9776
Latin (Schweers, Zhu 1991)	347, 173, 142, 98, 93, 59, 40, 39, 9	471.6391	0.3879	0.9541
Russian (Köhler 2012)	182,63,54,50,19,14,11,10,4	313.6476	0.6130	0.9343
German (Best 1997)	313,262,193,163,150,104,56, 46,1	419.0872	0.2539	0.9568
German (Best 2001) *Pestalozzi*	32,24,19,17,15,12,9,8	36.7889	0.2107	0.9842
German (Best 1994) *Bichsel 1*	313,262,193,163,150,104,56. 45,1	419.0872	0.2539	0.9568
German (Best 1994) *Bichsel 2*	229,172,144,132,120,89,80,79	253.8127	0.1657	0.9658
German (Best 1994) *Bobrowski 1*	238,168,167,158,155,125,110, 98	243.11184	0.1127	0.8926
German (Best 1994) *Bobrowski 2*	192,174,150,141,88,79,67,42	244.5220	0.1848	0.9472
German (Best 1994) *Bobrowski 3*	274,164,161,142,140,138,109, 89	269.3371	0.1405	0.8076
German (Best 1994) *De Bruyn 1*	221,162,125,115,92,83,75,66	248.0664	0.1902	0.9590
German (Best 1994) *De Bruyn 2*	249,193,176,154,108,108,105, 63	286.8540	0.1707	0.9635
German (Best 1994) *Kunert 1*	261,185,148,139,121,99,85,65	203.2338	0.1909	0.9622
German (Best 1994) *Kunert 2*	169,166,159,112,92,90,85,68	206.2586	0.1374	0.9207

German (Best 1994) *Bichsel 3*	168,151,121,105,85,79,74,47	199.7286	0.1628	0.9806
German (Schweers, Zhu 1991)	192,161,153,112,111,104,97, 70	213.3540	0.1305	0.9541
German (Best 1994)	2032, 1939, 1532, 1338, 1179, 974, 914, 761	2419.8121	0.1446	0.9841

As can be seen, the results are very good; the exponential function expresses well the representation of individual POS in texts. Needless to say, in other texts other researchers could have found other results. Also, the fact that all words of the texts were taken into account and not only words in some special positions, e.g. rhyme, could be a relevant factor.

Tuzzi, Popescu and Altmann (2010) computed the ranking of parts-of-speech in all New-Year Addresses of Italian presidents and though they still adhered to the Zipfian function they computed also tacitly the exponential function. The results with the reparametrization are presented in Table 3.2 and 3.3.

Table 3.2
Frequencies of parts-of-speech
in Italian End-of-Year Addresses

Text	**Parts-of-speech frequencies**
1949 Einaudi	41,37,33,30,17,15,14,6,1
1950 Einaudi	42,36,20,15,15,9,8,4,1
1951 Einaudi	50,41,40,34,21,18,15,11
1952 Einaudi	46,35,28,27,13,12,11,7
1953 Einaudi	47,42,34,24,15,12,9,6,1
1954 Einaudi	57,54,43,36,20,18,17,14,1
1955 Gronchi	83,78,64,51,31,30,29,21,1
1956 Gronchi	180,121,88,87,71,58,34,26
1957 Gronchi	267,241,170,126,93,84,79,59,6,5
1958 Gronchi	201,162,131,127,82,74,63,42,3,1
1959 Gronchi	181,135,92,80,72,71,36,29,1
1960 Gronchi	196,161,112,106,78,63,45,38,3,2
1961 Gronchi	304,244,184,162,111,105,75,65,2
1962 Segni	196,147,120,83,73,54,36,29
1963 Segni	257,219,170,131,93,68,52,45,14,8
1964 Saragat	102,85,84,64,42,40,28,17,3
1965 Saragat	267,211,141,138,85,79,78,45,6,3
1966 Saragat	324,239,185,144,109,75,66,50,5,2
1967 Saragat	263,207,167,145,96,64,59,36,14,3,2
1968 Saragat	304,243,176,134,95,86,70,56,8,2

1969 Saragat	394,284,232,222,165,103,99,72,8,3,2
1970 Saragat	490,389,272,257,186,113,112,86,17,5,2
1971 Leone	70,51,37,35,30,17,11,6,3,2
1972 Leone	182,149,134,111,69,45,45,24,5,3
1973 Leone	298,232,205,174,103,97,76,63,1,1
1974 Leone	197,141,139,120,66,59,42,35,1,1
1975 Leone	312,244,200,191,122,97,91,69,2
1976 Leone	321,239,211,196,113,112,97,73,3,1
1977 Leone	358,270,262,216,142,122,115,113,4,2
1978 Pertini	332,283,248,156,130,125,106,86,17,10
1979 Pertini	499,442,345,279,219,201,184,115,8,8,2
1980 Pertini	316,244,228,164,121,104,101,61,10,9,2
1981 Pertini	571,571,377,331,261,231,227,196,38,14,1
1982 Pertini	509,495,332,322,233,202,172,139,62,19,2
1983 Pertini	786,760,510,452,360,308,275,206,55,33,3
1984 Pertini	302,269,197,163,129,97,95,51,20,17
1985 Cossiga	612,427,404,289,207,192,120,93,10,3,2
1986 Cossiga	321,232,215,187,130,106,79,77,1,1
1987 Cossiga	501,414,349,248,184,163,107,103,11,11
1988 Cossiga	557,467,369,311,199,183,146,134,14,5
1989 Cossiga	441,399,302,231,154,145,102,101,31,6
1990 Cossiga	800,646,534,396,305,277,173,163,35,18
1991 Cossiga	95,71,64,57,48,29,26,22,4,2
1992 Scalfaro	656,472,435,360,250,231,208,151,4,3,2
1993 Scalfaro	684,501,469,387,247,236,218,168,22,8,1
1994 Scalfaro	866,633,590,482,284,267,248,207,15,12,1
1995 Scalfaro	994,741,682,523,357,332,290,246,38,22,3
1996 Scalfaro	535,348,326,313,183,128,115,110,16,11
1997 Scalfaro	1113,1048,712,522,429,397,368,329,54,33,10
1998 Scalfaro	972,775,577,415,399,289,254,251,35,23,5
1999 Ciampi	504,347,291,278,206,110,89,82,24,9,1
2000 Ciampi	432,338,291,273,168,124,95,88,23,12
2001 Ciampi	549,395,338,262,224,109,96,89,18,15,2
2002 Ciampi	556,389,312,304,209,132,112,98,10,7
2003 Ciampi	408,297,231,214,142,112,79,75,4,2,1
2004 Ciampi	455,353,268,265,147,111,93,88,19,8
2005 Ciampi	290,235,181,166,114,89,55,40,12,10,1
2006 Napolitano	502,377,356,286,191,169,159,146,10,7,1
2007 Napolitano	419,352,274,242,144,123,115,104,13,5,3
2008 Napolitano	409,328,281,220,135,127,120,86,4,2,1

Table 3.3
Fitting the exponential function to Italian parts of speech

Text	a	b	R^2
1949 Einaudi	55.4131	0.2270	0.9054
1950 Einaudi	59.6584	0.3345	0.9664
1951 Einaudi	56.1011	0.2981	0.9106
1952 Einaudi	57.2212	0.3552	0.9477
1953 Einaudi	66.9476	0.2875	0.9595
1954 Einaudi	78.2885	0.2402	0.9330
1955 Gronchi	112.0836	0.2259	0.9300
1956 Gronchi	214.7349	0.3653	0.9597
1957 Gronchi	364.2917	0.2688	0.9584
1958 Gronchi	262.1934	0.2346	0.9408
1959 Gronchi	228.9571	0.2616	0.9481
1960 Gronchi	260.2692	0.2634	0.9645
1961 Gronchi	393.3201	0.2482	0.9669
1962 Segni	248.7756	0.3754	0.9816
1963 Segni	352.5834	0.2860	0.9769
1964 Saragat	134.6237	0.2345	0.9270
1965 Saragat	350.8234	0.2744	0.9615
1966 Saragat	436.2819	0.2959	0.9840
1967 Saragat	358.7014	0.2774	0.9747
1968 Saragat	410.9684	0.2856	0.9798
1969 Saragat	512.3810	0.2627	0.9575
1970 Saragat	662.1603	0.2794	0.9738
1971 Leone	92.6163	0.3028	0.9644
1972 Leone	250.6557	0.2669	0.9483
1973 Leone	396.2084	0.2521	0.9510
1974 Leone	260.6625	0.2596	0.9394
1975 Leone	398.4730	0.2343	0.9506
1976 Leone	412.1600	0.2394	0.9371
1977 Leone	456.3595	0.2255	0.9205
1978 Pertini	434.2952	0.2468	0.9481
1979 Pertini	671.3088	0.2342	0.9402
1980 Pertini	417.2399	0.2496	0.9573
1981 Pertini	774.9393	0.2197	0.9208
1982 Pertini	689.4811	0.2228	0.9476
1983 Pertini	1068.0115	0.2295	0.9469
1984 Pertini	398.3592	0.2692	0.9640
1985 Cossiga	812.9602	0.2787	0.9670
1986 Cossiga	412.0190	0.2423	0.9426

1987 Cossiga	676.8828	0.2706	0.9666
1988 Cossiga	744.3419	0.2524	0.9606
1989 Cossiga	605.9822	0.2600	0.9673
1990 Cossiga	1058.5020	0.2653	0.9722
1991 Cossiga	120.0938	0.2367	0.9410
1992 Scalfaro	840.8369	0.2444	0.9418
1993 Scalfaro	880.3619	0.2405	0.9453
1994 Scalfaro	1131.8520	0.2527	0.9498
1995 Scalfaro	1288.6011	0.2468	0.9575
1996 Scalfaro	675.5472	0.2722	0.9426
1997 Scalfaro	1508.0736	0.2480	0.9437
1998 Scalfaro	1273.9222	0.2620	0.9653
1999 Ciampi	659.4942	0.2747	0.9634
2000 Ciampi	566.8884	0.2594	0.9542
2001 Ciampi	736.5069	0.2854	0.9746
2002 Ciampi	720.7311	0.2787	0.9622
2003 Ciampi	541.8572	0.2801	0.9703
2004 Ciampi	606.5151	0.2795	0.9653
2005 Ciampi	395.4018	0.2675	0.9719
2006 Napolitano	649.5857	0.2367	0.9381
2007 Napolitano	567.3769	0.2584	0.9584
2008 Napolitano	553.2141	0.2602	0.9571

In the first years of the presidential office, the parameter *b* had a certain oscillation but in the later years it seems to obtain a steady form. Of course, its value should be stated also in other texts. It is to be noted that the individual parts-of-speech do not have always the same rank but as a whole they behave according to the “prescription”.

Besides the New-Year Speeches of Italian presidents, Overbeck et al. (2010) considered also the POS in librettos by Italian authors, namely in:

L1. Alessandro Striggio the Younger (1573-1630), *La favola d’Orfeo* 1607
L2. Aurelio Aureli (ca. 1630- ca. 1709), *L’Orfeo* 1672
L3. Apostolo Zeno (1668-1750), *Teuzzone* 1719
L4. Pietro Metastasio (1698-1782), *L’Olimpiade* 1733
L5. Giambattista Varesco (1736-1805), *Idomeneo, re di Creta* 1781
L6. Gaetano Rossi (1774-1855), *Semiramide* 1823
L7. Felice Romani (1788-1865), *Norma* 1831
L8. Francesco Maria Piave (1810-1876), *Ernani* 1844
L9. Arrigo Boito (1842-1918), *Otello* 1887

and obtained the results won by fitting the above exponential function as shown in Table 3.4

Table 3.4
Rank-frequency sequences of POS in Italian librettos: 1607-1887
(Overbeck et al. 2010)

Author	Frequencies	a	b	R^2
Striggio	847,598,516,498,441,349,311,255, 79,72	981.2854	0.1990	0.9322
Aureli	1740,1737,1210,931,870,705,574, 480,253,107	2324.9707	0.2113	0.9576
Zeno	1632,1561,1215,821,792,695,547, 508,123,116	2166.8616	0.2114	0.9462
Metastasio	1959,1560,1403,921,900,768,676, 458,234,216	2429.3272	0.2096	0.9743
Varesco	957,775,651,463,413,338,309,234, 159,121	1181.6911	0.2114	0.9923
Rossi	991,934,845,598,480,393,333,256, 192,97	1323.9338	0.2001	0.9642
Romani	1035,846,773,434,372,367,318,199 193,88	1328.1702	0.2298	0.9664
Piave	942,903,615,446,403,391,304,152, 99,77	1265.4736	0.2322	0.9587
Boito	1342,1209,968,643,540,401,369, 293,128,114	1815.7245	0.2414	0.9787

An historical study of the parameter *b* would be interesting but to this end one would be forced to analyze many more texts than we have up to now – in different languages.

A. Ziegler (1998) performed a POS-analysis in Brasilian-Portuguese press text. The authors of the texts are named in his article; here we shall take merely the numbers. He found 9 POS classes (verb, noun, adjective, pronoun, adverb, preposition, conjunction, numeral, article) and computed the frequencies in 21 texts, as given in Table 3.5. Ziegler applied the mixed Poisson and the negative hypergeometric distributions which were not satisfactory in several cases. This is, of course, no reason for rejection because in each text there may be some subsidiary conditions which may destroy the "usual" course of numbers.

Table 3.5
POS classes in Brazilian-Portuguese press texts (Ziegler 1998)

Text No.	Frequencies of POS classes	a	b	R^2
1	168,165,103,82,67,52,41,34,11	232.5036	0.2548	0.9506
2	119,61,55,50,44,32,18,6,2	146.9609	0.3094	0.9159
3	255,141,130,88,80,75,40,32,2	320.7427	0.3087	0.9443
4	689,368,208,174,157,138,123,114,55	931.2671	0.3920	0.9123
5	238,144127,115,84,79,72,50,5	277.0486	0.2395	0.9229

6	308,197,193,105,96,82,79,68,50	371.0922	0.2541	0.9445
7	392,221,154,129,126,114,69,62,23	480.0875	0.3066	0.9264
8	338,250,147,142,139,104,72,66,15	421.9795	0.2656	0.9491
9	163,94,91,84,71,55,39,33,1	189.1529	0.2357	0.9060
10	129,90,54,36,33,27,19,17,11	179.9481	0.3613	0.9821
11	383,267,137,132,112,106,104,103,64	451.4686	0.2643	0.8730
12	128,123,98,59,52,34,34,32,1	179.8989	0.2534	0.9360
13	391,223,168,140,139,109,92,85,42	468.6189	0.2851	0.9245
14	496,276,164,156,138,132,71,69,36	656.9504	0.3643	0.9296
15	457,253,181,128,97,91,86,83,55	610.3322	0.3666	0.9526
16	298,151,97,96,80,78,60,47,37	371.6166	0.3420	0.9889
17	200,143,85,83,81,61,36,28,5	263.8475	0.3061	0.9332
18	203,100,97,88,70,53,46,25,10	237.4213	0.2764	0.9071
19	223,115,84,83,78,74,42,29,8	281.6976	0.2916	0.8828
20	146,141,85,75,70,59,59,38,8	186.2555	0.2106	0.9156
21	117,81,80,54,50,41,35,15,2	145.8296	0.2423	0.9424

For the evaluation of Portuguese, Ziegler (2001) used 20 texts. The data and the results are presented in Table 3.6

Table 3.6
POS classes in Portuguese press texts (Ziegler 2001)

Text No.	**Frequencies of POS classes**	**a**	**b**	**R^2**
1	96,67,47,45,34,30,28,20,19	111.1577	0.2323	0.9594
2	86,57,37,31,30,28,26,18,13	98.6781	0.2499	0.9163
3	100,67,45,34,30,27,21,14,11	127.4099	0.3055	0.9738
4	119,73,41,37,30,28,24,7,6	161.5774	0.3711	0.9534
5	115,83,54,48,39,38,36,33,30	127.6730	0.2132	0.9001
6	126,101,54,47,45,36,26,24,19	161.2315	0.2781	0.9537
7	92,74,52,48,37,33,28,23,8	112.7431	0.2260	0.9744
8	135,93,56,45,44,30,21,18,16	177.9563	0.3238	0.9745
9	131,99,70,49,47,46,46,45,42	141.4379	0.1886	0.8501
10	134,79,63,53,41,38,18,10,9	171.4751	0.3158	0.9646
11	164,107,61,45,43,36,23,20,8	228.4136	0.3749	0.9723
12	178,101,68,61,61,52,39,26,25	208.3041	0.2833	0.9063
13	140,91,46,45,37,34,24,22,13	184.6351	0.3458	0.9426
14	116,79,38,27,22,22,16,11,6	175.8951	0.4351	0.9726
15	163,100,39,35,35,25,22,18,17	251.4516	0.4726	0.9389
16	111,66,65,52,44,43,32,19,2	128.9268	0.2354	0.9132
17	108,65,40,39,34,27,24,17,11	132.1662	0.3050	0.9342
18	145,53,33,29,24,21,15,12,10	253.9931	0.6464	0.9162
19	146,84,53,45,43,34,22,14,10	192.4778	0.3574	0.9537
20	181,79,49,37,34,32,28,26,26	267.9148	0.4366	0.8750

Here, too, the formula yields good results.

Zhu and Best (1992) analyzed juridical and literary texts in Chinese and counted the numbers of mono- and bisyllabic words subdivided into parts of speech. The results are presented in Table 3.7. In the juridical texts there were no numbers or numerals as POS, and hence, there were only 7 POS classes.

Table 3.7
POS in Chinese juridical and literary text (Zhu, Best 1992)

Texts	**Frequencies**	a	b	R^2
Juridical texts	253,210,107,69.55,54,18	388.6421	0.3901	0.9671
Literary texts	227,194,173,105,78,44,32,20	325.5462	0.2864	0.9505

K.-H. Best (2001) analyzed 30 German press texts from *Der Spiegel*, He took into account only 8 POS-classes. In the texts from the *Eichsfelder Tagesblatt* the date of the edition is written in German style.

Table 3.8
POS classes in German press texts (Best 2001)

Der Spiegel				
Text	**Frequencies**	**a**	**b**	$\mathbf{R^2}$
Angebot an die Union (49,1988, p.18)	69,31,28,26,20,9,9,9	88.6590	0.3696	0.9039
Strafe für eilige Kunden (49,1988,p.17)	51,35,20,19,16,15,15,9	61.6432	0.2843	0.9157
Kranker Norden (50,1999,p.187)	67,35,29,28,26,23,14,7	75.7980	0.2664	0.8698
Gescheiterte Versöh nung (50,1999,p.189)	57,31,24,24,15,12,5,4	75.4554	0.3624	0.9504
Bedingungen für Leos (48,1999,p.17)	57,28,26,18,14,11,9,6	74.9058	0.3732	0.9440
Zweifel an Hillary (48, 1999, p.180)	84,32,31,31,27,15,13,13	99.8543	0.3416	0.8179
Lord der Lüge (48,1999,p.179)	92,45,36,34,31,21,18,17	107.4338	0.3012	0.8736
Sorge um Waffen-Deal (48,1999,p.180)	65,25,24,24,21,11,11,7	78.8667	0.3556	0.8337
Asiatisches Tauwetter (50,1999,p.188)	42,26,23,19,17,13,12,7	48.0123	0.2418	0.9481
Imagepflege für Riester (50,1999,p.18)	39,14,12,11,9,7,7,5	54.2080	0.4745	0.8433
Eichsfelder Tagesblatt				
Liebe Bürger, tretet zurück	36,31,26,22,21,16,14,12	41.2719	0.1627	0.9924

(16.12.99,p.9)				
Immer dünner (16.12.99,p.9)	52,36,32,28,21,21,17,13	58.1430	0.1977	0.9640
Pädagogisch wertvoll (18.12.99,p.13)	26,25,25,23,17,17,13,12	30.3514	0.1141	0.8937
Hilfe, Gefahr für Hessen (6.12.99,p.7)	39,25,17,15,15,15,12,8	42.9401	0.2382	0.8782
Kein Geld für Tiere (14.12.99,p.7)	51,40,29,24,21,21,18,16	57.3212	0.1948	0.9455
Zukunft zerredet (15.12.99, p.7)	59,28,26,22,21,18,11,11	66.9577	0.2769	0.8541
Ein System steht gerade (2.12.99,p.9)	34,33,28,24,20,15,14,13	41.1471	0.1567	0.9679
Grenzen der Instal-lation (7.12.99,p.7)	36,28,21,16,16,14,13,13	39.5595	0.1900	0.9241
Nadel- oder Nudel holz (20.12.99,p.7)	45,31,31,25,25,20,19,18	45.9067	0.1404	0.9094
Krieg der Lichter (21.12.99,p.7)	37,26,26,25,17,15,12,8	42.2023	0.1883	0.93454
Großprojekt ohne Tabus (18.12.99,p.1)	46,22,21,21,17,6,4,3	58.3004	0.3452	0.8875
Fangquoten werden erheblich gekürzt (18.12.99, p.3)	47,28,25,25,24,9,8,6	55.8876	0.2597	0.8845
Das Millenium ist in aller Munde (18.12.99,p.12)	56,24,22,16,14,9,9,7	74.7012	0.4100	0.9062
Fronten bleiben starr (20.12.99,p.2)	42,18,18,17,16,6,4,3	52.4455	0.3572	0.8640
Zwei AKW vom Netz? (22.12.99,p.1)	31,22,13,12,8,7,6,6	40.8265	0.3370	0.9715
“Gespräche am Ende” (24.12.99,p.2)	61,31,29,22,16,14,7,5	79.1755	0.3546	0.9468
Weihnachten wird nass (21.12.99,p.12)	44,22,20,16,15,13,9,8	50.1392	0.2860	0.8842
Deutsche spenden wie nie (27.12.99,p.1)	59,29,25,20,17,9,8,4	78.6369	0.3839	0.9441
Müller will Zeichen Setzen (27.12.99,p.2	50,35,25,19,16,14,11,10	62.3293	0.2850	0.9788
Eichelhäher “pflanzt” fleißig Bäume (27.12.99,p.6)	68,27,26,24,16,14,13,10	83.3429	0.3600	0.8467

K.-H. Best (2000) studied the POS in magazine advertisements and fitted the negative hypergeometric distribution to the data with very good results. The sources can be found in his article. In Table 3.9 we display his data and the fit of the exponential function. It must be remarked that these texts are very short and purposeful, and each of them contains an unknown subsidiary condition inserted according to the kind of advertisement. Some of the texts have a determination coefficient smaller than 0.8 (T 1, T 4 and T 11). We may try to explicate this anomaly as a strong interference of some editors of the journal in the original text. At the same time, it would be a good stimulus for earching for searching for other functions.

Table 3.9
Parts-of-speech frequencies in German magazine advertisements
(Best 2000)

Text	**Frequencies**	**a**	**b**	R^2
T 1	25,12,11,10,10,10,7,1	26.9315	0.2683	0.7822
T 2	48,26,17,14,13,12,8,4	63.8676	0.3861	0.9349
T 3	28,17,13,7,6,5,4,4	39.9528	0.4190	0.9827
T 4	25,12,11,10,8,8,7,7	25.8247	0.2528	0.7898
T 5	12,11,10,7,7,7,3,3	14.3092	0.1954	0.8924
T 6	16,9,7,7,7,5,4,2	17.8158	0.2849	0.8860
T 7	9,9,7,7,6,6,5,3	9.7378	0.1372	0.8888
T 8	19,16,13,10,10,7,5,1	24.0279	0.2460	0.9315
T 9	20,13,12,11,10,9,3,2	23.0370	0.2390	0.8732
T 10	22,11,11,10,9,7,7,4	22.8653	0.2390	0.8390
T 11	33,10,10,10,8,7,7,5	42.4792	0.4437	0.7478
T 12	19,17,16,14,13,11,9,2	22.6490	0.1613	0.8144
T 13	32,14,12,11,11,9,8,5	35.5355	0.3056	0.8052
T 14	18,17,15,11,10,8,7,5	21.8823	0.1844	0.9656
T 15	22,16,9,8,7,6,4,3	28.9367	0.3434	0.9709
T 16	35,20,14,13,12,10,9,8	39.6354	0.2788	0.8815
T 17	21,14,14,12,12,11,9,5	21.3898	0.1569	0.8672
T 18	39,16,15,14,11,11,8,7	44.1998	0.3184	0.8086
T 19	29,18,15,13,10,9,8,6	33.1230	0.2561	0.9500
T 20	19,18,13,10,9,6,5,2	25.2548	0.2574	0.9528
T 21	22,17,13,12,11,10,9,7	23.3109	0.1725	0.9514
T 22	19,17,12,9,6,6,4,4	25.1438	0.2814	0.9761
T 23	80,34,31,22,21,18,16,2	104.4691	0.3925	0.8798

J. Laufer and E. Nemcová (2009) studied the language of SMS. They found the rank order of individual POS as given in Table 3.10. It is to be remarked that the authors used for some fittings also the exponential function as shown above.

Table 3.10
POS in German SMS (Laufer, Nemcová 2009)

Origin	Frequencies	a	b	R^2
German SMS	2815, 2550, 2316, 1606, 1459, 767, 541, 175	3938.1901	0.2384	0.9162

For Czech, M. Kubát (2016) performed a text type analysis of POS. Using a great number of texts, he studied the occurrence of 10 individual POS. The results are presented in Table 3.11. The ranking is not equal for all text types; we show them in decreasing order. As can be seen, the model functions also with large numbers. The determination coefficient is very satisfactory.

Table 3.11
Ranking of POS frequency in Czech text types (Kubát 2016)

Text type	Frequencies of POS	a	b	R^2
Novel	83039, 79208, 59441, 36517, 35111, 28376, 28128, 4328, 3303, 644	118391.2470	0.2609	0.9328
Short story	33287, 31091, 26235, 14810, 13378, 10772, 9831, 1831, 941, 368	48169.1247	0.2658	0.9346
Travel book	22948, 14502, 11773, 10842, 10796, 8230, 7500, 924, 481, 95	27946.9195	0.2515	0.8963
Study	18750, 10529, 9745, 8193, 6768, 5341, 5298, 393, 323, 25	23301.0977	0.2854	0.9196
Press	15788, 12034, 8826, 6811, 6237, 6007, 5096, 815, 421, 109	20525.0095	0.2630	0.9384
Fairy tale	10910, 9871, 7571, 5467, 4366, 3753, 2697, 730, 297, 169	15459.2164	0.2658	0.9573
Letter	9853, 7986, 7796, 5148, 4860, 3526, 3501, 522, 403, 53	13311.8412	0.2381	0.9179
Poetry	904, 768, 588, 375, 316, 272, 250, 42, 22, 17	1267.1526	0.2840	0.9620

For Chinese, L.Wang (2016) obtained the results presented in Table 3.12. She used the right truncated modified Zipf-Alekseev distribution but the result was not persuasive. In Table 3.12 we again apply the exponential function.

Table 3.12
Number of words in Chinese POS classes (Wang 2016)

Ranked classes	Number of words	Exp
1	29602	30567.64
2	17773	14494.52

3	5557	6873.27
4	1157	3259.57
5	460	1546.09
6	217	733.62
7	203	348.38
8	163	165.71
9	100	79.10
10	85	38.03
11	68	18.57
12	36	9.33
a = 64464.6379, b = 0.7462, R^2 = 0.9801		

For the POS in Chinese there are a number of sources. Here, we shall use the numbers given by B. Yin (1986) and used by C. Schindelin (2005). Yin did not tell which dictionary was used but older counts are surely available. The results are presented in Table 3.13.

Table 3.13
Frequencies of POS in Chinese (Yin 1986; Schindelin 2005)

Number of words in individual POS classes	**a**	**b**	**R^2**
23267, 11603, 3116, 239, 202, 86, 83, 74, 64, 37, 32	57024.3685	0.8779	0.9879

According to the type of language, a word can belong to several POS classes at the same time, In this way classes of POS-combinations are created, e.g. in German each verb can be at the same time a noun, or each adjective can be an adverb, etc. In English, this phenomenon is well known. L. Wang (2016) created all classes belonging to 2,3,4,5 and 6 POS simultaneously, e.g. there are words belonging at the same time to nouns, verbs, adverbs and prepositions. The counts are presented in Table 3.14. There are only two words belonging to 6 classes at the same time and we omit them. Wang's Table 3a is not clearly printed but it can be restored.

Wang's procedure is the best image of the subjectivity of our classificatory thinking. Even if a whole linguistic school accepts a given classification, it does not mean that we are nearer to the reality ot the truth. We stick with tradition. The only corroborating entities are the results of some tests performed quantitatively using special hypotheses. But if exceptions occur, one must check the data, the hypotheses, the models and its derivation, the computation, etc. Each exception extends the theory. Table 3.14 is to be read e.g. "There are 1630 words belonging to the noun+verb class; there are 540 words belonging to the noun+adjective class", etc. The combinations are not given but it is sure that in every language, we obtain different numbers expecting that they will behave according to our model.

Table 3.14
Number and kinds of combinations of Chinese POS
to which the words may belong (Wang 2016)

Number of POS classes	**Number of words**	**a**	**b**	**R^2**
2	1630, 540,459,143,136,93,78,27, 19,19,17,16,12,12,11,9,8,8,8,7,6, 5,4,4,3,3,3,3,2,2,2,2,1,1,1,1,1,1,1	3532.8318	0.8018	0.9775
3	86,85,37,31,20,14,9,9,6,6,6,4,4,4, 4,4,3,3,3,3,3,2,2,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1	132.9350	0.3611	0.9596
4	16,10,4,3,2,2,2,2,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1	29.0837	0.6474	0.9857
5	3,2,2,1,1,1,1,1,1,1,1,1,1	3.8321	0.6383	0.9333

It can be conjectured that no writer is focused on the POS he uses but to the contents he wants to express. Hence, the counts considered are true random phenomena. Be warned of mixing text; consider corpora as wholes. If the function holds true inspite of mixing one must reflect on the the role of its parts, on the role of subsidiary conditions, on the result of mixing concerning some kinds of words, etc.

Wang's numbers can also be added and one obtains an overall table of polyfunctionality presented in Table 3.15. As above, the last class (6) contains only 2 cases and was omitted. Wang used here the well fitting Waring distribution.

Table 3.15
Polyfunctional classes in Chinese (Wang 2016)

Polyfunctionality	**Number of words**	**Exp**
1	47414	47413.33
2	3299	3317.97
3	364	233.065
4	60	17.245
5	17	2.146
6	2	1.08
$a = 677705.4950$, $b = 2.6598$, $R^2 = 1.0000$		

Though in the last classes, the difference between the observed and the computed numbers is rather large, the fit can be accepted. It would be better to study individual texts and show all the results separately.

If one considers the number of properties that can be found in POS, one sees at once that 10 languages are too few to demonstrate something. There are surely many data sets that were not accessible to the author but the present demonstration is sufficient.

4. Morphology

It is well known that the morphemes expressing a special category, e.g. case, are not unique in highly synthetic languages where they must be adopted also into other categories, e.g. gender.

The 9 morphemes used as a plural ending in German have been analyzed by N. Brüers and A. Heeren (2004) in the letters of the German writer Heinrich von Kleist (1777-1811) in the years 1800 and 1801. Some letters contain more different endings than other ones but in general one can expect many exceptions. In Letter 4 we added a zero because the number of classes was too small. In Letter 11, the endings are almost identical, in any case their number is too small. Another investigation has been performed by K. Meuser, J.M. Schütte and S. Stremme (2008) using the short stories by W. Schnurre (1920-1989). The results are presented in the second part of Table 4.1. Two cases consisting of few plural endings give no satisfactory results but this is surely caused by the too small sample.

Table 4.1
Plural endings in German (Meuser, Schütte, Stremme 2008)
(First part: Texts by H. von Kleist, second part: Texts by W. Schnurre)

Type	**Frequencies**	**a**	**b**	**R^2**
H.v. Kleist				
Letter 1. Würzburg, September 13,1800 to Wilhelmine von Zenge	17,7,6,5,3,1	28.3951	0.6206	0.9395
Letter 2. Würzburg, September 14, 1800 to Wilhelmine von Zenge	8,6,4,3,2,1,1	11.8448	0.4694	0.9751l
Letter 3. Würzburg, September 15,1800 to Wilhelmine von Zenge	10,3,2,1,1,1,1	36.6268	1.4062	0.9959
Letter 4. Würzburg, September 18,1800 to Wilhelmine von Zenge	6,3,1, (0)	17.6003	1.2406	0.9176
Letter 5. Würzburg, September 19,1800 to Wilhelmine von Zenge	12,12,4,4,2,1	19.9582	0.4843	0.8511
Letter 6. Würzburg, Septem-ber 23,1801 to Heinrich Lohse	6,3,2,1	12.5198	0.9160	0.9898
Letter 7. Berlin, October 27, 1800 to Ulrike von Kleist	5,4,2,2	6.8738	0.5066	0.9204
Letter 8. Berlin: November 22,1800 to Wilhelmine von Zenge	9,3,2,1	29.1396	1.2956	0.9942
Letter 9. Berlin, November 27,1800 to Ulrike von Kleist	11,7,5,4,4,2	14.4043	0.3976	0.9665
Letter 10. Berlin, January 21, 1801 to Wilhelmine von Zenge	6,4,4,2,1	8.1299	0.4678	0.8864

Letter 11. Berlin, March 23, 1801 to Ulrike von Kleist	5,2,2,2,2,2			0.6800
Letter 12. Berlin, 1. April 1801 to Ulrike von Kleist	16,10,4,3,2,2,1	29.6558	0.6626	0.9850
Letter 13. Berlin: April 14,1801 to Wilhelmine von Zenge	8,6,3,2,1	13.2968	0.5936	0.9532
Letter 14. Dresden: May 4, 1801 to Wilhelmine von Zenge	4,4,3,2,1	4.9579	0.3728	0.8029
Letter 15. Straßburg: June 28, 1801 to Wilhelmine von Zenge	7,5,2,1	12.9476	0.7246	0.9240
Letter 16. Paris: October 27, 1801 to Wilhelmine von Zenge	6,4,2,2,2	8,8778	0.5754	0.9504
Letter 17. Paris. November 1801 (probably) to Ludwig von Brockes	7,6,4,3	8.7922	0.3424	0.9542
Letter 18. Paris, November 1801, to Adolfine von Werdeck	18,17,10,5,5,4, 4,2	26.4896	0.3618	0.9330
Letter 19. Frankfurt a.M., November 29, 1801, to Adolfine von Werdeck	3,3,3,3,1			0.3984
Letter 20. Basel, December 16, 1801, to Ulrike von Klerist	7,2,1,1,1	37.9247	1.8432	0.9991
Letter 21. Würzburg, September 20, 1800 to Wilhelmine von Zenge	23,10,8,8,6,4	31.9958	0.4622	0.8853
W. Schnurre				
Als Vater sich den Bart abnahm	9,9,7,6,5,5,3,1	11.0780	0.2175	0.8753
Fritzchen	28,16,14,13,8,3	36.7785	0.3543	0.9246
Das Zeichen	17,7,7,6,4,3,2,2	23.1081	0.4605	0.9105
Herr Kellotat oder die Weite des Meeres	20,14,4,4,3,2,1	36.8717	0.6292	0.9557
Rückkehr ins Paradies	11,8,7,6,5,2,1	13.8934	0.3114	0.9096
Glück und Glas	20,7,5,3,2,2,2,2	46.3114	0.9104	0.9752
Laterne, Laterne	21,18,10,8,5,3, 2,2	31.6831	0.3961	0.9692
Abstecher ins Leben	59,33,20,18,18, 12,10,7	77.8567	0.3740	0.9304
Erfahrungen mit Zwergen	12,9,8,8,4,3,3,1	15.0692	0.2860	0.9122
Aller Glanz für Willi	51,25,14,13,5,3, 2,1	91.4667	0.6225	0.9863
Des Hasen Heimgang	20,13,10,9,8,3, 1,1	26.9720	0.3613	0.9373
Veitel und seine Gäste	20,7,7,6,5,4,4	27.4119	0.4853	0.8200
Flieder	19,9,9,7,6,5,1	24.0449	0.3785	0.8884
Kalünz ist keine Insel	125,80,63,43,	177.7745	0.3697	0.9754

	42,13,8,8,4			
Die Leihgabe	10,10,4,3,2,2,2,1,1	15.8317	0.4567	0.9013
Die Flucht nach Ägypten	39,27,20,14,12,11,4,1,1	53.9704	0.3519	0.9737
Der Verrat	18,13,11,4,2,1	28.7026	0.4709	0.9221
Wovon man lebt	18,12,8,8,6,4	22.9433	0.3369	0.9655
Der Morgen der Welt	33,18,15,13,8,7,7,3	42.2099	0.3505	0.9512
Onkel Aluco, einige Vögel, die Zeit	63,19,18,17,14,9,8,2,2	98.5704	0.5650	0.8677
Die Verbündeten	11,11,3,3,3,1	18.2624	0.4986	0.8235

The letters 11 and 19 with Kleist do not present a clearly decreasing tendency, hence the exponential function does not work. There is a subsidiary condition which could be found by literary scientists, but one simply sees that the sequences are almost invariable..

K.-H. Best (2006: 81) analyzed the allomorphs of the plural of German nouns in the text F. Borsch, "Zensiert oder unkontrollierbar?" *al-Journal 5, 2000, 6-9* and obtained the results for the sequence *-{en}, -{e}, -{0}, -{n}, -{s}, Umlaut + -{e}, Umlaut+-{er}, -{er}* and the results are presented in Table 4.2.

Table 4.2
Frequencies of plural morphs in German (Best 2006: 81)

Rank	1	2	3	4	5	6	7	8	a = 45.6009
Frequency	33	28	24	20	12	6	5	4	b = 0.2785
Computed	35.52	27.13	20.78	15.97	12.33	9.57	7.49	5.91	R^2 = 0.9366

P. Steiner and C. Prün (2007) used the inflectional classes of nouns as analyzed by Engel (1988). Taking the numbers from CELEX (cf. Baayen, Piepenbrock, van Rijn 1993) they obtained 37 inflectional noun classes; their frequencies are as presented in Table 4.3. Though mixing of texts (using a complete corpus) is not a recommended way of collecting data, the results are acceptable, even if the differences between the frequencies in some classes and the computed values are not quite satisfactory. A new counting taking into account separate texts would surely lead to better results.

Table 4.3
German inflectional noun classes (Steiner, Prün 2007)

Type	**Frequencies**	**a**	**b**	**R^2**
German inflectional noun classes	5934, 4062, 2685, 2109, 1671, 1267, 686, 655, 530, 508,428, 423, 418,	7763.6867	0.3124	0.9808

(Steiner, Prün 2007)	413, 340, 320, 316, 269, 269, 241, 165, 155, 126, 102, 97, 87, 81, 62, 41, 28, 27, 21, 20, 18, 4, 4, 3			

It can be shown that even the morphological productivity of Indonesian stems (cf. Wimmer, Altmann 1995; Prün, Steiner 2005) could be modelled by the exponential function. The determination coefficient is very high but some individual frequencies deviate strongly from the observed one. Hence the proposed Pólya distribution – from which the exponential function can be derived – captures well also the specificity of this phenomenon. This is a case where the subsidiary conditions can be captured by a more complex function or distribution and their character can be explicated on the basis of a differential equation or a stochastic process. The results of computation are presented in Table 4.4.

Table 4.4
Morphological productivity of Indonesian stems
(Wimmer, Altmann 1995)

Productivity	**f_x**	**Exp**
0	6970	6889.08
1	1961	2357.07
2	1109	806.90
3	622	276.66
4	391	95.29
5	250	33.25
6	173	12.03
7	99	4.77
8	81	2.29
9	44	1.44
10	32	1.15
11	14	1.05
12	12	1.02
13	6	1.01
14	4	1.00
15	1	1.00
16	3	1.00
18	3	1.00
39	2	1.00
a = 6888.0791, b = 1.0728, R^2 = 0.9882		

J. Sambor (1989) studied the distribution of inflection classes of nouns and verbs in the Polish frequency dictionary. She took two aspects into account: the number of nouns and verbs belonging to a given class and the frequency of the entire class in the dictionary. In this way she obtained four diversification cases:

Number of nouns belonging to a special inflection class

Number of verbs belonging to a special inflection class
Frequencies of individual noun classes
Frequencies of individual verb classes.

The results are given in Table 4.5.

Table 4.5
Diversification of Polish inflection classes (Sambor 1989)

Classes	**Data**	**a**	**b**	**R^2**
Nouns, number	2624,2557,1473,1407,1196,1011, 729,569,403,357,273,159,149,56, 54,40,32,31,27,21,20,14,9,8,8,5,4,1;	3490.5423	0.2287	0.9818
Verbs, number	1728,1115,970,456,331,218,203, 200, 112,108,103,78,78,63,62,61, 56, 54,16, 9,8,8,6,5,2,1,1;	2459.7834	0.3666	0.9824
Nouns, frequency	34288,20034,17370,14755,11677, 8326,7626,6475,6121,6021,2553, 2121,1263,1081,683,341,245,175, 171, 107, 84,73,39,38,25,7,6,4;	39465.3018	0.2463	0.9754
Verbs, frequency	19292,13865,10413,4804,3935, 3224,3120,2485,1859,1184,1107, 1084,1011,886,633,541,529,450, 317,251,244,132,57,15,6,1,1;	27122.3564	0.3457	0.9825

K.-H. Best (2012) studied the „Ablaut“ series of strong verbs in German and presented the ranked frequencies in Table 4.6. As can be seen, the exponential function is a satisfactory expression of this trend. The German strong verbs have different forms in present, perfect and participle which make its learning difficult. The change towards unification (i.e. leaving the class of irregular verbs and becoming a member of the class of regular verbs) is a regular historical process which can be modelled with the aid of the Piotrowski law. But for the time being, the individual forms are represented according to the exponential function.

Tabelle 4.6

Fitting the exponential function to the „Ablaut"-series of strong verbs in German (Best 2012, *Duden-Grammatik* 1998, 127)

Rank	Ablautseries	f_x	Exp	Rank	Ablautseries	f_x	Exp
1	ei - i - i	23	21.80	21	au - i: - au	2	1.64
2	i - a - u	19	18.48	22	au - o: - o:	2	1.54
3	ei - i: - i:	16	15.69	23	a - i - a	2	1.45
4	i: - o - o	11	13.35	24	i - a: - e	1	1.38
5	i: - o: - o:	11	11.38	25	i - u - u	1	1.32
6	e - a - o	9	9.72	26	i - a: - e:	1	1.27
7	e - o - o	7	8.33	27	i: - a: - e:	1	1.23
8	i - a - o	6	7.16	28	a - o - o	1	1.19
9	a: - u: - a:	6	6.18	29	e: - u - o	1	1.16
10	e: - a: - e:	6	5.35	30	e: - a: -o	1	1.13
11	e - a: - o	5	4.66	31	o - a: - o	1	1.11
12	e - a: - e	5	4.08	32	o: - i: - o:	1	1.10
13	e: - o: - o:	5	3.59	33	u: - i: - u:	1	1.08
14	a – u: - a	4	3.17	34	ä - i - a	1	1.07
15	a: - i:- a:	4	2.83	35	ä: - a: - o:	1	1.06
16	a - i:- a	3	2.54	36	ö - o - o	1	1.05
17	e: - a: - o:	3	2.29	37	ö: - o: - o:	1	1.04
18	ä: - o: - o:	3	2.08	38	au - o - o	1	1.03
19	ü: - o: - o:	3	1.91	39	ei - i: - ei	1	1.03
20	i - o - o	2	1.77				
$a = 24.7441, b = 0.1738, R^2 = 0.9801$							

Legend: The colon behind a vowel shows length

The high value of the determination coefficient is a sufficient reason to believe in the law-like character of this phenomenon.

4.1. Morphological complexity

Applying a special kind of quantification, it is possible to measure the morphological complexity of words (cf. Roelcke, Altmann 2014). Of course, one can modify the procedure and we hope that it will be at least tested in other languages. G. Altmann (2014) analyzed 5 Indonesian texts and obtained the ranked frequencies of individual complexities as shown in Table 4.7. The author fitted the zeta distribution but here we apply the exponential function.

Table 4.7
Rank-frequencies of morphological complexities of words
in five Indonesian texts (Altmann 2014)

	T 1		T 2		T 3		T 4		T 5	
1	153	151.70	312	311.00	252	250.91	186	186.37	112	111.58
2	35	43.42	67	75.19	45	54.76	66	63.14	33	35.49
3	24	12.94	35	18.76	33	12.56	15	21.83	15	11.76
4	11	4.36	6	5.25	9	3.49	12	7.98	5	4.36
5	7	1.95	6	2.02	6	1.54	5	3.34	4	2.05
6	5	1.27	5	1.24	3	1.12	3	1.78	2	1.33
7	5	1.07	3	1.06	2	1.02	3	1.26	2	1.10
8	2	1.02	2	1.01	2	1.01	1	1.09	1	1.03
9	1	1.01	2	1.00	1	1.00	1	1.03		
10	1	1.00	1	1.00	1	1.00	1	1.01		
11					1	1.00				
12					1	1.00				
	a = 535.3693 b = 1.2677, R^2 = 0.9849		a = 1295.2083 b = 1.4299 R^2 = 0.9956		a = 161.7867 b = 1.5366 R^2 = 0.9899		a = 552.8514 b = 1.0929 R = 0.9975		a = 354.4838 b = 1.1649 R = 0.9978	

The high determination coefficients in all cases hint at the fact that the method of measuring morphological complexity is preliminarily satisfying.

4.2. Derivation

This phenomenon can be studied only in languages which display some degree of synthetism. A.Nagórko-Kufel (1984) stated that in Polish in some cases, the basic word occurs either more frequently than the derivatives, or it occurs more seldomly than the derivatives. There are also cases where the occurrence is equal or cases in which the basic word does not occur at all, only the derivates. The author uses the size of frequency difference as the independent variable X and shows the number of derivatives (f_x) occurring with such a difference to the basic word. Since one cannot search for the cause with each word, one may set up a table (x,f_x) and conjecture that the greater the difference in frequency, the smaller is the number of such derivates. Hence theoretically one must obtain a decreasing function of numbers. The second triad means that the number of derivates is more frequrnt than the basic word; the third triad means that the derivates are equally frequent as their basic words. A. Nagórko-Kufel used the Polish frequency dictionary by Kurcz et al. (1976). The results of fitting the exponential function to her data are presented in Table 4.8. We restrict the first triad to x = 30.

Table 4.8
Frequency of basic word and derivatives in Polish (Nagórko-Kufel 1984)

Derivates more seldom			Derivates more frequent			Derivates equally frequent		
Dif. x	f_x	Exp	Dif. x	f_x	Exp	Dif. x	f_x	Exp
1	78	69.79	1	38	37.46	1	86	85.91
2	60	58.59	2	11	14.18	2	19	19.71
3	48	49.21	3	11	5.76	3	6	5.12
4	33	41.35	4	1	2.72	4	3	1.91
5	29	34.78	5	1	1.62	5	5	1.21
6	19	29.28	6	3	1.22			
7	28	24.67	7	2	1.08			
8	16	20.82	8	1	1.03			
9	17	17.59	9	1	1.01			
10	20	14.89	11	1	1.00			
11	14	12.62	13	1	1.00			
12	14	10.73	29	1	1.00			
13	9	9.15						
14	11	7.82						
15	7	6.71						
16	12	5.78						
17	9	5.00						
18	14	4.35						
19	9	3.80						
20	7	3.35						
21	5	2.96						
22	5	2.64						
23	4	2.38						
24	4	2.15						
25	6	1.96						
26	3	1.81						
27	2	1.68						
28	5	1.57						
29	2	1.47						
30	1	1.40						
a = 82.1775, b = 0.1778 R^2 = 0.9360			a = 100.8544, b = 1.0176 R^2 = 0.9645			a = 385.3464,b = 1.5126 R^2 = 0.9966		

U. Rothe (1990) studied the denominal verbs in German constructed by means of a certain affix, e.g. *–ier* or *–isier-*. She used the classification proposed by V.D. Kaliuščenko (1988) and stated the number of verbs formed from nouns with the aid of the given affixes. Her results are presented in Table 4.9. There are, of course, many more affixes of this sort. It would be very interesting to study their increase both in German and in other languages.

Table 4.9
German verbs formed from nouns with the aid of the given affixes
(Rothe 1990)

Affix	**Frequency**	**a**	**b**	**R^2**
-ier	83,31,30,29,21,21,19,17,16,14,12,11,9,6,5, 5,4,3,1,1	78.2614	0.2329	0.8329
-isier	18,10,10,7,5,3,3,3,2,2,2,1,1	22.8771	0.3535	0.9694

4.3. Cases

Cases are nothing else but a diversification of the relations in the reality transferred into language. Some languages do not differentiate them at all, and join the words by means of prepositions, but more synthetic languages use affixes attached to nouns. They are not easy to learn but they show a clear diversification process. R. Köhler (2012: 114 ff) has shown that the above formula is adequate for the diversification of Finnish cases. We present them here in percentage form because Köhler obtained them in this form from the work by Väyrynen, Noponen and Seppänen (2008). The individual cases and their percentages are presented in the Table 4.10.

Table 4.11
Diversification of Finnish cases (Köhler 2012)

Case	**Rank**	**Percentage**	**Exp**
Nominative	1	29.5	29.60
Genitive	2	20.3	19.77
Partitive	3	13.7	13.32
Inessive	4	7.1	9.09
Illative	5	6.3	6.31
Elative	6	4.4	4.48
Adessive	7	4.4	3.27
Accusative	8	3.1	2.50
Essive	9	2.6	1.99
Allative	10	2.3	1.65
Translative	11	2.2	1.42
Instructive	12	1.9	1.28
Abessive	13	0.2	1.18
Comitative	14	0.1	1.12
Ablative	15	0.1	1.07
	a = 43.5686, b = 0.4210, R^2 = 0.9891		

Automatically the question arises whether or not in strongly synthetic languages all categories obey the exponential distribution. One can also perform an historical study concerning a language which slowly loses its syntheticity and strives against analyticity.

Of course, even counting in languages having fewer cases than Finnish or Hungarian would be a way to continue the forming of a theory of case.

5. Semantic Diversification

Semantic diversification means simply the increase of meanings or functions of a linguistic entity. One can imagines a great number of possibilities, most of them appear with prepositions, conjunctions and affixes.

K.-H. Best (1991) examined the work *Der Spiegel im Spiegel* by M. Kuby (München: Heyne 1987) and found 71 meanings of the German particle *von*. In the above mentioned work there were 53 ordered as in 5.1. Fuchs (1991) studied the German preposition *auf* in the novel *Tonio Kröger* (*Mario und der Zauberer*) by Th. Mann (Frankfurt 1980) (see (1) in the table) and in *Kindheitsmuster* by Ch. Wolf (Darmstadt und Neuwied 1986) (see (2) in the table). As shown by Wolf, there are still other possibilities of ranking, if one considers different functions of the preposition *auf*. A. Hennern (1991) found 44 different uses of the English preposition *in* in the first 5 chapters of the novel *The Hound of the Baskervilles* by C. Doyle (1981). R. Hammerl and J. Sambor (1991) analyzed the Polish preposition *w* in the novel *Droga wiodla przez Narvik* by K. Pruszyński (1956) and presented various aspects; we consider merely the first, main data. E. Beöthy and G. Altmann (1994a,b) studied the diversification of some Hungarian verbal prefixes in the novel *A rózsa kiállítás* by J. Örkény. E. Nemcová (1991, 2007) studied the meanings of Slovak verbal prefixes. R. Fuchs used two different texts, namely (1) Th. Mann: *Tonio Kröger, Mario und der Zauberer*. Frankfurt 1980 and (2,3,4) C. Wolf, *Kindheitsmuster*. Darmstadt und Neuwied 1886. Hammerl und Sambor used two different classifications: (1) *Slownik poprawnej polszczyzn*y. Warszawa 1973 and (2) *Mały Slownik Języka Polskiego. Warszawa 1968.* The results are presented in Table 5.1.

Table 5.1
Semantic diversification of prefixes, prepositions and postpositions

Type	**Frequencies**	**a**	**b**	**R^2**
Hungarian prefix *meg-* (Beöthy, Altmann 1991)	107,8,8,5,5,4,2,1,1	1395.3573	2.5778	0.9912
Hungarian prefix *föl-* (Beöthy, Altmann 1984a)	11,7,7,6,5,4,3,3,3,3	11.3196	0.2156	0.9423
Hungarian prefix *el-* (Beöthy, Altmann 1984a)	83,9,3,2,2,1,1,1,1,	813.0530	2.2942	0.9994
Hungarian prefix *be-* (Beöthy, Altmann 1984a)	20,11,10,7,5,3,3,3, 2,1,1,1,1	26.9134	0.3964	0.9782
Hungarian prefix *ki-* (Beöthy, Altmann 1984b)	12,10,8,8,5,4,3,3,2, 2,2	14.6620	0.2485	0.9722
Slovak prefix *do-* Nemcová (1991)	22, 20, 6, 3	39.4120	0.5430	0.8217
Slovak prefix *na-* (Nemcová 1991)	30,23,17,16,10,7	38.4631	0.2784	0.9776

Slovak prefix *o-* (Nemcová 1991)	17,15,14,11,4	21.8139	0.2383	0.7790
Slovak prefix *od-* (Nemcová 1991)	25, 8, 5, 4, 4, 4	61.0983	0.9565	0.9378
Slovak prefix *po-* (Nemcová 1991)	59, 54, 29, 24, 18	84.0363	0.3160	0.9257
Slovak prefix *pre-* (Nemcová 1991)	33, 16, 13, 7	55.2442	0.5707	0.9682
Slovak prefix *roz-* (Nemcová 1991)	26,26,25,22	27.0931	0.0534	0.7663
Slovak prefix *s-/z-* (Nemcová 1991)	71,38,18,15,7,5	129.9827	0.6248	0.9934
Slovak prefix *u-* (Nemcová 1991)	56, 39, 33, 2	92.2172	0.4762	0.8223
Slovak prefix *vy-* (Nemcová 1991)	61, 47, 33, 32, 23	75.9748	0.2502	0.9717
Slovak prefix *za-* (Nemcová 1991)	77, 35, 21, 8	158.8685	0.7427	0.9949
German particle *von* (Best 1991) in Kuby 1987	54,38,21,21,19,16, 15,13,12,11,11,10, 9,9,8,8,7,7,6,6,5,5, 5,5,4,4,4,3,3,3,3, 3,3,3,2,2,2,2,2,2,2, 1,1,1,1,1,1,1,1,1,1, 1,1	48.7459	0.1600	0.9065
German particle *von* as a preposition in adverbial role (Best 1991)	10,6,6,5,3,3,2,2,1, 1	12.0045	0.3305	0.9540
German *von* in a prepos-itional object (Best 1991)	38,21,12,1,9,4,3,3, 1	68.9710	0.6253	0.9481
German *von* as a pre-position in a nominal part of sentence (Best 1991)	21,19,16,11,8,8,7, 7,5,5,4,4,3,2,2,1,1, 1,1	28.5203	0.2082	0.9787
German preposition *auf* (Fuchs 1991) (1)	24,12,12,6,6,6,4,3, 3,2,2,2,2,2,2,2,2,2, 1,1,1,1,1,1,1,1,1	30.3920	0.3705	0.9532
German preposition *auf* (Fuchs 1991) (2)	312,152,123,41,38, 34,33,30,30,26,25, 24,23,20,12,11,10, 10,10,9,9,8,7,7,6,6, 6,6,5,5,5,5,4,3,3,3, 2,2,2,2,2,2,1,1,1,1, 1,1,1,1,1,1,1,1	497.9016	0.5134	0.9493
German preposition *auf* in a prepositional attribute (Fuchs 1991) (3)	30,12,10,6,3,1,1,1, 1	56.8739	0.7014	0.9761

German preposition *auf* in fixed constructions (Fuchs 1991) (4)	26,25,20,9,7,6,5,5, 2,2,2,2,1,1,1,1	38.4158	0.3169	0.9506
English *in* (Hennern 1991)	51,49,14,14,14,10, 8,7,6,5,5,5,5,4,4,4, 4,4,4,4,3,2,2,2,1,1, 1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1	74.9520	0.3652	0.9111
Polish *w* (Hammerl, Sambor 1991) (1)	199,100,55,40,21,2 1,20,15,13,8, 7,1	340.1610	0.5706	0.9797
Polish *w* (Hammerl, Sambor (1991) (2)	299,55,40,39,20, 11,9,8,7,7,2,1,1,1	1218.1900	1.4163	0.9678

Though in two cases we obtained a determination coefficient smaller than 0.8 (Slovak *o*- and *roz*-) we may either decide to search for another function or to check the data, namely to take longer texts in which the given prefixes are more frequently represented. The result is, nevertheless, acceptable. There are six analyzed languages but, of course, their number should be greatly increased.

U. Roos (1991) differentiated individual texts and studied the semantic diversification of the Japanese postposition *ni* in works of Japanese scientists. The results are presented in Table 5.2. She considered individual functions; we shall perform simple ranking (cf. also Prün 1995).

Table 5.2
Semantic/functional diversification of the postposition *ni* in Japanese (Roos 1991)

Texts	**Ranked frequencies**	**a**	**b**	R^2
Hayashi	17, 13, 10, 9, 7, 6, 4, 2, 2, 1, 1	21.5746	0.2833	0.9730
Matsumura	21, 13, 10, 6, 2, 2, 2,	33.1878	0.4969	0.9799
Umegaki	7, 4, 4, 2, 2, 1, 1	9.8715	0.5123	0.9469
Shibata	8, 5, 5, 4, 3, 2, 1, 1	9.8102	0.3560	0.9346

Distinguishing the individual functions and taking all the text together, Roos (1991) obtained the data presented in Table 5.3.

Table 5.3
Postposition *ni* in Japanese ranked according to its functions (Roos 1991)

Japanese postposition *ni* (Roos 1991)	40,32,31,27,14,11, 6,6,4,4,2,1	54.8167	0.2658	0.9425

Altmann, Best and Kind (1987) studied the diversification of meaning of some German adjectival affixes like *–al/-ell, -ar/-är*, etc. and of prefixes and published two of them explicitly, namely *–os/-ös* and *–al/-ell*. The results are presented in Table 5.4.

Table 5.4
Semantic diversification of some German adjectival affixes
(Altmann, Best, Kind 1987)

Affix	**Frequencies**	**a**	**b**	**R^2**
-os/-ös	59, 31, 20, 6, 5, 4, 2, 2, 2,	110.7682	0.6453	0.9927
-al/-ell	93, 86, 56, 46, 32, 23, 20, 19, 11, 7, 7, 6, 6, 4, 4, 1	128.1519	0.2701	0.9854

H. Kuße (1991) analyzed the meanings and functions of Russian conjuntions *a* and *no* in the quoted work by Karamzin and found the frequencies as given in Table 5.5. He applied the negative binomial distribution with very good results. In Table 5.5. it will be shown that the exponential functions also yield very good results.

Table 5.5
Polysemy of conjunction *a* and *no* in Russian (Kuße 1991)

	Frequencies	a	b	R^2
a	22,11,8,6,6,5,4,3,3,2,2,2,2,1,1,1,1,1	28.3747	0.4915	0.9483
no	19,16,13,10,9,8,8,7,6,5,4,3,3,1,1,1	21.4502	0.1860	0.9733

E. Nemcová (2009) considered the polysemy and function of nominal affixes in German as given by H. Becker (1995) in four German newspapers: *Frankurter Allgemeine Zeitung* (Germany, FAZ), *Neue Zürcher Zeitung* (Switzerland, NZZ), *Die Presse* (Austria, PRESSE) and *Neues Deutschland* (former GDR, ND),. They are *–ung, -e, -ion, -schaft, -t, -ie, -er, -ent, -nis, -enz, -zeug, -it, -keit, -heit, -el, -ik, pref.+0-morpheme, -al, -ität, -tum, -anz, -ur, -ium, -ismus, -ement, -um, -eur, -or, -ant, -är, -ei, -ial, -age, -ier, -ose, -at, -igkeit, -elle, -ar, -ist, -arier, -ling, -ül, -itis, -ler, -in, -chen, -lein.* Not all can be found in all newspapers but Nemcová could set up a ranking. We omit those that have zero frequency. The results can be seen in Table 5.6.

Table 5.6
Nominal affixes in German press (Nemcová 2009)

Newspaper	Frequencies	a	b	R^2
FAZ	260,78,42,34,34,21,18,17,17,11, 11,10,9,8,8,8,6,6,5,5,4,4,3,3,3,2,2, 2,2, 2,1,1,1,1,1,1,1,1	58.9180	0.9552	0.9550
NZZ	299,106,71,29,27,26,18,15,15,14, 14,13,12,11,11,,8,8,6,5,5,5,4,4,4, 4,3, 2,2,2,2,2,1,1,1,1,1,1,1,1	653.7091	0.8097	0.9686
PRESSE	241,49,44,30,28,20,18,18,16,12,8, 8,8,7,7,7,6,5,5,4,4,3,3,3,3,3,2,2,2, 2, 1,1,1,1,	783.7115	1.2006	0.9334

ND	324,58,50,41,40,33,32,28,26,23, 21,21,17,12,12,9,8,7,7,5,5,5,5,5,4, 4,3,3,2,2,2,1,1,1,1,1,1,1,1,	1248.8631	1.3638	0.9o84

A special attention has been devoted to cases in some languages. They are present only in synthetic languages. Popescu et al. (2009) analyzed the frequency of cases in German (4), Slovenian (6), Slovak (7) and Russian (6) and obtained the results in Table 5.7.

Table 5.7
Rank-frequencies of cases in four European languages (Popescu et al. 2009)

German		**a**	**b**	**R^2**
Text	**Data**			
01	45,40,11,6;	82.2776	0.5450	0.8330
02	50,47,33,7;	77.9565	0.3677	*0.7587*
03	48,44,27,10;	74.1799	0.3754	0.8460
04	40,32,24,3;	64.4072	0.4377	0.8148
05	30,28,27,4;	43.7324	0.3068	*0.5741*
06	73,49,42,13;	110.9958	0.4160	0.9091
07	58,32,26,4;	102.2089	0.5796	0.9320
08	32,21,20,4;	48.8350	0.4351	0.8376
09	51,39,34,7;	77.2494	0.3872	0.8090
10	49,35,27,4;	79.9370	0.4695	0.8702
11	78,53,28,2;	146.1120	0.5998	0.9274
12	48,37,20,1;	86.4089	0.5474	0.8752
13	48,45,35,3;	76.0505	0.3761	*0.6828*
14	43,29,24,2;	71.2971	0.4950	0.8509
15	32,26,23,3;	48.8251	0.3870	*0.7339*
16	46,34,28,3;	71.9454	0.4387	0.8065
17	56,46,38,2;	89.8929	0.4169	*0.7377*
18	46,45,28,6;	73.5385	0.3842	*0.7583*
19	43,39,34,4;	65.5963	0.3501	*0.6584*
20	64,61,38,6;	104.5182	0.4035	*0.7639*
Slovenian		**a**	**b**	**R^2**
01	64,54,52,21,20,7;	93.2201	0.3114	0.8656
02	89,65,49,39,23,13;	124.2891	0.3314	0.9824
03	86,78,45,29,27,7;	130.6258	0.3552	0.9320
04	81,64,42,26,17,15	115.4391	0.3716	0.9934
05	26,16,14,11,9,1;	35.0608	0.3593	0.9130
06	82,77,52,31,27,16	118.6707	0.3022	0.9496
07	78,34,28,23,20,12;	110.9525	0.4483	0.9028
08	43,34,28,15,14,5;	61.0379	0.3298	0.9499
09	67,59,42,23,16,14;	98.1359	0.3288	0.9507
10	133,97,91,48,28,16;	192.1239	0.3399	0.9394
Slovak				
Text	Data	**a**	**b**	**R^2**
01	26,18,16,13,8,2;	34.9087	0.3245	0.9222

02	14,13,5,4,2,2,2;	23.1300	0.4861	0.9149
03	53,52,47,46,42,6;	68.4038	0.1623	*0.5564*
04	36,31,30,25,10,5,4;	51.1355	0.2762	0.8372
05	67,50,44,28,15,15,2;	95.8032	0.3339	0.9508
06	39,36,22,12,10,4,2;	61.5526	0.3362	0.9367
07	27,22,10,7,4,2,	44.9479	0.4875	0.9538
08	28,20,10,9,7,3;	44.4313	0.4561	0.9779
09	163,105,72,43,24,22,2;	256.8889	0.4530	0.9921
Russian				
Text	Data	**a**	**b**	**R^2**
01	134,74,43,24,19,15;	229.0380	0.5548	0.9942
02	64,25,19,8,7,6;	126.7926	0.7260	0.9729
03	36,31,28,27,6,3;	50.9864	0.2816	*0.7493*
04	54,28,21,14,4,2;	91.2274	0.5557	0.9786
05	35,15,15,12,7,4;	49.9067	0.4569	0.9100
06	66,35,21,11,11,6;	114.3263	0.5788	0.9914
07	36,28,28,22,17,15;	41.4355	0.1768	0.9579
08	47,33,23,18,16,10;	62.1620	0.3206	0.9882
09	83,31,21,15,12,12;	157.5656	0.6948	0.9400
10	51,17,10,9,7,2;	117.9497	0.8820	0.9615

In some cases the fit seems to be not satisfactory but this is caused by the great jump in the last frequency. Here a subsidiary condition should be taken into account and one should, perhaps, use mixed functions.

For the German cases, K.-H. Best (2006: 64) analyzed the novel *Hühner, Adler und Mäuse* by Pestalozzi and obtained the order Nominative, Dative, Accusative, Genitive. The results are presented in Table 5.8.

Table 5.8
Ranking the cases in a German text (Best 2006)

Rank	1	2	3	4	a = 24.3391 b = 0.3910 R^2 = 0.9703
Frquency	17	13	9	5	
Computed	17.46	12.13	8.53	6.09	

Concerning one sole entity having many meanings and functions, U. Rothe (1986) studied the French conjunction *et* and obtained the results presented in 5.9.

Table 5.9
Diversification of the French conjunction *et*

		a	**b**	**R^2**
French *et* (Rothe 1986)	17,17,13,11,9,6,5,5,5,5,4,4,4,4,4,4,3,3, 3,3,3,3,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1	18.4498	0.1515	0.9451

M. Hug (2001: 219) studied the diversification of the French "que" and found 22 different functions and environments. K.-H. Best (2006) fitted to the data the 1-displaced Hyperpoisson distribution. Another diversification of "que" concerning its use has been presented by M. Hug (2000). Here K.-H. Best (2006) applied the modified negative binomial distribution. In both cases, we fit the exponential function.

Table 5.10
Functions, environments and uses of the French "que" (Hug 2000, 2001)

		a	**b**	**R^2**
French *que* (Hug 2001; Best 2006)	501,408,290,223,165,128,98,69,50, 38,22,10,9,6,6,3,2,1,1,1,1,1	685.1859	0.2858	0.9973
Use of *que* (Hug 2000; Best 2006)	870,325,318,212,173,54,49,25,6,1	1374.8830	0.5293	0.9427

V. Levickij, V.V. Drebet and S.V. Kiiko (1999) analyzed the number of certain POS according to the numbers of meanings in the German Duden dictionary. Peculiarly enough, the number of POS is almost monotonically decreasing; hence we may try to fit the exponential function. The results are presented in Tables 5.11 Here we shall show the computed numbers.

Table 5.11
Number of meanings of words according to POS in the German Duden dictionary (Levickij, Drebet, Kijko 1999)

Meanings	**Verbs**	**Exp**	**Nouns**	**Exp**	**Adjectives**	**Exp**
1	1385	1395.54	1571	1595.55	213	215.29
2	721	690.37	913	821.34	128	120.35
3	341	341.78	342	423.03	62	67.47
4	145	169.46	214	218.12	36	38.02
5	62	84.28	108	112.70	23	21.62
6	37	42.17	51	58.47	12	12.48
7	24	21.35	28	30.56	6	7.40
8	9	11.06	20	16.21	6	4.56
9	5	5.97	9	8.82	2	2.98
10	6	3.46	8	5.03	1	2.10
11	5	2.22	6	3.07	1	1.62
12	5	1.60	2	2.07		
13	2	1.30	2	1.55		
14	2	1.15	1	1.28		
15	5	1.07	2	1.15		
16	2	1.04	1	1.07		

17	1	1.02				
18	1	1.01				
20	2	1.00				
22	2	1.00				
23	1	1.00				
24	1	1.00				
26	1	1.00				
	a = 2821.0487 b = 0.7045 R^2 = 0.9990		a = 3099.4518 b = 0.6646 R^2 = 0.9944		a = 384.7588 b = 0.5853 R^2 = 0.9977	

As can be seen, even in this case, the exponential function is a sufficient model. Needless to say, in many other investigations one can use either the Zipfian power function or Zipf-Mandelbrot’s function (i.e. without normalization)

L. Wang (2016) studied also the diversification of individual POS in Chinese and ranked the numbers of diversifications into other classes. For the 12 POS types, she obtained the ranking presented in Table 5.12, where already the above exponential function proposed in Popescu, Altmann, Köhler (2009) has been used. Of course, for every type of POS, the ranking of the resulting POS is different. We simply quote the given results.

Table 5.12
Ranking of Chinese POS diversifying into other POS classes (Wang 2016)

POS	**Frequencies**	**a**	**b**	**R^2**
Noun	1946,712,276,271,54,51,42,29,29,14,3	4852.5348	0.9228	0.9911
Verb	1846,606,188,140,63,38,27,24,10,7,7	6008.6364	1.1306	0.9970
Adjective	712,606,150,27,12,9,7,5,3,2,2	1461.0868	0.6336	0.9154
Auxiliary	29,24,9,8,6,6,5,4,3,3,1	43.4432	0.4147	0.9409
Adverb	271,188,150,37,19,17,15,9,7,1	458.2651	0.4839	0.9566
Quantifier	276,140,27,15,10,6,5,2,1,1	669.0348	0.8718	0.9853
Preposition	63,54,19,18,10,9,6,1,1	104.1285	0.4600	0.9450
Pronoun	42,17,11,8,7,5,2,2,1	77.7696	0.6781	0.9688
Conjunction	51,38,37,18,11,7,5,5	74.4796	0.3481	0.9419
Onomatopoeia	27,14,12,7,3,1,1,1	44.5978	0.5504	0.9741
Number	29,7,7,2,2,1,1	92.5652	1.2081	0.9697
Interjection	10,7,4,3,2	15.2906	0.5111	0.9906

K.-H. Best (1990) studied the classes of nouns formed by the prefix *ge-* in Early New High German relying on the analysis of Habermann, Müller 1987). K.-H. Best distinguished 9 classes (cf. also Altmann, Best, Kind 1987) and obtained the results presented in Table 5.13. K.-H. Best remarks that one could set up the classes also in a different way.

Table 5.13
Frequencies of nouns formed by *ge-* in Early New High German
(Best 1990)

Frequencies								a	b	R^2
12	11	11	6	6	4	1	1	16.5447	0.2805	0.8531
13.50	10.44	8.13	6.39	5.07	4.07	3.32	2.75			

5.1. Definition chains

Definition chains are known under the name Martin's law. Martin (1974) took individual words from the dictionary and analyzed their hypernyms. Each hypernym lies at a higher abstraction level. Increasing the level, the number of words automatically decreases because each more abstract word contains a number of words with more concrete meaning (hyponyms). For example a *revolver* is a *hand weapon*, which is a *weapon*, which is an *instrument*, which is an *object* etc. A good explanation can be found in Kisro-Völker (1984). Up to the most abstract concept each word has a definition chain. The length of chains can be computed, and the number of words at the individual levels can be computed; if one takes several words then a tree is formed; a word may have several definition chains, it depends on its polysemy, etc. The study of definition chains developed into a discipline which was practiced especially by Polish linguists (cf. e.g. Sambor, Hammerl 1991). In Table 5.14 we present several results and fit to the numbers the exponential function. R. Hammerl developed a special formula directly for definition chains and applied it in many cases successfully. There are also different ways of counting, e.g. with and without the repetition of a hypernym, because some of them may occur with different words. In the last line, one finds the number of words at the x abstractness level in the German dictionary as computed by S.Schierholz (1989). For some theoretical views, see Altmann, Kind (1983). In the following table, the first number means always the number of words that were analyzed.

Table 5.14
Definition chains

Source	Frequencies	a	b	R^2
French, Martin (1974) (with repetition)	1723, 348, 108, 39, 13, 3	8149.9883	1.5556	0.9992
French Martin (1974) (without repetition)	1375, 240 69, 26, 10, 3	7567.3637	1.7068	0.9993
Polish Hammerl (1987) (with repetition)	1000, 618, 271, 110, 44, 16, 9, 3, 1	1961.1755	0.6465	0.9891
Polish	382, 347, 161, 66, 28, 7, 6,	694.8913	0.4981	0.9311

Hammerl (1987) (without repetition)	2, 1			
German Schierholz (1989)	1483, 1039, 443, 192, 99, 49, 30, 20, 15, 12, 9, 9, 8, 6, 6, 2	2773.0314	0.5838	0.9832

R. Hammerl (1989) studied the number of paths in the definition chains from the most concrete to the most abstract in 100 German chains using the Handwörterbuch (1984) and obtained the result presented in Table 5.15.

Table 5.15
Number of ways in 100 chains in German (Hammerl 1989)

Number of ways	**1**	**2**	**3**	**4**	**5**	**6**	**a**	**b**	**R^2**
Frequencies	53	27	11	4	4	1	115.2526	0.7852	0.9946
Exp	53.56	24.97	11.93	5.98	3.27	2.04			

The second Martin's law has been treated especially by J. Sambor (2005). In that version, she counted the number of notions at the given level of generality. Analyzimng Polish and German she obtained the numbers presented in Table 5.16. Here, the sequence is not monotonically decreasing and we must operate with the formula (4). Hammerl and Sambor (1993) applied the negative binomial distribution but it is easier to interpret the respective parameter as the result of a writer's/speaker's effort.

Table 5.16
Second Martin's law concerning the generality levels (Sambor 2005)

Level of generality	Polish f_x	Polish Menz.	Level of generality	German f_x	German Menz.
1	259	245.52	1	273	267.35
2	438	462.94	2	406	420.21
3	440	419.18	3	368	349.52
4	291	281.12	4	218	224.12
5	142	160.09	5	120	124.79
6	84	82.40	6	67	63.73
7	31	39.75	7	24	30.86
8	6	18.50	8	14	14.58
9	1	8.59	9	4	6.96
			10	1	3.54
a = 761.7762, b = 1.1363, c = 2.5572, R^2 = 0.9922			a = 787.6128, b = 1.0841, c = 2.2185, R^2 = 0.9968		

It would be, of course, scientifically very interesting if we had at our disposal several other languages. The testing of this hypothesis is not sufficient so far.

A number of further problems associated with definition chains are shown in Schierholz (1991) and Bagheri (2002) where the approaches are quantified but not yet simplified.

5.2. Semantic associations

In several languages, there are dictionaries consisting of individual words to which test persons uttered their associations. Thus saying "music", one can have various associations like "melody, tones, Chopin, Bach, violin, orchestra, jazz, etc." The examiner collects the responses and states how many words are there having x = 1,2,3,… associations. Here we shall show only one example, namely the word "high", taking the numbers from Palermo, Jenkins (1964), Altmann (1992), Hřebíček (1995). There were a number of distributions applied to this phenomenon; we show here merely the exponential function. The results are presented in Table 5.17

Table 5.17
Associations of the word "high" (Palermo, Jenkins 1964)

x	f_x	Exp	x	f_x	Exp
1	129	128.70	18	2	1.00
2	16	19.77	19	2	1.00
3	14	3.76	20	2	1.00
4	12	1.41	21	2	1.00
5	6	1.06	22	2	1.00
6	5	1.01	23	2	1.00
7	4	1.00	24	1	1.00
8	4	1.00	25	1	1.00
9	3	1.00	26	1	1.00
10	3	1.00	27	1	1.00
11	3	1.00	28	1	1.00
12	2	1.00	29	1	1.00
13	2	1.00	30	1	1.00
14	2	1.00	31	1	1.00
15	2	1.00	32	1	1.00
16	2	1.00	33	1	1.00
17	2	1.00	34	1	1.00
			35	1	1.00
a = 868.8017, b = 1.9174, R^2 = 0.9802					

5.3. Pure polysemy

G. Wimmer and G. Altmann (1999a) analyzed the Maori dictionary data and stated the polysemy as given in Table 5.18. The problem should be analyzed in other languages, too, but the work is not simple. The result depends on the size of the dictionary. In any case, the exponential function is a satisfactory model.

Table 5.18
Polysemy of words in the Maori dictionary (Wimmer, Altmann 1999)

Degree of polysemy	Number of words	Exp
1	5111	5097.48
2	1558	1632.93
3	604	523.56
4	215	168.33
5	112	54.58
6	41	18.16
7	19	6.49
8	7	2.76
9	14	1.56
10	3	1.18
11	1	1.06
12	3	1.02
23	1	1.00
a = 15916.1573, b = 1.1388, R^2 = 0.9992		

H. Sanada (2007) studied the polysemy of Japanese scholarly terms using three Japanese dictionaries and Japanese magazines. The definition of a scholarly term depends always on the researcher; here we accept Sanada's choice. Her results are presented in the Table 5.19. In the magazines, Sanada employed intervals of occurrences; here we shall use the ranks.

Table 5.19
Polysemy of Japanese scholarly terms (Sanada 2007)

Japanese dictionaries			90 Japanese magazines		
Polysemy	Frequency	Exp	Polysemy Rank	Frequency	Exp
1	743	742.35	1	191	188.19
2	232	236.35	2	88	95.33
3	84	75.72	3	47	48.54
4	21	24.72	4	31	24.96
5	9	8.53	5	19	13.07
6	6	3.39	6	9	7.08
7	3	1.76	7	8	4.07
10	1	1.02	8	1	2.55
			9	1	1.78
			10	2	1.39
			11	3	1.20
			12	1	1.10
			13	5	1.05

			14	1	1.03
			15	1	1.01
			16	1	1.01
a = 2335.2185,b = 1.1474, R^2 = 0.9998			a =371.4438, b = 0.6853, R^2 = 0.9953		

Krylov (1982) analyzed the polysemy of Russian words in Ožegov's dictionary and obtained the results which were further processed by Kempgen (1995). Krylov computed simply the number of meanings of individual words and obtained the results presented in Table 5.20

Table 5.20
Polysemy of Russian words (Krylov 1982)

Number of meanings	**Number of words**	**Exp**
1	23456	23448.232
2	6288	6326.202
3	1682	1707.307
4	649	461.299
5	232	125.172
6	110	34.497
7	55	10.036
8	34	3.438
9	23	1.658
10	16	1.177
11	7	1.048
12	2	1.013
13	2	1.003
14	0	1.001
15	2	1.000
16	1	1.000
a = 86917.8089, b = 1.3102, R^2 = 0,9999		

A similar investigation has been performed by G. Altmann (2002) concerning Indonesian stems using the Waring distribution. In Table 5.21 it will be shown that the exponential function is sufficient, too. Altmann used the Echols-Shadily dictionary (1963)

Table 5.21
Polysemy of Indonesian stems (Altmann 2002)

Number of meanings	**Number of words**	**Exp**
1	16439	16428.57
2	5209	5268.36
3	1763	1689.93
4	555	542.54
5	236	174.64
6	59	56.68
7	37	18.85
8	13	6.72
9	12	2.84
10	5	1.59
11	1	1.19
12	5	1.06
13	1	1.02
16	1	1.00
a = 51233,4634, b = 1.1374, R^2 = 0,9999		

S. Rensinghoff considered 1000 French lexemes from MICRO ROBERT and stated their polysemy as presented in Table 5.22. Here too, we consider simply the number of meanings as above.

Table 5.22
Polysemy in 1000 French stems (Rensinghoff 1988)

Number of meanings	**Number of words**	**Exp**
1	549	552.46
2	256	240.12
3	89	104.69
4	37	45.96
5	24	20.50
6	14	9.45
7	13	4.67
8	3	2.59
9	3	1.69
10	2	1.30
11	3	1.13
12	1	1.06
13	2	1.02
14	1	1.01

15	1	1.00
23	1	1.00
a = 1271.7582, b = 0.8356, R^2 = 0,9978		

6. Ranking

Ranking is a conceptual process and epistemologically, it is not worse than any other classificatory process. Some scientists protested against ranking in linguistics because the rank-variable has been set up a posteriori. They simply forgot that also the dependent variable, e.g. the individual classes, their numbers, etc. are results of our classificatory effort, and not given by reality. The ascribing of ranks is not worse than any kind of classification – our daily bread.

Here we shall consider only some types of ranking in order to illustrate this procedure.

H. Sanada and G. Altmann (2009) analyzed the occurrence of Japanese postpositions in their grammatical functions in three texts and obtained the results presented in Table 6.1. They applied function (2) and presented the results which are shown in the same table. The text analyzed were *Jinseiron Note* (essay, Miki 1941, 1995), *Kusa no hana* (novel, Fukunaga 1954, 1995) and *Shinbashi Karasumori Guchi Seishunhen* (novel, Shiina 1987, 1995)

Table 6.1
Fitting function (2) to postposition data (Sanada, Altmann 2009)

Rank	**Jinseiron Note**		**Kusa no hana**		**Shinbashi Karasumori Guchi**	
r	f_r	Exp	f_r	Exp	f_r	Exp
1	2614	2702.95	4279	4531.84	4399	3844.26
2	2076	2172.93	3534	3834.68	2747	3182.28
3	1832	1746.89	3270	3244.78	2016	2634.31
4	1493	1404.41	3124	2745.66	1945	2180.74
5	1239	1129.12	2887	2323.34	1938	1805.29
6	1127	907.83	2366	1966.00	1901	1494.51
7	929	729.94	1975	1663.64	1892	1237.26
8	477	586.95	1320	1407.81	988	1024.32
9	347	472.01	830	1191.34	877	848.06
10	266	379.62	780	1008.18	662	702.15
11	190	305.35	632	853.20	458	581.38
12	174	245.65	616	722.08	302	481.41
13	69	197.66	452	611.12	298	398.67
14	44	159.08	344	517.24	282	330.17
15	42	128.07	271	437.81	237	273.47
16	41	103.14	238	370.60	209	226.54
17	36	83.11	186	313.73	185	187.69
18	30	67.00	179	265.61	159	155.53

19	26	54.05	175	224.89	107	128.92
20	22	43.65	171	190.44	101	106.88
21	22	35.28	132	161.29	96	88.65
22	15	28.56	117	136.63	93	73.55
23	14	23.15	114	115.76	90	61.05
24	11	18.81	97	98.10	89	50.71
25	8	15.31	79	83.16	61	42.15
26	8	12.51	69	70.52	54	35.06
27	7	10.25	69	59.82	54	29.19
28	7	8.43	67	50.77	53	24.34
29	5	6.98	65	43.11	44	20.32
30	4	5.80	64	36.63	41	16.99
31	4	4.86	61	31.15	37	14.24
32	3	4.10	57	26.51	36	11.96
33	2	3.50	55	22.58	30	10.07
34	1	3.01	53	19.26	25	8.51
35	1	2.61	41	16.45	24	7.21
36			39	14.08	22	6.14
37			31	12.06	16	5.26
38			31	10.36	14	4.52
39			27	8.92	14	3.92
40			24	7.70	13	3.41
41			23	6.67	12	3.00
42			23	5.80	11	2.65
43			22	5.06	10	2.37
44			21	4.44	9	2.13
45			20	3.91	6	1.94
46			16	3.46	6	1.78
47			15	3.08	6	1.64
48			13	2.76	5	1.53
49			12	2.49	4	1.44
50			10	2.26	3	1.36
51			7	2.07	3	1.30
52			7	1.90	3	1.25
53			5	1.76	2	1.21
54			5	1.65	2	1.17
55			2	1.55	2	1.14
56			2	1.46	1	1.12
57			1	1.39	1	1.10

58					1	1.08
	a = 3361.3072 b = 0.2184 R^2 = 0.9851		a = 5354.7916 b = 0.1671 R^2 = 0.9785		a = 4642.9967 b = 0.1890 R^2= 0.9592	

The exponential function seems to be adequate in all cases. It would be interesting to analyze each postposition separately and perform the same procedure also in other languages having this grammatical means.

6.1 Grammatical categories

In a very thorough investigation, U. Rothe (1991a) analyzed the functions of the German genitive in the texts by S. Lenz, *Die Erzählungen 1965-1984*, Deutscher Taschenbuchverlag, München 1986, using official German grammars. She obtained the results presented in Table 6.2.

Table 6.2
Functions of the German genitive (Rothe 1991a)

Type	**Frequencies**	**a**	**b**	**R^2**
German genitive (Rothe 1991a)	47,31,28,27,24,23,21,21,13,9,9, 8,8,7,6,6,6,6,5,4,4,4,3,3,3,3,3,3, 3,2,2,2,2,2,2,2,2,2,2,2,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1	46.4457	0.1399	0.9722

Rothe (1990) used the 16 semantic classes of nouns defined by Köpcke (1982). Based on the results of Hirsch-Wierzbicka (1971) she stated the number of monosyllabic masculine, feminine and neuter nouns and their ascription to the given classes. They are presented in Table 6.3.

Table 5.3
Numbers of semantic classes of monosyllabic nouns belonging
to different genders (Rothe 1990)

Type	**Frequencies**	**a**	**b**	**R^2**
Masculine nouns	65,23,15,11,10,7,8,5,5,4,4	127.8466	0.7348	0.9335
Feminine nouns	18,12,5,5,4,3,1,1,1	29.1105	0.5310	0.9735
Neuter nouns	37,18,12,6,5,4,4,3,2,2,1,1,1	64.3508	0.6060	0.9867

Levickij and Lučak (2005) studied the frequency of 16 tenses ascribed to English verbs subdivided into 20 semantic categories according to Levin (1998). The results are presented in Table 6.4. Some classes are not presented here because the frequencies were either too small or too few or none. In the last case, we added a zero, for computational reasons but this was necessary only in one case. As can be seen, the parameter *b* attains very variegated values depending most probably on the association of some verbs with a special meaning to a special tense.

Table 6.4
Tenses ascribed in texts to semantic classes of verbs
(Lučak 2003, Levickij, Lučak 2005)

Semantic class	**Frequencies**	**a**	**b**	**R^2**
Existence	1092,572,103,100,49,31,30,29,6,2	2530.7074	0.8256	0.9814
Perception	1321,327,122,55,54,41,40,18,7,3	4926.3614	1.3200	0.9948
Circumstance	736,412,200,77,58,47,23,21,5	1411.8491	0.6442	0.9959
Directed motion	657,118,103,75,59,35,26,22,6,5,3	2566.2820	1.3744	0.9572
Mental process	493,405,38,19,13,10,7,5,1	1070.5796	0.7006	0.8910
Ownership	177,135,24,21,19,18,7,1	345.0816	0.6226	0.9238
Manner of motion	283,27,22,17,17,9,7,2,2,1,1	2605.6969	2.2244	0.9863
Change of ownership	126,30,26,26,20,10,3	307.1942	0.9367	0.8999
Change of position	157,23,19,6,6,6,3	891.3976	1.7459	0.9855
Psychological state	108,71,9,2,2,2,1	250.4135	0.8043	0.9363
Position	73,68,6,1,1,1	159.3712	0.6847	0.8159
Change of physical state	71,21,11,7,7,6,5	202.2350	1.0725	0.9735
Impact/Effect	68,23,16,9,2,1,1,1	162.3616	0.9010	0.9850
Removal	46,19,7,3,2	117.8826	0.9597	0.9991
Ingestion	41,14,6,6,3,1,1	109.8361	1.0185	0.9889
Exchange	15,8,4, (0)	33.2109	0.8446	0.9551

It must be remarked that the semantic verb classes (just as any other linguistic ones) are not objective entities. They are defined differently by each researcher and the assignment of a verb to a special class is a problem of the given school, of the given text or text type, or of the given language.

Best and Altmann (2015) studied the frequencies of simple attributes in German using the works of writers G. Grass, C. Hein and E.-W. Händler. They found 9 classes of attributes and differentiated the position of attributes (in front of, f, and behind, b). There are types like Adjective attribute (f), Genitive attribute (b and f), Apposition (b) etc. They applied the Zipfian function with very good results. Since they ranked the attributes, it is also possible to use the above function. The results are presented in Table 6.5.

Table 6.5
Ranking of simple attributes in German (Best, Altmann 2015)

	G. Grass		C. Hein		E.-W. Händler	
Rank	f_x	Exp	f_x	Exp	f_x	Exp
1	474	438.40	467	445.73	263	241.73
2	213	280.00	180	237.51	91	142.24
3	169	178.96	134	126.78	90	83.87
4	114	114.51	90	67.89	64	49.62
5	107	73.40	80	36.57	51	29.53
6	95	47.18	27	19.92	26	17.74
7	45	30.46	16	11.06	25	10.82
8	20	19.79	11	6.35	10	6.76
9	3	12.99	5	3.85	5	4.38
	a = 685.7451, b = 0.4496, R^2 = 0.9415		a = 836.2772, b = 0.6315, R^2 = 0.9632		a = 410.3004, b = 0.5332, R^2 = 0.9192	

L. Uhlířová (2002) studied the frequency of possessive constructions in a Czech corpus and found 37 individual nouns which were heads of these constructions. The frequencies are displayed in Table 6.6. She used for both cases the negative hypergeometric distribution very successfully; however, one can also use the simple exponential function.

Table 6.6
Frequency of individual nouns in possessive constructions in a Czech corpus (Uhlířová 2002)

Subclasses of possessive constructions in Czech (Uhlířová 2002)	**a**	**b**	**R^2**
171,131,61,54,52,51,51,48,45,33,33,30,27,27,24, 19,16, 15,14,13,11,11,10,10,9,8,8,7,4,4,4,4,4,3, 3,2,1	161.6612	0.1681	0.8878
Nouns in possessive constructions in Czech (Uhlířová 2002)			
295,280,238,225,216,213,158,142,139,128,96,90, 88,81,77,77,76,61,51,43,38,37,27,25,24,23,22,19, 14,12,11,9, 7,6,4,4,2	337.1924	0.1022	0.9902

6.2. Adnominals

Adnominals represent a large chapter of text linguistics. One defines the types and studies the occurrence of individual types in individual texts. One obtains numbers which can be ranked; however, the individual types do not obtain necessarily the same rank in each text. But one can apply to the series a distribution or a function, usually one used the Zipfian family (power/zeta, Zipf-Mandelbrot, Zipf-Alekseev, Menzerathian, etc.). Here we present the ranking of Russian adnominals as they were observed in 42 Russian texts (cf. Andreev, Popescu, Altmann 2017). The titles of the texts and the authors can be found in the above reference. We fit the above defined exponential function and present the data in Table 6.7. There are no exceptions; the exponential function is a sufficient model.

Table 6.7
Fitting the exponential function to Russian adnominals in 42 texts
(Andreev, Popescu, Altmann 2017)

Text	**Frequencies**	**a**	**b**	R^2
T 1	229,83,59,59,36,32,29,24,21,8,6,6,5,4,4,3,3,1,1,1	358.9040	0.5434	0.9244
T 2	198,107,50,43,34,30,26,24,21,18,13,10,8,8,6,4,3, 3,2,2,2	301.5836	0.4812	0.9488
T 3	213,107,45,45,28,16,16,15,15,13,9,7,7,6,3,3,2,2, 2, 1,1	380.2554	0.6092	0.9744
T 4	219,108,56,46,22,19,17,17,16,13,13,12,9,8,7,5,3, 2, 2,2,2,1,	382.2223	0.5902	0.9741
T 5	197,90,44,43,41,40,39,31,25,15,9,9,8 7,6,5,2,2,2, 1	269.9918	0.4366	0.8923
T 6	103,54,46,35,29,22,17,17,14,8,6,5,4,2,2,2,2,1,1	125.6455	0.3148	0.9662
T 7	198,113,69,68,40,30,30,20,13,12,8,8,7,5,4,4,3,3, 2	273.0013	0.3883	0.9776
T 8	350,137,71,55,36,35,29,24,22,18,11,10,9,9,8,6,6, 4, 3,3,2	709.6540	0.7391	0.9652
T 9	267,121,43,41,29,29,26,20,19,17,12,10,5,5,2,2,2, 2,1	537.405	0.7222	0.9617
T 10	349,121,94,52,35,32,29,24,23,23,22,15,13,8,6,6, 2, 1,1,1,1	682.6698	0.7160	0.9509
T 11	194,58,54,42,27,21,20,19,18,17,9,8,5,4,3,2,2,2,1, 1,1	340.3420	0.6434	0.9102
T 12	220,92,77,50,44,38,34,24,22,14,9,5,4,4,4,3,3,2,1, 1,1	303.5008	0.4328	0.9401
T 13	209,89,68,65,46,35,29,22,14,14,13,10,9,7,4,4,3, 2,1	276.2242	0.4017	0.9364
T 14	192,87,61,43,33,28,23,21,20,19,13,8,7,6,3,2,2,2, 2	282.5778	0.4721	0.9460
T 15	252,138,53,49,37,32,28,26,23,22,22,10,8,8,8,7,4, 2,2,2,1	422.2681	0.5510	0.9507
T 16	358,106,72,49,45,45,43,42,37,34,16,10,7,5,4,4,2, 2,1,1,1	809.4124	0.8542	0.9090
T 17	405,140,72,69,45,45,40,30,28,28,11,10,6,6,5,5,4, 4,2,2	868.8798	0.7986	0.9456
T 18	211,60,47,41,35,31,30,26,24,21,15,13,10,6,5,4,3, 2,1,1,1	396.2996	0.7056	0.8639
T 19	168,50,41,35,32,27,20,18,17,14,10,7,6,5,4,4,3,3, 2,1	283.0097	0.6179	0.8792
T 20	272,51,49,29,27,23,20,19,18,8,8,8,7,6,5,3,3,3,2,	976.1063	1.2949	0.9394

	2,2,2			
T 21	259,114,32,32,30,26,23,19,16,16,10,9,5,4,3,3,2,2,1,1,1	570.2240	0.8031	0.9624
T 22	320,63,42,38,34,27,23,22,21,17,12,10,6,3,3,2,2,2,1,1	1224.3741	1.3544	0.9432
T 23	368,89,70,51,35,31,29,22,22,17,10,8,6,4,4,4,3,2,1,1	1016.8348	1.0422	0.9443
T 24	139,60,36,31,30,19,8,7,7,3,2,2,2,2,2,1,1,1,1,	226.1784	0.5545	0.9621
T 25	297,84,43,37,34,28,27,25,24,22,14,9,5,5,4,4,3,2,2,2,1,1	814.8671	1.0300	0.9406
T 26	231,113,87,50,50,45,36,30,28,25,16,12,9,8,7,6,3,3,2,2,2,1,1	302.7430	0.3810	0.9449
T 27	144,67,34,23,19,18,15,13,12,5,4,3,3,3,3,3,2,2,1,1	258.6211	0.6252	0.9690
T 28	531,196,120,95,74,69,42,41,33,30,9,9,8,7,7,6,6,3,3,2,1	992.0104	0.6758	0.9511
T 29	316,165,67,45,31,25,17,16,14,13,12,10,5,5,4,4,4,3,2,2,1,1	607.7747	0.6627	0.9883
T 30	213,71,44,25,20,15,14,12,11,8,8,6,5,3,2,1,1,1,1	493.3146	0.8672	0.9721
T 31	174,57,40,26,18,13,9,8,8,8,7,6,4,3,2,2,2,1	378.1955	0.8136	0.9657
T 32	201,80,38,35,34,20,20,19,15,12,8,6,6,4,3,2,2,1	383.6211	0.9619	0.9446
T 33	216,79,58,47,39,36,23,19,17,12,9,6,5,4,2,2,1,1,1	340.8033	0.5476	0.9275
T 34	156,97,30,24,23,20,16,9,9,8,7,6,5,5,4,3	277.3671	0.5811	0.9629
T 35	204,77,65,34,29,28,19,14,10,10,9,8,7,5,4,3,1,1	344.5901	0.5909	0.9504
T 36	152,77,27,21,20,19,18,18,17,17,6,6,5,2,2,2,1,1,1,1,1	283.9115	0.6513	0.9418
T 37	256,96,65,59,37,34,28,28,20,14,8,8,8,7,2,2,1,1,1,1	422.5573	0.5785	0.9353
T 38	156,69,57,36,36,36,31,23,15,5,5,5,3,2,2,1,1,1,1	199.2217	0.3770	0.9286
T 39	207,146,57,45,37,26,24,19,14,10,8,7,7,7,1,1	330.0258	0.4678	0.9719
T 40	230,63,56,53,42,33,28,28,23,23,7,7,7,5,4,3,3,1,1,1	387.0156	0.6211	0.8687
T 41	226,109,78,56,49,45,27,17,13,11,11,4,4,4,2,2,2,2,1,1	317.8866	0.4272	0.9656
T 42	173,56,37,32,32,30,24,22,19,12,5,3,3,3,2,1,1, 1,1	307.2369	0.6543	0.8871

Another set of texts has been processed by Andreev, Popescu, Altmann (2016) and the Mandelbrot distribution has been successfully applied. In the next table, we show that the exponential function is adequate, too. The results are presented in Table 6.8 which shows rather the development of Russian texts because the texts are dated and ordered historically.

Table 6.8
Russian adnominals in the history of texts (Andreev, Popescu, Altmann 2016)

Text	**Frequencies**	**a**	**b**	R^2
Pushkin: *Vadim* 1822	72,40,28,12,6,5,5,4,3,3,1	123.9595	0.5597	0.9924

Pushkin: *Graf Nulin* 1825	70,44,18,13,8,6,4,4,4,2,2,1,1,1	124.6769	0.5802	0.9901
Pushkin: *Mednij vsadnik* 1833	142,78,41,34,14,11,10,7,5,4,4,2,2	242.6812	0.5570	0.9930
Derzhavin: *Felitsa* 1782	27,26,23,21,15,9,6,5,5,4,3,1,1,1,1	38.0224	0.2331	0.9377
Derzhavin: *Vodopad* 1791-1794	110,76,62,52,28,26,25,18,18,10, 5,4,3,2,2,2,1,1,1,	140.7249	0.2810	0.9883
Karamzin: *Bednaja Liza* 1792. Paragraphs 1-8	93,40,28,15,13,11,9,7,4,3,3,2,2, 1,1,1,1,	162.0550	0.6082	0.9746
Pushkin: *Vystrel* 1830. Part 1	93,73,36,30,18,17,16,14,12,5,4, 3,2,2,1,1,1,1,1,1,	133.3088	0.3649	0.9774
Lermontov: *Demon* 1839. Part 1	55,30,23,10,9,7,7,7,7,4,4,3,2,1, 1,1,1,1,	82.4230	0.4630	0.9700
Lermontov: *Fatalist* 1837-1840	112,74,43,25,22,18,17,15,8,5, 4,4,2,1,1,1,1,	164.8666	0.4149	0.9842
Turgenev: *Dvorjanskoe gnezdo* 1856-1858. Part 1	84,38,24,18,12,7,7,7,6,6,4,2,2, 2,1,1,1,1,	142.5053	0.5831	0.9748
Nekrasov: *Zheleznaja doroga* 1864	40,29,16,12,6,4,4,4,4,3,3,3,2,2, 1,1,1,	61.5948	0.4389	0.9859
Tolstoy: *Vojna i mir* 1863-1869. Chapter 2	81,34,29,21,15,13,11,11,4,2,2,2, 2,1,1,1,	115.5439	0.4584	0.9477
Tolstoy: *Anna Karenina*, 1873-1877. Chapt. 1-2	83,54,36,22,14,10,7,6,4,3,2,2,2, 1,1,1,1,1,1	127.5053	0.4401	0.9993
Chekhov: *Zhenshchina bez predrassudkov* 1883	30,29,9,7,6,4,4,3,2,2,2,2,1,1,1,1	49.6391	0.4520	0.9231
Chekhov: *Dama s sobachkoj* 1899. Chapters 1-2	107,39,30,27,23,20,6,5,5,5,3,2,1, 1,1,1,	164.5896	0.5306	0.9281
Kuprin: *Chari*, 1897	92,35,31,20,15,9,9,8,6,1,1,1,	148.8780	0.5601	0.9492
Kuprin: *Junkera*	125,39,22,13,13,12,9,7,6,4,3,	319.3346	0.9635	0.9696

1928-1932. Chapter 2	2,2,1,1,1,1,			
Bunin: *Kavkaz* 1937	100,31,18,16,13,6,5,5,3,2,2,1,1,	240.8851	0.9137	0.9652
Bunin: *Stepa* 1938	103,33,29,26,18,4,4,4,3,2,1,1,1, 1,1,1	184.4412	0.6554	0.9413
Kuprin, *Domik* 1929	125,21,19,16,10,6,5,5,4,4,3,3,2,1,	542.7748	1.4845	0.9594
Bunin: *Antonov-skie jabloki* 1900. Parts I-II	192,47,36,23,21,20,13,12,12,12, 4,3,2,1,1,1,1,1,	541.8644	1.0636	0.9442

The adnominals in Chinese have been studied by Pan and Liu (2014). Their results are presented in Table6.9. The titles of analyzed texts (here merely their abbreviation) can be found in the original article.

Table 6.9
Rank-frequency sequence of adnominals in Chinese (Pan, Liu 2014)

ID	Frequencies	a	b	R^2
LA1	28,23,17,16,12,10,7,7,6,5,5,5,3,3,2,2,2,2,1,1,1	33.1278	0.2149	0.9921
LA2	23,18,16,14,13,13,12,10,7,7,6,6,6,6,4,3,3,2,2, 4,2,1,1,1,1,1	24.0088	0.1365	0.9743
LA3	17,16,16,14,10,10,10,7,7,6,5,4,4,3,3,2,2,1,1,1, 1,1,1,1	20.9004	0.1511	0.9669
LA4	33,12,12,12,10,9,9,8,8,6,6,6,6,4,2,2,1,1,1,1,1, 1,1,1,1	28.3794	0.2032	0.8255
LA5	33,22,21,15,11,11,9,6,5,5,5,4,4,4,3,2,2,2,2,2, 2,1,1,1,1,1,1	37.7904	0.2328	0.9833
LA6	33,20,15,12,10,9,8,7,7,7,6,6,5,5,5,5,4,2,2,1,1, 1,1,1,1	32.6509	0.2118	0.9129
LA7	20,15,14,14,14,12,11,9,8,8,8,6,4,4,2,2,2,1,1,1, 1,1	21.4882	0.1340	0.9379
LA8	19,14,12,9,9,8,6,4,4,4,3,3,3,2,2,2,2,2,2,1,1,1,1	21.0114	0.2083	0.9861
LA9	23,20,16,15,12,11,11,10,9,6,4,3,3,3,3,2,2,2,2, 2,1,1,1,1,1	26.4094	0.1664	0.9786
LA10	26,22,20,12,11,10,9,8,7,6,5,5,5,5,4,4,3,3,2,2, 1,1	29.1110	0.1791	0.9707
LC1	11,11,11,11,9,5,4,4,4,4,4,3,3,3,2,2,2,2,1,1,1,1, 1	13.8307	0.1596	0.9202
LC2	47,25,16,13,13,12,9,9,8,7,7,7,7,6,6,5,4,4,3,3, 3,3,2,2,1,1,1	48.3167	0.2587	0.8684
LC3	39,27,20,15,12,11,9,8,8,7,7,5,5,5,3,3,3,2,2,2, 2,1,1,1	43.4612	0.2361	0.9660
LC4	63,30,22,16,12,11,11,11,10,9,7,7,6,5,5,4,4,3, 3,3,2,2,1,1,1,1,1,1,1	78.0105	0.3650	0.9042
LC5	28,28,15,14,12,12,11,11,10,7,7,6,4,3,3,3,2,2, 2,2,2,1,1,1,1,1,1,1,1,1,1	31.6906	0.1751	0.9540

LC6	59,17,16,16,14,11,11,11,10,9,9,8,8,6,5,4,4,2, 2,2,2,2,2,1,1,1,1	59.2037	0.3074	0.7644
LC7	48,16,15,11,10,9,9,9,7,6,6,5,4,3,2,1,1,1,1,1,1, 1,1,1	60.4094	0.4193	0.8483
LC8	67,26,24,17,16,16,14,13,11,9,8,8,8,7,7,4,4,4, 2,2,2,2,1,1,1,1,1	68.2987	0.2787	0.8512
LC9	27,16,14,13,11,10,9,6,6,5,5,4,4,3,2,2,2,1,1,1, 1,1,1,1, 1,1	27.6271	0.2022	0.9648
LC10	19,16,16,14,14,12,10,9,7,7,6,5,4,4,3,3,2,2,2,2, 2,2,2,1,1,1,1	21.7846	0.1368	0.9786
XSH1	9,5,4,4,2,2,1,1,1,1,1,1,1,1,1,1	12.3088	0.4687	0.9673
XSH2	12,10,7,6,5,3,3,2,2,2,1,1,1,1,1,1,1,1,1	15.4460	0.3019	0.9888
XSH3	6,6,3,2,2,2,2,2,2,1,1,1,1,1	7.6729	0.3587	0.8945
XSH4	11,8,6,3,3,2,2,2	15.4183	0.4153	0.9801
XSH5	11,10,5,4,3,2,2,2,2,2,2,2,1,1	15.4601	0.3799	0.9461
XSH6	7,6,4,3,3,2,2,2,2,1,1,1,1	8.4880	0.3189	0.9716
XSH7	7,3,2,2,1,1,1	15.0870	0.9353	0.9800
XSH8	7,6,4,4,4,4,3,3,2,2,1,1,1	7.2618	0.2951	0.9165
XSH9	16,10,10,7,7,4,4,3,3,2,2,2,1,1,1	18.6646	0.2725	0.9719
XSH10	9,6,5,4,4,3,2,2,2,1,1,1,1,1,1,1	10.4103	0.3065	0.9764
Y1	25,16,14,11,10,9,9,7,6,6,6,5,5,5,4,3,2,2,2,2,2, 1,1,1,1	23.9435	0.1697	0.9512
Y2	28,26,23,22,19,13,13,13,10,10,10,9,6,5,5,5,4, 4,4,3,1,1,1	32.5286	0.1347	0.9786
Y3	14,13,13,11,9,8,8,7,6,6,5,5,5,5,4,4,4,2,2,2,2,1, 1, 1,1,1,1,1,1,1,1	15.5991	0.1212	0.9730
Y4	25,22,20,12,12,12,9,9,8,8,7,7,7,6,6,6,5,5,3,3, 2,1,1,1,1,1	25.9539	0.1370	0.9471
Y5	37,13,13,11,10,10,7,7,7,5,5,4,4,3,3,2,2,1,1,1, 1,1,1,1	37.1807	0.2840	0.8473
Y6	36,33,33,26,21,18,14,13,12,9,7,6,6,6,3,3,2,2, 2,2,1,1	45.6101	0.1694	0.9804
Y7	27,20,17,15,15,15,11,10,10,6,6,5,5,5,4,4,4,4, 2,1,1,1,1,1,1,1	28.0329	0.1501	0.9718
Y8	27,22,19,19,14,12,11,11,11,10,9,8,7,6,5,5,4,3, 3,3,2,2,1,1,1	28.1316	0.1310	0.9778
Y9	18,17,16,15,14,13,9,8,7,6,5,4,4,3,3,2,3,2,2,2, 1,1	22.2119	0.1439	0.9618
Y10	37,27,22,14,13,12,5,5,5,4,4,4,3,3,3,3,2,2,2,2, 2,2,1	46.5440	0.2781	0.9834

The text LC6 can be better captured by a simple Menzerathian function with a = 54.6685, b = -1.0106, c = -0.0138, R^2 = 0.9256 (without adding 1) or with the power function (with 1) with a = 54.4881, b = -1.0245, R^2 = 0.9264.

6.2. Motifs of adnominals

As is well known, sequences of some qualitative entities – given in form of symbols – may be restructered in Köhlerian motifs (Köhler 2008, 2015; Köhler, Naumann 2008, 2010) in which no symbol may be repeated in the same motif, and the next motif may not contain two elements of the preceding motif. Since we have a number of Russian texts, we may try to construct the motifs of adnominals and fit to them the above exponential function. Of course, qualitative motifs may have at least two properties: the type of the motif and its length measured in terms of element numbers. Here we shall consider merely the first alternative. Andreev et al. (2017) fitted the power function (they mention also the exponential); here we shall apply the exponential. The results are presented in Table 6.10 and 6.11. They are to be read as follows (see T 1): „there are exactly 98 motif types occurring exactly once"; or, in the same line „there is exactly 1 motif-type occurring 76 times". Since we have to do with spectra – not with ranking – we must show both the variable X (= number of times) and the frequency f_x (number of motif types)

Table 6.10
Russian adnominal motifs in female and male texts
(Andreev, Popescu, Altmann 2017)

Text	Frequency									
T 1	1	2	3	4	6	7	8	11	12	76
f_x	98	11	6	2	1	1	2	2	1	1
Exp	97.95	11.88	2.21	1.14	1.00	1.00	1.00	1.00	1.00	1.00
	a = 863.8401, b = 2.1871, R^2 = 0.9978									
T 2	1	2	3	4	5	7	12	15	53	
f_x	123	11	4	2	2	1	1	1	1	
Exp	122.98	11.38	1.88	1.08	1.00	1.00	1.00	1.00	1.00	
	a = 1433.1329, b = 2.4637, R^2 = 0.9995									
T 3	1	2	3	4	5	6	15	19	73	
f_x	97	9	1	2	1	3	1	1	1	
Exp	97.00	8.91	1.65	1.05	1.00	1.00	1.00	1.00	1.00	
	a = 11.64.8577, b = 2.4960, R^2 = 0.9993									
T 4	1	2	3	4	5	6	17	20	69	
f_x	89	16	4	3	2	2	1	1	1	
Exp	88.97	16.28	3.66	1.46	1.08	1.01	1.00	1.00	1.00	
	a =506.3666, b =1.7503, R^2 = 0.9993									
T 5	1	2	3	4	5	6	7	17	49	
f_x	120	10	5	1	2	4	2	1	1	
Exp	119.98	10.52	1.76	1.06	1.00	1.00	1.00	1.00	1.00	
	a = 1487.0279, b = 2.5256, R^2 = 0.9983									
T 6	1	2	3	4	8	9	25			
f_x	71	12	4	1	1	1	1			
Exp	70.97	12.35	2.84	1.30	1.00	1.00	1.00			
	a = 431.2818, b = 1.8187, R^2 = 0.9971									

T 7	1 2 3 4 5 7 9 18 24 67
f_x	106 12 3 4 2 2 1 1 1 1
Exp	105.98 12-73 2.21 1.13 1.01 1.00 1.00 1.00 1.00 1.00
	a = 977.6226, b = 2.2313, R^2 = 0.9989
T 8	1 2 3 4 5 7 8 11 16 18 28 144
f_x	123 11 1 3 4 1 1 2 1 1 1 1
Exp	123.00 10.91 1.81 1.07 1.01 1.00 1.00 1.00 1.00 1.00 1.00 1.00
	a = 1501.4078, b = 2.5101, R^2 = 0.9999
T 9	1 2 3 4 5 6 8 11 27 99
f_x	95 12 1 3 1 2 1 2 2 1
Exp	95.01 11.79 2.24 1.14 1.02 1.00 1.00 1.00 1.00 1.00
	a = 819.0966, b = 2.1648, R^2 = 0.9988
T 10	1 2 3 4 6 7 8 10 11 13 16 18 150
f_x	122 15 5 2 1 3 1 1 1 1 1 1 1
Exp	121.96 15.58 2.76 1.21 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00
	a = 1003.9166, b = 2.1162, R^2 = 0.9992
T 11	1 2 3 4 5 6 7 8 9 10 76
f_x	85 7 3 2 2 1 1 1 1 1 1
Exp	84.99 7.24 1.46 1.03 1.00 1.00 1.00 1.00 1.00 1.00 1.00
	a = 1129.7008, b = 2.5990, R^2 = 0.9993
T 12	1 2 3 4 5 6 8 9 10 11 13 20 73
f_x	107 12 3 2 2 2 1 1 1 1 1 1 1
Exp	106.99 12.20 2.18 1.13 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00
	a = 1002.7842, b = 2.2472, R^2 =0.9998
T 13	1 2 3 4 5 7 8 9 17 20 60
f_x	107 13 5 2 1 2 1 2 1 1 1
Exp	106.96 13.63 2.51 1.18 1.02 1.00 1.00 1.00 1.00 1.00 1.00
	a = 889.0107, b = 2.1270, R^2 = 0.9991
T 14	1 2 3 4 6 8 9 13 60
f_x	102 10 4 3 1 2 1 1 1
Exp	101.98 10.45 1.88 1.08 1.00 1.00 1.00 1.00 1.00
	a = 1079.4311, b = 2.3693, R^2 = 0.9989
T 15	1 2 3 4 6 7 10 23 26 89
f_x	119 13 6 3 1 1 1 1 1 1
Exp	118.95 13.85 2.40 1.15 1.00 1.00 1.00 1.00 1.00 1.00
	a = 1082.6836, b = 2.2169, R^2 = 0.9995
T 16	1 2 3 4 5 6 7 8 9 12 13 18 28 129
f_x	116 11 7 3 3 1 1 1 1 3 1 1 1 1
Exp	115.95 12.01 2.05 1.19 1.01 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
	a = 1200.5462, b = 2.3460, R^2 = 0.9969
T 17	1 2 3 4 5 6 7 8 9 11 12 26 123
f_x	116 17 11 1 1 2 3 1 1 1 1 1 1
Exp	115.82 19.22 3.89 1.46 1.07 1.01 1.00 1.00 1.00 1.00 1.00 1.00 1.00
	a = 723.4494, b = 1.8406, R^2 = 0.9949
T 18	1 2 3 4 5 6 8 10 86
f_x	108 8 3 5 1 1 2 1 1
Exp	107.99 8.26 1.49 1.03 1.00 1.00 1.00 1.00 1.00
	a = 1576.8622, b = 2.6905, R^2 =0.9981

T 19	1 2 3 4 5 6 7 75
f_x	93 12 1 1 3 1 1 1
Exp	93.02 11.72 2.25 1.15 1.02 1.00 1.00 1.00
	a = 790.1305, b = 2.1502, R^2 = 0.9992
T 20	1 2 3 4 5 6 7 11 13 133
f_x	77 13 4 1 2 2 1 1 1 1
Exp	76.97 13.32 3.00 1.32 1.05 1.01 1.00 1.00 1.00 1.00
	468.5434, b = 1.8192, R^2 = 0.9994

Table 6.11 (Andreev et al. 2017)

T 21	1 2 3 4 6 7 8 12 20 32 97
f_x	75 13 4 2 1 1 2 1 1 1 1
Exp	74.97 13.37 3.07 1.35 1.01 1.00 1.00 1.00 1.00 1.00 1.00
	a = 442.4148, b = 1.7886, R^2 = 0.9995
T 22	1 2 3 4 5 6 9 11 21 161
f_x	78 7 4 4 5 1 1 1 1 1
Exp	77.98 7.49 1.55 1.05 1.00 1.00 1.00 1.00 1.00 1.00
	a = 913.6220, b = 2.4739, R^2 = 0.9940
T 23	1 2 3 4 5 7 8 9 12 15 20 79
f_x	92 10 3 3 4 1 1 3 1 1 1 1
Exp	91.98 10.29 1.95 1.10 1.01 1.00 1.00 1.00 1.00 1.00 1.00 1.00
	891.3649, b = 2.2821, R^2 = 0.9976
T 24	1 2 3 4 5 6 7 8 11 53
f_x	57 6 3 1 1 2 1 1 1 1
Exp	56.99 6.28 1.50 1.05 1.00 1.00 1.00 1.00 1.00 1.00
	a = 593.3790, b = 2.3607. R^2 = 0.9988
T 25	1 2 3 4 5 7 8 9 18 139
f_x	102 12 3 3 2 2 1 1 1 1
Exp	101.99 12.24 2.25 1.14 1.02 1.00 1.00 1.00 1.00 1.00
	a = 907.1837, b = 2.1954, R^2 = 0.9993
T 26	1 2 3 4 7 10 11 13 17 23 61
f_x	138 9 5 4 1 1 1 1 1 1 1
Exp	137.99 9.46 1.52 1.03 1.00 1.00 1.00 1.00 1.00 1.00 1.00
	a = 2217.1264, b = 2.7841, R^2 = 0.9987
T 27	1 2 3 4 5 6 10 14 61
f_x	74 3 2 1 1 1 1 1 1
Exp	734.00 3.05 1.06 1.00 1.00 1.00 1.00 1.00 1.00
	a =2595.6661, b = 3.5711, R^2 = 0.9998
T 28	1 2 3 4 5 6 7 9 10 11 15 16 22 37 41 203
f_x	145 26 6 3 2 2 1 1 3 2 1 1 1 1 1 1
Exp	144.07 26.33 5.46 1.78 1.14 1.02 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0 1.0
	a = 818.3264, b = 1.7377, R^2 = 0.9996
T 29	1 2 3 4 6 8 10 15 37 42 127
f_x	100 10 2 4 2 2 1 1 1 1 1
Exp	99.99 10.10 1.84 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00
	a =1076.3767, b = 2.3863, R^2 = 0.9988
T 30	1 2 3 4 6 10 14 101

f_x	74 6 2 2 2 2 1 1
Exp	74.00 6.10 1.36 1.02 1.00 1.00 1.00 1.00
	a = 1043.9007, b = 2.6603, R^2 = 0.9993
T 31	1 2 3 4 5 6 9 10 14 82
f_x	54 6 1 2 1 2 1 1 1 1
Exp	54.00 5.94 1.46 1.04 1.00 1.00 1.00 1.00 1.00
	a = 568.7328, b = 2.3731, R^2 = 0.9991
T 32	1 2 3 4 5 6 8 9 14 18 86
f_x	81 8 3 1 2 1 1 1 1 1 1
Exp	80.99 8.24 1.66 1.06 1.01 1.00 1.00 1.00 1.00 1.00 1.00
	a = 883.1317, b = 2.4016, R^2 = 0.9995
T 33	1 2 3 4 5 8 11 12 78
f_x	91 14 7 1 1 2 1 2 1
Exp	90.91 15.16 3.23 1.35 1.05 1.00 1.00 1.00 1.00
	a = 570.8594, b = 1.8483, R^2 = 0.9976
T 34	1 2 3 4 5 11 18 58
f_x	69 7 2 2 2 2 1 1
Exp	68.99 7.11 1.55 1.05 1.00 1.00 1.00 1.00
	a = 757.0482, b = 2.4100, R^2 = 0.9992
T 35	1 2 3 4 5 6 11 13 82
f_x	83 9 3 1 2 1 2 2 1
Exp	82.99 9.24 1.83 1.08 1.01 1.00 1.00 1.00 1.00
	a = 816.0695, b = 2.2979, R^2 = 0.9992
T 36	1 2 3 4 5 13 17 59
f_x	78 10 1 1 1 1 1 1
Exp	78.01 9.77 2.00 1.11 1.01 1.00 1.00 1.00
	a = 676.4277, b = 2.1728, R^2 = 0.9998
T 37	1 2 3 4 5 6 7 9 10 16 83
f_x	99 14 2 5 1 1 1 2 2 1 1
Exp	99.00 14.01 2.73 1.23 1.03 1.00 1.00 1.00 1.00 1.00 1.00
	a = 738.3434, b = 2.0195, R^2 = 0.9980
T 38	1 2 4 5 7 8 10 49
f_x	88 13 1 4 1 1 1 1
Exp	88.00 13.02 1.23 1.03 1.00 1.00 1.00 1.00
	a = 629.7669, b = 1.9795, R^2 = 0.9986
T 39	1 2 3 4 5 6 7 10 24 30 73
f_x	97 11 3 2 1 1 1 1 1 1 1
Exp	96.99 11.22 2.09 1.12 1.01 1.00 1.00 1.00 1.00 1.00 1.00
	a = 901.1661, b = 2.2395, R^2 = 0.9998
T 40	1 2 3 4 5 6 7 9 10 13 87
f_x	109 8 2 2 2 2 1 1 1 1 1
Exp	109.00 8.08 1.46 1.03 1.00 1.00 1.00 1.00 1.00 1.00 1.00
	a = 1646.4472, b = 2.7243, R^2 = 0.9997
T 41	1 2 3 4 5 6 9 10 13 21 75
f_x	102 13 4 3 1 2 3 1 1 1 1
Exp	101.97 13.44 2.53 1.19 1.02 1.00 1.00 1.00 1.00 1.00 1.00
	a = 819.2186, b = 2.0935, R^2 = 0.9988
T 42	1 2 3 4 6 9 73

f_x	79 4 6 6 3 1 1
Exp	78.99 4.46 1.15 1.01 1.00 1.00 1.00
	a = 1758.1756, b = 3.1155, R^2 = 0.9893

Other types of motifs can be found in Köhler's works. He called them (e.g. in 2015) L-motif = a set of non-decreasing length values, F-motif = a set of non-decreasing frequency values, P-motif = a set of non-decreasing polysemy values, T-motif = a set of non-decreasing polytextuality values, D-motif = a set of uninterrupted depth-first path of elements in a tree structure and in general, for qualitative motifs he defined R-motifs = uninterrupted sequence of unrepeated elements. As a matter of fact, any property presented in the form of a qualitative or quantitative sequence, can be transformed into some kind of motifs. The fact that for some motif types regularities were found is an encouraging idea for investigating sequential phenomena of any form. Motifs have some properties that can be exploited for text analysis. For parts-of-speech in an Italian text, he found 602 different R-motifs and set up a very long table which cannot be reprinted here. It would be very interesting to study several types of motifs and differentiate text types or even languages, comparing the frequencies of identical motifs. A very stimulating book concerning these problems has been edited by Mikros and Mačutek (2015).

Here we shall show merely the ranking of L-motifs. The data were found by P. Grzybek and taken from the Internet by Köhler (2012: 119). The L-motifs concern word lengths in the original Russian version of Dostoevsky's *Crime and Punishment.* Köhler ranked them decreasingly and fitted to them the right truncated modified Zipf-Alekseev distribution and the Zipf-Mandelbrot distribution with good results. Here we show merely the fitting of the exponential function presented in Table 6.12.

Table 6.12
Fitting the exponential function to word length motifs in Dostoevsky's novel (Köhler 2012)

Frequencies	**a**	**b**	**R^2**
825,781,737,457,389,274,269,247,245,235, 207,177,164,135,132,122,98,93,87,77	981.0193	0.1555	0.9464

Motifs have themselves various properties which can be further examined and one can define them based on any linguistic property. One can obtain secondary properties; e.g. with quantitative motifs, one can consider the length, the range, or the mean of them, and with qualitative motifs, one can consider their length, their in-advance-defined weight, etc.

6.3. Syllable types

Every language has a restricted number of syllable types. Usually, one writes them as CV, CVC, VC, … where C = consonant, V = vowel; however, one can define them more specifically. We conjecture that every language uses them with specific preferences. There are languages which do not have a combination of two consonants, e.g. CCV, there are languages without consonants at the end of a syllable but there are also languages having "syllabic" consonants and syllables like CCC, e.g. Slovak having [r], [l] etc. as syllabic (e.g. words like *krk, vlk, smrt`*). Whatever the situation, one can conjecture that the ranked frequency of syllable types in a text abides by the above mentioned exponential function. In order to illustrate this conjecture, we present the syllable types in Hungarian and German. To this end, we use the poem *Szeptember végén* by S. Petöfi in Hungarian and its translation into German by M. Remané (*September-Ausklang*). The results are presented in Table 6.13. They are, of course, not representative for the languages but merely for the given text.

As can be seen, the frequency of individual types differs in the two languages but the law holds true. In German, inspite of 11 types, the concentration of the main type is greater than in Hungarian. It can be seen on the representation of the most frequent type and the jump to the next. It is expressed also by the parameter *b* whose computation in many languages would be recommended. One can characterize the sequences in many various ways but here it would be premature.

Table 6.13
Syllable types in Hungarian and German

Rank	**Hungarian**			**German**		
	Syllable	**f_x**	**Exp**	**Syllable**	**f_x**	**Exp**
1	CV	105	113.06	CVC	120	116.17
2	CVC	87	67.24	CV	59	60.52
3	VC	34	40.16	VC	29	31.76
4	V	26	24.15	CVCC	26	16.90
5	CVCC	6	14.68	CCV	15	9.22
6	VCC	2	9.09	CVCCC	7	5.25
7	CCV	1	5.78	CCVCCC	6	3.19
8	CCVC	1	3.83	CCVC	5	2.13
9				VCC	4	1.59
10				V	3	1.30
11				CCVCCC	1	1.16
	a = 189.5724, b = 0.5257 R^2 = 0.9451			a =222.8492, b = 0.6601 R^2 = 0.9793		

7. Various Word Classes

7.1. Special word frequencies

Though we omit the fitting of the above function to word frequencies, we use it for some particular classes of words. Taking the results concerning **color** names obtained by A. Pawlowski (1999) from various frequency dictionaries, we present the fit in Table 7.1. Pawlowski analyzed 12 basic colours and if some of them did not occur he marked the frequency by zero. Since we apply a function ignoring the zeros, we can omit them.

Table 7.1
Frequency of colour names in 10 languages (Pawlowski 1999)

Language	**Data**	***a***	***b***	R^2
Czech	604,519,416,206,205,158,146,127,96,16, 4,3;	833.4061	0.2758	0.9638
English	365,203,197,176,143,116,55,48,23,13,12,7	445.7312	0.2718	0.9467
French (Juilland 1971)	136, 113, 74, 58, 35, 20, 17, 7;	200.4302	0.3414	0.9799
French (Engwall 1984)	298,278, 170, 134, 101, 98, 77, 61, 12, 9;	411.4612	0.2689	0.9600
Italian	155, 122, 115, 91, 79, 55, 40, 22, 16, 11;	201.0900	0.2263	0.9607
Polish	93, 87, 52, 39, 36, 29, 24, 19, 8, 2, 2;	128.1275	0.2739	0.9622
Romanian	165, 104, 78, 75, 64, 60, 18;	198.7026	0.2635	0.9059
Russian	473,471,371,317,216,116,109,88,49,30, 22,16;	689.7777	0.2461	0.9496
Slovak	473,461,315,275,181,104,71,58,43,19,19,7	699.5725	0.2800	0.9618
Spanish	141, 102, 71, 51, 44, 41, 21, 6, 6, 3;	193.8425	0.3263	0.9814
Ukrainian	310,282,193, 175, 118, 114, 50, 36, 23, 10, 4	436.3971	0.2679	0.9638

As can be seen, the parameter *b* is relatively stable. In two cases (French and Spanish) it is somewhat greater (> 0.3) but it may depend on the frequency dictionary. Further languages should be analyzed.

L. Uhlířová (2007) studied the frequencies of referential adjectives in Czech texts which are, unfortunately, not named. The 17 adjective types are given. She obtained the results presented in Table 7.2. She fitted the modified Zipf-Alekseev distribution but it may be shown that the exponential function is sufficient, too.

Table 7.2
Frequency of individual referential adjectives in Czech texts (Uhlířová 2007)

Type	**Frequency**	**a**	**b**	**R^2**
Czech referential adjectives (Uhlířová 2007)	221,201,80,28,19,9,4,4,3, 2,2,2,1,1,1,1,1	408.2034	0.5196	0.9408

7.2. Borrowings

Another closed class is the class of borrowings from foreign languages. The set of borrowings is usually called "etymological spectrum" (cf. Finkenstaedt & Wolff 1973). One can find them either in the etymological dictionary or, if one knows them, one can state their frequency in any texts, even in a corpus. The number of borrowings changes every year. In general their number increases – until purists begin to work! The most innovations can be observed in press texts which concern events in other countries, too, and make reference to all views of life. Hence, an historical study of borrowings comparing, say, each January issue of a magazine since its beginning up to now would be a good picture of language evolution. This change abides by the Piotrowski law but surely one will find a number of exceptions. Fortunately, using an etymological dictionary one can perform the work mechanically.

K.H. Best (2008c) showed the numbers of borrowings in Turkish from 24 languages (cf. Table 7.3).

Wolff (1969) examined the etymological spectrum in English press texts taking into account each token (cf. Best 2006). We conjecture that the exponential function does not work for text mixtures in which tokens are examined. Such a counting does not concern a population and does not represent a sample. For setting up the „population" of borrowings, it is better to consider a dictionary and consider types, or to take merely a single text in order to get a sample.

H. Körner (2004) analyzed the pure borrowings in German; Altmann (1993) applied a special function with very good results (cf. also Best 2006: 90); here we fit the exponential function. Only „pure" borrowings are accepted. The results can be found in Table 7.4.

Table 7.3
Number of borrowings in Turkish (Best 2008c)

Source language	Rank	Number of borrowings	Computed
Arabic	1	6463	6844.35
French	2	4974	3748.28
Persian	3	1374	2052.94
Italian	4	632	1124.60
English	5	538	616.26
Greek	6	399	337.91
Latin	7	147	185.48
German	8	85	102.02
Russian	9	40	56.32
Spanish	10	36	31.29
Slavic (without Russian)	11	24	17.59
Armenian	12	23	10.08
Hungarian	13	19	5.97
New Greek	14	14	3.72
Mongolian	15	13	2.49
Hebrew	16	9	1.82
Bulgarian	17	8	1.46
Japanese	18	7	1.24
Portuguese	19	4	1.13
Finnish	20	2	1.07
Norwegian	21	2	1.04
Albanian	22	1	1.02
Korean	23	1	1.01
Soghdic	24	1	1.01
$a = 12496.4279$, $b = 0.6022$, $R^2 = 0.9607$			

K.H. Best (2010) studied the borrowings in Japanese and obtained the results presented in Table 7.5. Here the last row consists of the information „Andere Geber-Sprachen“ (*Other Source Languages*) and the number of words is 70 but in this form it is not possible to study it. There are surely languages (how many?) which have small numbers of words. In this form it must be omitted. The fit is optically not quite satisfactory but the greatest number shows the prevalence of English.

Table 7.4
Fitting the exponential function to borrowings in German
(Körner 2004)

Source language	Rank	Number of borrowings	Computed
Latin	1	2031	2077.73
French	2	1424	1247.83
Low German	3	545	749.57
English	4	519	450.43
Italian	5	286	270.83
Greek	6	144	163.00
Dutch	7	87	98.26
Slavic	8	44	59.39
Spanish	9	43	36.06
Rotwelsh	10	41	22.05
North German	11	12	13.64
Japanese	12	10	8.59
Hungarian	13	10	5.56
Turkish	14	8	3.73
Jiddish	15	6	2.64
Arabic	16	5	1.99
Portuguese	17	5	1.59
Celtic	18	4	1.36
Eskimo	19	3	1.21
Gothic	20	3	1.13
Hindi	21	3	1.08
Afrikaans	22	2	1.05
Islandic	23	2	1.03
Malay	24	2	1.02
Chinese	25	1	1.01
Finnish	26	1	1.01
Hebrew	27	1	1.00
Hunnic	28	1	1.00
Ladino	29	1	1.00
Persian	30	1	1.00
Polynesian	31	1	1.00
$a = 3459.0228$, $b = 0.5102$, $R^2 = 0.9863$			

Table 7.5
Fitting the exponential function to borrowings in Japanese (Best 2010)

Source language	**Rank**	**Number of borrowings**	**Computed**
English	1	2853	2851.08
German	2	260	296.22
French	3	194	31.58
Dutch	4	73	4.17
Latin	5	73	1.33
Italian	6	69	1.03
Greek	7	37	1.00
Russian	8	20	1.00
Spanish	9	11	1.00
Portuguese	10	8	1.00
a = 27514.4068, b = 2-2674 , R^2 = 0.9937			

7.3. Parts of speech

Usually, one classifies the individual POS cognitively, i.e. one considers a whole class, e.g. adjectives, and performs a semantic classification. Ascribing each adjective to a class, one obtains a classification according to another criterion. Needless to say, one can further subdivide each class and study the frequencies. The first step has been done by N. Yesypenko (2009) who found subclasses of adjectives, nouns, verbs and adverbs and analyzed the work of three writers, namely E. Waugh, *A Handful of Dust*, J. Swift, *Gulliver's Travels* and M. Twain, *The Adventures of Tom Sawyer*. The adjectives were ordered in the following classes: 1. Traits of characterization, 2. Physical/natural condition, 3. Intellectual capacity, 4. Appearance, 5. Senses, 6. Age/time, 7. Temperature/sound, 8. Shape/ size, 9. Flavour, 10. Weight, 11. Degree/intensity, 12. Color, 13. Actions done to the object, 14. Positive evaluation, 15. Evaluation of length/ distance /position of the object, 16. Evaluation of value/function of the object, 17. Materia, 18. Negative evaluation. The nouns were subdivided in 25 classes, the verbs in 27 and the adverbs in 7. In the linguistic literature one can find a great number of similar classifications. It is again to be remarked that any classification of this kind – one can find a number of books and articles about this theme – represents our human view, and differs from language to language and from linguist to linguist. The results of Yesypenko's analysis are presented in Table 7.6.

Table 7.6
Semantic subclasses of POS-types in English

Author	Frequencies	a	b	R^2
Nouns				
Waugh	281,133,119,118,102,88,70,70,45,45, 35,33,30,24,24,21,15,12,12,12,10,8,8, 7,3	271.6542	0.1796	0.9174
Swift	222,154,138,102,102,92,86,68,62,54, 52,36,34,32,32,26,20,18,12,10,6,6,4, 4,2,	227.8855	0.1534	0.9756
Twain	142,128,122,112,106,100,86,80,70, 54,50,28,24,24,22,20,20,18,10,10,6, 6,4,2,2,	175.4079	0.1243	0.9619
Verbs				
Waugh	271,148,131,113,91,66,58,50,41,39, 31,30,24,23,23,18,17,16,15,9,7,4,3, 3,3,2	291.6524	0.2307	0.9544
Swift	174,148,104,84,70,66,52,50,48,32,30, 30,28,24,22,22,22,18,18,8,6,4,2,2	192.5670	0.1728	0.9713
Twain	202,184,132,122,78,68,58,52,44,42,3 6,30,30,30,26,26,20,20,20,12,8,8,6,6 , 4,2	239.3513	0.1825	0.9754
Adjectives				
Waugh	122,118,100,68,67,59,48,45,44,42,22, 22,19,7,6,2,2	152.0448	0.1662	0.9646
Swift	120,99,93,75,66,63,60,57,54,33,18, 12,6,3	140.5740	0.1508	0.9209
Twain	93,87,81,75,69,63,54,48,42,42,21,18, 12,12	113.8698	0.1236	0.9295
Adverbs				
Waugh	215,153,141,66,66,49,5	304.5463	0.3320	0.9430
Swift	135,111,108,63,18,3	205.6041	0.3365	0.8361
Twain	111,111,96,75,54,15,12,	162.5641	0.2470	0.8341

7.4. Verbs

Köhler (2005, 2012) studied the polysemy of German verbs and their valency using the data given in Helbig and Schenkel (1991). His results concerning polysemy are presented in Table 7.7. Köhler fitted the negative binomial distribution.

Table 7.7
Variants of German verbs (Köhler 2005,2012)

Type	**Frequencies**	**a**	**b**	**R^2**
Semantic variants	218,118,73,42,18,8,4,2,2,1	390.7289	0.5892	0.9977

N. Yesypenko (2008) studied also the classification of verbs in 27 classes in the literature written in English and obtained for individual writers the results presented in Table 7.8.

Table 7.8
Types of verbs in English (Yesypenko 2008)

Author	**Frequencies**	**a**	**b**	**R^2**
E.Waugh	553,324,301,259,212,187,172, 141,73,71,66,66,54,53,49,41,29, 27,24,17,15,13,10,10,4,3,1	588.4105	0.1985	0.9685
E.Hemingway	683,452,408,378,297,265,257, 184,121,110,98,95,83,53,44,37, 35,33,32,30,18,14,11,10,9,6,6,	742.3368	0.1797	0.9737
K. Vonnegut	742,372,286,223,192,186,180, 126,108,107,64,62,51,50,45,40, 39,28,27,24,19,17,12,9,2,1	802.1791	0.2633	0.9213
I. Show	467,418,393,360,265,178,170, 142,140,113,86,83,78,72,63,62, 61,55,51,42,38,28,16,13,8,5,2,	569.4047	0.1552	0.9784
N.Mailer	631,395,307,305,295,278,233, 194,156,142,114,101,69,64,62, 58,56,50,6,42,42,35,34,22,13, 10,7	618.8359	0.1537	0.9566
F.M.Ford	559,346,313,241,227,220,216, 117,,98,96,94,85,66,63,58,49, 43, 40,38,23,21,19,17,15,11,2,2,	572.6376	0.1753	0.9611

Laufer and Nemcová (2009) studied various properties of German verbs and ordered them in each case. In order to show the classification we present the results explicitly in Table 7.9. Some computed numbers deviate strongly from the observed ones but the determination coefficient is satisfactory.

Table 7.9
Types of German verbs in SMS (Laufer, Nemcová 2009)

Type	**fr**	**Exp**	**Modal verbs**	**fr**	**Exp**	**Tense**	**fr**	**Exp**
Full verb	1956	1940.96	Können	92	90.13	Present	1902	1900.87
Copula	323	443.91	Müssen	53	56.95	Perfect	240	256.20
Auxiliary	296	102.12	Wollen	33	36.13	Preterit	89	35.28

Modal Modifying	230 10	24.06 6.27	Sollen Mögen Dürfen	31 17 5	23.05 14.85 9.69	Future Pluperfect	38 3	5.60 1.62
a = 8497.2013, b = 1.4771 R^2 = 0.9618			a = 141.9663, b = 0.4650 R^2 = 0.9750			a = 14143.9291, b = 2.0075 R^2 = 0.9984		

The semantic subcategories of actants of German verbs have been studied by Köhler (2012: 98 ff.). He applied successfully the Poisson distribution and used a German valency dictionary. Here, we show the fit of the exponential function. The results are presented in Table 7.10.

Table 7.10
Frequencies of semantic subcategories of German verbs (Köhler 2012)

Number of alternatives	**Number of verbs**	**Exp**
1	1796	1816.13
2	821	736.82
3	242	299.29
4	73	121.92
5	9	50.02
6	1	20.87
a = 4488.5645, b = 0.9029, R^2 = 0.9940		

7.5. Adverbs

Pelegrinová and Altmann (2017) studied the ranking of adverbial classes in various Czech texts. They applied the Zipf-Alekseev function with good results. They considered the following classes: place, time, manner, means, aspect, condition, measure, cause, purpose, concession, originator, result, and origin. The fitting of the exponential function is presented in Table 7.11. The texts used are:

1. **ŠKODA, Jan.** Nejslavnější padouch komiksových příběhů zastíní i hrdinného Batmana. In: *Reflex* [online]. [cit. 22. 12. 2016]. Accessible from: http://www.reflex.cz/clanek/causy/75224/nejslavnejsi-padouch-komiksovych-pribehu-zastini-i-hrdinneho-batmana.html
2. **BYSTROV, Michal**. Producent George Martin: Pátý Beatle se nebál odhazovat hudební konvence. In: *Reflex* [online]. [cit. 22. 12. 2016]. Accessible from: http://www.reflex.cz/clanek/causy/75229/producent-george-martin-paty-beatle-se-nebal-odhazovat-hudebni-konvence.html
3. **ŠKODA, Jan.** Eliška Junková. In: *Reflex* [online]. [cit. 22. 12. 2016]. Accessible from: http://www.reflex.cz/clanek/causy/76003/eliska-junkova.html
4. **ŠAFRÁNEK, Šimon.** Hollywood v Číně. In: *Reflex* [online]. [cit. 22. 12. 2016].Accessible from: http://www.reflex.cz/clanek/causy/75999/hollywood-v-cine.html

5. **BYSTROV, Michal** 2016: Accessible from: http://www.reflex.cz/clanek/causy/74555/nejslavnejsi-cesky-loupeznik-vaclav-babinsky-dozil-v-klastere.html
6. **MATYÁŠOVÁ, Judita** 2016, Accessible from: http://www.lidovky.cz/udelam-tam-tropickou-lasku-ddg-/kultura.aspx?c=A161223_132845_ln_kultura_hep
7. **VÁŇA, Jan** 2016, Accessible from: http//www.h7o.cz/zit-technologie
8. **KLÍČOVÁ, Eva** 2016, Accessible from: http://www.h7o.cz/uzitecny-idiot-z-prazske-kavarny/
9. **PUSKELY, Martin** 2016, Accessible from: http://www.h7o.cz/obrana-kritiky/#sthash.vczHeCO2.dpuf
10. **VYBÍRAL, Zbyněk** 2016, Accessible from: http://www.h7o.cz/kriticky-ano-ale-proc-vulgarne/#sthash.n9ob6oBO.dpuf
11. **ŠRÁMEK, Fráňa** 1916, *Splav*. Praha: Mladá fronta, 1972, p. 431-436.
12. **ŠRÁMEK, Fráňa** 1916, *Romance*. Praha: Mladá fronta, 1972, p.413-417
13. **ČAPEK, Josef** 1946. *Na spánek.* From: ČAPEK, Josef, *Básne z koncentračního tábora*, p. 19-26. Praha: Fr. Borový.
14. **WOLKER, J.** *Svatý kopeček.* From: Wolker, J. *Host do domu.* [online]. V MKP 1. vyd. Praha: Městská knihovna v Praze, 2011. [cit. 7. 2. 2017]. Accessible from:
http://web2.mlp.cz/koweb/00/03/37/00/11/host_do_domu.pdf
15. **ERBEN, K. J.** *Svatební košile*. From: Erben, K. J. *Kytice*. Praha: Albatros, 1965, p. 31-43.
16. **ČELAKOVSKÝ, F.L.** *Prokop Holý.* From: Čelakovský, F. L. *Ohlas písní českých.* [online]. V MKP 1. vyd. Praha: Městská knihovna v Praze, 2011, p. 23-26. [cit. 7. 2. 2017] Accessible from:
http://web2.mlp.cz/koweb/00/03/37/00/52/ohlas_pisni_ceskych.pdf
17. **BŘEZINA, O.** *Se smrtí hovoří spící…* From: Březina, O. *Stavitelé chrámů.* [online]. V MKP 1. vyd. Praha: Městská knihovna v Praze, 2011, p. 24-29. [cit. 7. 2. 2017] Accessible from:
https://web2.mlp.cz/koweb/00/03/59/20/93/stavitele_chramu.pdf
18. **KAINAR, J.,** *Lazar a píseň.* Praha: SNKLU 1960.

Table 7.11
Adverbials in Czech texts: ranking of class sizes
(Pelegrinová, Altmann 2017)

Text	Frequency	a	b	R^2
T1	56,52,38,19,6,5,4,3,2,1,1	92.1902	0.3950	0.9311
T2	77,64,37,14,4,3,2,1,1,1	136.5372	0.5001	0.9491
T3	77,55,34,16,7,2,1,1,1,1	133.5204	0.5149	0.9785
T4	65,34,26,17,9,7,2,1,1	104.3342	0.5108	0.9876
T5	100,79,34,18,6,3,3,3,2	181.7207	0.5419	0.9616
T6	128,46,24,9,4,4,3,2,1	326.0979	0.9494	0.9971
T7	66,58,31,23,2,2,1,1	113.6935	0.4626	0.9248

T8	36,35,30,7,4,3,3,2,1	60.7193	0.4008	0.8518
T9	44,28,13,8,8,4,3,3,2,2,2	74.7915	0.5465	0.9894
T10	61,40,18,15,6,3,1,1,1	107.7572	0.5620	0.9873
T11	72,47,41,22,6,3,2,2,2,1	115.6037	0.4547	0.9614
T12	100,53,32,11,4,3,1,1	195.0494	0.6693	0.9934
T13	38,37,22,13,4,2,2,1,1,1	64.4763	0.4249	0.9222
T14	91,44,40,19,4,4,2,2,1	156.8240	0.5690	0.9710
T15	95,54,37,6,2,1,1,1	183.6763	0.6475	0.9718
T16	71,26,23,3,2,2,1,1,1,1,1	158.7471	0.8316	0.9739
T17	113,39,19,12,11,5,2,1,1	284.0612	0.9427	0.9890
T18	43,33,18,9,5,5,4,2	70.7278	0.4753	0.9765

Adverbs have the ability to modify verbs or adjectives in different ways. Again, the results depend on the classification of verbs or adjectives – which is not unique. V.V.Silnickij (1989: 66) defined 20 adjective properties and searched for the strength of their specification by adverbs in five directions: intensity, quality, time, condition, modality. In Table 7.12, we shall order the numbers in the five specification categories, that is, we ignore the 20 properties of adjectives and obtain for each category the rank-order.

Table 7.12
Specification of adjectives by adverbs in 5 dimensions (Silnickij 1989)

	Intensity		**Quality**		**Time**		**Condition**		**Modality**	
Rank	**f_x**	**Exp**	**f_x**	**Exp**	**f_x**	**Exp**	**f_x**	**Exp**	**f_x**	**Exp**
1	666	687.51	99	99.73	53	51.70	7	7.82	30	30.72
2	554	580.09	80	87.09	42	39.72	7	5.35	26	21.56
3	534	489.48	68	76.06	24	30.57	4	3.77	11	15.22
4	404	413.05	63	66.45	23	23.59	2	2.77	9	10.84
5	394	348.58	63	58.07	18	18.25	2	2.13	7	7.81
6	314	294.20	61	50.76	17	14.17	1	1.72	6	5.71
7	243	248.32	54	44.39	12	11.06	1	1.46	5	4.26
8	180	209.66	53	38.83	8	8.69			5	3.25
9	167	176.98	52	33.99	7	6.87			4	2.56
10	166	149.45	34	29.77	7	5.48			4	2.08
11	142	126.22	22	26.08	6	4.42			3	1.75
12	119	106.63	16	22.87	5	3.61			3	1.52
13	114	90.10	16	20.07	4	3.00			2	1.36
14	47	76.16	16	17.63	3	2.53			1	1.25
15	36	64.40	4	15.50	1	2.16				
16	31	54.48	3	13.64	1	1.89				
17	23	46.11	3	12.02	1	1.68				
18	17	39.05	3	10.61	1	1.52				
19	13	33.10	2	9.38						

20	7	28.08	1	8.31						
	a = 813.8455, b = 0.1702, R^2 = 0.9847		a= 113.2261 b = 0.1370, R^2 = 0.9182		a = 66.3828, b = 0.2695, R^2 = 0.9808		a = 10.7019 b = 0.4504 R^2 = 0.8853		a = 42.9636 b = 0.3685, R^2= 0.9446	

7.6. Pronouns

Laufer and Nemcová (2009) studied the rank order of pronoun types in SMS and obtained the results presented in Table 7.13.

Tabelle 7.13
Pronouns in German SMS (Laufer, Nemcová 2009)

Rank	Pronoun type	Fr	Exp
1	Personal	1683	1679.40
2	Indefinite	216	259.09
3	Demonstrative	148	40.69
4	Possessive	122	7.10
5	Interrogative	114	1.94
6	Reflexive	107	1.14
7	Relative	26	1.02
a = 10914, b = 1.8723, R^2 = 0.9758			

The jump from the first to the second rank is too great but the rest of the frequencies are very similar; hence, the higher ranks do not show a good agreement. Nevertheless, one can search for a function which is “better”.

Laufer and Nemcová (2009) analyzed also the adverb in two dimensions: syntactic and semantic and obtained the results presented in Table 7.14.

Table 7.14
Syntactic and semantic aspects of German adverbs in SMS
(Laufer, Nemcová 2009)

Rank	**Synt. function**	**fr**	**Exp**	**Rank**	**Sem. function**	**Fr**	**Exp**
1	Modal	4189	4173.93	1	Temporal	896	879.57
2	Sentential	680	830.94	2	Modal	209	295.57
3	Deictic	541	166.06	3	Interrogative	180	99.76
4	Relational	49	33.83	4	Local	123	34.11
				5	Consecutive	27	12.10
				6	Concessive	10	4.72
				7	Causal	9	2.25
				8	Instrumental	5	1.42

	a = 20981.4584, b = 1.6150, R^2 = 0.9849		a = 2620.3924, b = 1.0928, R^2 = 0.9644

Laufer and Nemcová (2009) studied also the frequency of particles in SMS and obtained the results presented in Table 7.15.

Table 7.15
Particles in SMS (Laufer, Nemcová 2009)

Rank	Type of particle	fr	Exp
1	Prepositions	629	650.21
2	Conjunctions	538	462.08
3	Tuning particles	249	328.46
4	Negations	221	233.56
5	Answers	220	166.17
6	Conjunctional	114	118.30
7	Intensive	103	84.31
8	Focusing	38	60.17
9	Modal words	21	43.02
a = 914.1168, b = 0.3422, R^2 = 0.9535			

7.7. Individual words

Busse (1998: 486) studied the occurrence and forms of the word „pray" in all the stage plays by Shakespeare. He found ten environments, and stated their frequency, and K.-H. Best fitted to the empirical distribution the mixed Poisson distribution. It can, again, be shown that the exponential function is sufficient. The data and the computation are presented in Table 7.16.

Table 7.16
Occurrence of „pray" in all environments in Shakespeare's dramas (Best 2006)

Rank	1	2	3	4	5	6	7	8	9	10
Frequency	253	134	132	101	71	54	21	3	3	1
Computed	241.19	169.64	119.40	84.13	59.36	41.98	29.77	21.20	15.18	10.96
a = 342.1116, b = 0.3537, R^2 = 0.9521										

7.8. Personal names

Family names are not unique in any language. There are always variants which developed historically and remained in the new form even if the bearers moved from one place to another. K.-H. Best (2007) studied the distribution of variants of personal names. Using the data collected by Kunze (2004), he fitted the negative binomial distribution to the number of individual variants stated according to telephone connections. The name *Burghard* occurred in 68 variants, the name *Schmidt* only in 5. The number of occurrences is presented in Table 7.17. In the last line the frequencies of the name *Lang – Lange – Langen – Langer* are presented as stated by Seibicke (1982: 174) in four German cities. K.-H. Best (2006: 84) fitted to the last line the Hirata-Poisson distribution, but here we apply the exponential function.

Table 7.17
Variants of personal names in telephone

Name	**Frequencies**	**a**	**b**	**R^2**
Burghard (Best 2007)	6834,4348,3230,2383,1855,1581, 1456,1213,1001,762,677,616,540, 525,511,450,446,410,395,287, 192,192,138,121,98,95,86,85,73, 61,60,59,58,54,46,32,31,30,24,23, 22,21,19,19,18,18,16,13,12,11,10, 6, 5,5,4, 4,4, 3,3,3,1,1,1,1,1,1,1,1	7971.3202	0.2601	0.9715
Weißflog (Best 2007)	530,91,60,51,19,13,13,8,3,2,2,2,2, 1	2477.1644	1.5488	0.9826
Schmidt (Best 2007)	390,33,18,10,8	4356.7860	2.4162	0..9971
Jakobus (Best 2007)	16156,13157,10068,8626,8576, 3433,3422,2804,2620,2029,1967, 1674	20653.5479	0.2309	0.9683
Family names	9707, 5537, 1880, 704, 140,77, 7 *according to day names* (Best 2007)	20930.4492	0.7441	0.9860
Fries (Best 2007)	4728,3018,2938,1844,1708,1379, 413,177,135,117,60,36,26,24,19, 17,16,10,9,9,8,8,5,5,5,4,4,4,3,3, 2,2, 1,1,1,1,1,1,1	6424.2882	0.3126	0.9759
Lang (Seibicke 1982; Best 2006)	118, 30, 20, 1	402.3291	1.2406	0.9857

8. Compounds

Compounds present a number of problems, some of which can be expressed quantitatively. First of all, it is the number of components of a compound which is especially conspicuous in chemistry. Further, there is a POS of the individual components. There are compounds like noun + noun or noun + adjective, etc. Some languages prefer a special kind. The compounds, when written, are expressed either with a blank between the components or with a hyphen or they are written together. In some compounds, one finds different joining: some components are placed directly behind one another – as is usual in analytic languages, other ones use case endings or even conjunctions or prepositions between the components. All these properties can be scaled and the frequency of individual classes can be counted. One can also consider only one type of compounds and study its semantic diversification.

The last aspect was examined by Rather and Rothe (1991) who analyzed the semantic diversification of German compounds *noun* + *noun* and found 18 categories presented in Table 8.1. The resulting parameters of the fitting of the above function are presented in the Table 8.1.

Table 8.1
Types of compounds (Rather, Rothe 1991)

Type	**Frequencies**	**a**	**b**	**R^2**
German:Noun + noun compounds (Rather, Rothe 1991	60,59,54,51,29,26,20,20,19, 15,15, 8,6, 5,3,1,1,1	80.6349	0.1825	0.9513

All possible types of Ukrainian compounds in two computer texts were collected and their frequencies stated by Ishutin and Gnatchuk (2017). They found 11 types (such as noun+noun, noun+verb,…). Shmidt and Gnatchuk (2016) studied the types of compounds in the German technical book "Wirtschaftsinformatik" by H.R. Hansen et al. (2015) and found 221 compounds. The compounds in English newspapers (*The New York Times)* have been examined by Gnatchuk (2015a), in English scientific texts by Gnatchuk (2015b) and in the English prose by Gnatchuk (2015c).

In the above articles the function $y = a*b^x$ was applied with very good results. Applying the exponential function we obtain the results presented in Table 8.2.

Table 8.2
Frequencies of compound types in Ukrainian and English (Gnatchuk et al.)

	Frequencies	a	b	R^2
Ukrainian compounds	50,33,11,8,4,4,3,2,1,1,1	94.6612	0.6301	0.9803
German compounds	159,10,10,6,5,3,3,3,3,2,2,2,2, 1,1,1,1,1,1,1,1,1,1,1	2461.4218	2.7461	0.9944
English prose	528,96,91,45,43,31,27,26,21,21, 11,10,6,5,3,2,1,1	2027.3047	1.3579	0.9577
English scientific texts	172,30,15,5,5,4,4,4,4,3,3,3,3,2,2, 2,2,2,2,2,2,1,1,1	898.5151	1.5511	0.9944
English press	64,12,7,4,4,3,3,3,3,2,2,2,2, 2,2,2,1,1,1,1,1,1,1,1,1,1,1	307.8439	1.5897	0.9863

A similar investigation has been performed by J. Mackuliak (2007) who counted the different types of nominal compounds in three German text types: literature, press and science. In scientific texts, she found 44 nominal compound types. Of course, the presence of compounds of the same type is not always given in all text types but they can be separately ordered and the function can be fitted to them. The results of counting are presented in Table 8.3. Here not only two-component compounds are taken into account, but also there are some consisting of 6 components. The parameter *b* is in German slightly greater than that in English.

Table 8.3
Frequencies of German nominal compounds in three text types
(Mackuliak 2007)

Text type	Frequencies of compound types	a	b	R^2
Literature	6954, 998, 807, 339, 300, 74, 74, 52, 52, 49, 42, 27, 25, 17, 13, 13, 10, 10, 8, 8, 7, 7, 5, 5, 5, 5, 4, 3, 3, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1	38664.6503	1.7191	0.9875
Press	1797, 221, 137, 115, 79, 44, 39, 31, 21, 18, 18, 15, 12, 11, 10, 8, 7, 7, 6, 6, 5, 4, 3, 3, 3, 3, 2, 2, 2, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,	12558.5145	1.9463	0.9887
Science	2152, 262, 225, 148, 69, 59, 45, 39, 27, 27, 23, 23, 18, 17, 9, 9, 7, 7, 6, 6, 6, 4, 3, 2, 2, 2, 2, 1, 1, 1, 1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,	14186.6843	1.8887	0.9850

Sowinski (1979) studied the length (in terms of the number of components) of compounds in ten German advertizing texts (cf. also Best 2006: 47) and stated their frequency. The result and the fit of the exponential function are presented in Table 8.4. Here "Length" is the number of components of the compound.

Table 8.4
Length of compounds in ten German advertizing texts (Sowinski 1979)

Length	2	3	4	5	6	7	a = 2236.2530
Frequency	192	63	10	1	1	1	b = 1.2277
Computed	192.91	57.22	17.47	5.83	3.41	1.41	R^2 = 0.9960

Sowinski (1979) studied also the types of compounds and obtained the classes [Noun + noun, Adjective + noun, Noun + adjective, Verb + noun, Adjective + adjective, Verb + adjective] but considered merely the "main" ones, Best (2006: 82) fitted to the ranks the negative binomial distribution. The fit of the exponential function is presented in Table 8.5.

Table 8.5
Frequencies of compound types in German (Best 2006)

Length	1	2	3	4	5	6	a = 723.8360
Frequency	117	16	13	8	7	3	b = 1.8335
Computed	116.70	19.49	3.96	1.47	1.08	1.01	R^2 = 0.9820

The fact that the components of a compound have a different cohesion degree is well known. One can classify them in various ways, e.g. beginning with "written separately", "joined with a hyphen", "written together", using various grammatical means like cases, conjunctions, prepositions, abbreviating some components, etc. One may obtain a scale and study a dictionary of a language. Fan and Altmann (2007) developed a method for scaling the cohesion and analyzed three languages: English, German and Hungarian. They ranked the frequencies of individual classes and obtained the results presented in Table 8.6.

Table 8.6
Cohesions within compound in English, German and Hungarian
(Fan, Altmann 2007)

English			German			Hungarian		
Rank	f_x	Exp	Rank	f_x	Exp	Rank	f_x	Exp
1	496	548.60	1	547	552.70	1	766	765.70
2	451	316.30	2	242	210.58	2	76	81.46
3	165	182.55	3	37	80.62	3	32	9.47
4	36	105.53	4	27	31.25	4	21	1.89
5	26	61.19	5	21	12.49	5	17	1.09
6	13	35.66	6	16	5.36	6	12	1.01
7	9	20.95	7	10	2.66	7	11	1.00
8	7	12.49	8	8	1.63	8	11	1.00

9	6	7.62	9	8	1.24	9	5	1.00
10	5	4.81	10	8	1.09	10	4	1.00
11	4	3.19	11	7	1.03	11	4	1.00
12	3	2.26	12	5	1.01	12	3	1.00
13	3	1.73	13	5	1.00	13	2	1.00
14	3	1.42	14	4	1.00	14	2	1.00
15	3	1.24	15	3	1.00	15	2	1.00
16	2	1.14	16	3	1.00	16	2	1.00
17	2	1.08	17	3	1.00	17	2	1.00
18	2	1.05	18	3	1.00	18	2	1.00
19	2	1.03	19	3	1.00	19	1	1.00
20	1	1.02	20	2	1.00	20	1	1.00
21	1	1.01	21	2	1.00	21	1	1.00
22	1	1.01	22	2	1.00	22	1	1.00
23	1	1.00	23	2	1.00	23	1	1.00
24	1	1.00	24	2	1.00	24	1	1.00
25	1	1.00	25	2	1.00	25	1	1.00
26	1	1.00	26	2	1.00	26	1	1.00
			27	2	1.00	27	1	1.00
			28	2	1.00	28	1	1.00
			29	2	1.00	29	1	1.00
			30	2	1.00	30	1	1.00
			31	1	1.00	31	1	1.00
			32	1	1.00	32	1	1.00
			33	1	1.00	33	1	1.00
			34	1	1.00	34	1	1.00
			35	1	1.00	35	1	1.00
			36	1	1.00	36	1	1.00
			37	1	1.00	37	1	1.00
			38	1	1.00	38	1	1.00
			39	1	1.00	39	1	1.00
			40	1	1.00	40	1	1.00
			41	1	1.00	41	1	1.00
			42	1	1.00	42	1	1.00
			43	1	1.00	43	1	1.00
			44	1	1.00	44	1	1.00
			45	1	1.00			
			46	1	1.00			
			47	1	1.00			
			48	1	1.00			
a = 951.0475, b = 0.5520, R^2 = 0.9333			a = 1452.2974, b = 0.9679, R^2 = 0.9899			a = 7267.9289, b = 2.2517, R^2 = 0.9973		

Though placewise the expected numbers strongly differ from the observed ones, the fit is acceptable. We are sure that for the same languages other researchers would obtain other numbers but the parameters would probably remain very similar. In any case, one sees that compounds diversify in their cohesion and the diversification can be captured by the exponential function.

9. Syntax

9.1. Sentence types

Sentences can be classified according to various criteria. They stay at the top of the pyramid of speech entities and can attain different properties as grammatical entities, and as speech acts. They can be classified according to the kind of text (e.g. poetic, scientific, receipt, instruction, etc.). One can try various approaches, the mystery of sentences will never be disclosed.

We present here the simple classification of sentences to 4 types according to Dshurjuk and Levickij (2002). Their names will be abbreviated in English as SS = simple sentence, SSu = sentence succession, JS = joint sentences, CS = syntact-ically complex sentences. The authors analyzed German texts according to text type and obtained the results presented in Table 9.1.

Table 9.1
Occurrence of four sentence types in 6 text types (Dshurjuk, Levickij 2002)

Tex types	**Sentence type**			
	SS	***SSu***	***JS***	***CS***
Novels	4841	1734	3839	290
Short stories	1946	691	1727	215
Politics	4268	1117	2677	62
Economics	2631	575	1373	12
Science	2110	460	1100	10
Culture	2670	647	1769	61

If one orders the individual lines, one obtains the results presented in Table 9.2.

Table 9.2
Sentence types in different texts

Type	**Frequencies**	**a**	**b**	**R^2**
Novels	4841, 3839, 1734, 290	8794.3177	0.5342	0.8947
Short stories	1945, 1727, 691, 215	3437.8209	0.4890	0.8540
Politics	4268, 2677, 1117, 62	8667.5751	0.6755	0.9528
Economics	2631,1373, 575, 12	5833.9772	0.7797	0.9782
Science	2110, 1100, 460, 10	4682.1255	0.7806	0.9784
Culture	2670, 1769, 647, 61	5406.7940	0.6659	0.9419

As can be seen, the results are acceptable even if the last empirical value is usually too small. In cases where the above function does not work, one should reorganize the classification, i.e. define more classes and reorder the data.

Köhler (2005a) analyzed the sentence structures and used the results from Helbig, Schenkel (1991). The results are presented in Table 9.3

Table 9.3
Sentence structures in German (Köhler 2005a)

Sentence structures	**a**	**b**	**R^2**
286,78,73,71,58,45,32,22,17,15,11,11,11,10,9, 7,7,7,7,7,6,6,6,6,5,5,5,5,5,4,4,4,4,4,4,3,3,3,3,3, 3,3,3,3,3,3,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2,2, 2,2,2,2,2, 2,2,2,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1, 1,1,1,1,1,1,1,1,1,1,1,1,1,1,1	454.3323	0.5762	0.9025

Levickij, Pavlyčko and Semenyuk (2001) studied the frequency of types of subordinate clauses with German writers Böll, Kant, Mann and Remarque (the works are given in the References), and obtained 13 types with different ranking with individual writers. The frequencies and the results of computing the exponential function are displayed in Table 9.4.

Table 9.4
Subordinate clauses in German (Levickij, Pavlyčko, Semenyuk 2001)

Writer	**Frequencies**	**a**	**b**	**R^2**
Böll	194,124,122,42,28,20,18,16,14,10,6,4,2	292.3684	0.4036	0.9603
Kant	158,146,84,82,42,30,20,8,12,6,6,4,2	236.1260	0.3242	0.9677
Mann	282,96,92,20,20,18,18,16,16,12,6,2,2	583.3383	0.7607	0.9571
Remarque	166,134,74,66,28,28,28,24,16,12,8,8,8	237,3865	0.3441	0.9755

Levickij and Romanova (1997) studied the ascription of tenses to verbs joined with an adverb in English. The types of adverbs are given by the frequency classes, the tenses are shown separately. The results are presented in Table 9.5.

Table 9.5
Tenses with verbs joined with an adverb in English (Levickij, Romanova 1997)

Tense	**Frequencies**	**a**	**b**	**R^2**
Indefinite	2998,1362,1326,774, 286,191	4786.5416	0.5063	0.9504
Continuous	145,94,46,31,15,11	250.6630	0.5374	0.9930
Perfect	383,271, 248, 65, 41,31	614.4508	0.4287	0.9092
Perfect continuous	6,6,5,4,4	6.1107	0.1484	0.8905

K.-H. Best (2006: 52) analyzed the depth of German sentence members (= number of attributes) in Pestalozzi's *Hühner, Adler und Mäuse* and obtained the results presented in Table 9.6. A member without any attributes has depth x = 1, etc.

Table 9.6
Depth of German sentence members (Best 2006)

Depth	1	2	3	4	a = 162.3800
Frequency	46	14	4	1	b = 1.2815
Computed	46.08	13.51	4.47	1.96	R^2 = 0.9989

K.-H. Best (2006) analyzed also the frequencies of sentence members classified according to their categories in the same text. He obtained nominal, verbal, prepositional, adverbial and adjectival sentence members. The distribution (in the given order) is presented in Table 9.7.

Table 9.7
Number of sentence members according to their category (Best 2006)

Rank	1	2	3	4	5	a = 48.6826
Frequency	28	19	11	4	3	b = 0.5568
Computed	28.90	16.99	10.16	6.25	4.01	R^2 = 0.9739

Computing the function of the sentence members, Best (2006: 67) obtained for the same text the results presented in Table 9.8. Here the sequence of ranks is Predicate, Adverbial determination, Subject, Object, Predicative noun, Prepositional object, Vocative.

Table 9.8
Ranked functional sentence members (Best 2006)

Rank	1	2	3	4	5	6	7	a = 27.7440
Frequency	19	15	14	9	4	2	1	b = 0.3569
Computed	20.42	14.59	10.51	7.65	5.66	4.26	3.28	R^2 = 0.9022

9.2. Syntactic alternatives

The number of properties one can find in a sentence is just as infinite as with any other linguistic entity. The research has begun and develops very slowly because it is still held back by the shadows of the past. But one can continue evens, if one introduces quantification step by step. Here we shall touch merely on two further properties. The first is the counting of syntactic alternatives in dependence on . their position in the mother constituent. The study was proposed by R. Köhler

(2012) who analyzed the Susanne corpus. He obtained the results presented in Table 9.9. He applied the function $log\ y = a + bx^c$. Here, we shall try to do it with the Mezerathian function though the exponential function used everywhere yields a satisfactory $R^2 = 0.8892$ (a = 53.1320, b = 0.1893). The problem is the very slow decrease of values.

Tabel 9.9
Number of syntactic alternatives in dependence on the position in their mother constructions (Köhler 2012).

Position	**Number of alternatives**	**Menz.**
1	38	36.29
2	38	40.48
3	35	36.87
4	33	30.90
5	25	24.75
6	22	19.26
7	16	14.70
8	12	11.05
9	7	8.22
10	3	6.06
11	2	4.43
12	1	3.23
a = 53.0209, b = 0.7049 $R^2 = 0.9787$		

9.3. Semantic roles

A special study by Haitao Liu (2012) has been dedicated to the examination of semantic roles of autosemantic words in the sentence. Liu considers the semantic roles a result of diversification. He performed the analysis using 8 texts of the news from China Central Television and applied the modified Zipf-Alekseev distribution with an excellent result. The great distance between the first and second values show that there is some subsidiary condition which can be well captured by the Zipf-Alekseev distribution. Here we shall use the exponential function which yields also satisfactory results, as shown in Table 9.10. The frequencies are ranked.

Table 9.10
Distribution of semantic roles in dependency trees (Liu 2012)

Rank	Frequency	Exp
1	168	167.83
2	16	19.02
3	15	2.95
4	12	1.21
5	8	1.02
6	6	1.00
7	5	1.00
8	5	1.00
9	4	1.00
10	3	1.00
11	2	1.00
12	2	1.00
13	2	1.00
14	2	1.00
15	1	1.00
a = 1544.4447, b = 2.2255 R^2 = 0.9842		

10. Environment

B.I. Ginka (1983) and V.V. Levickij (1989) studied the (functional) environment (i.e. something like valency) of German words meaning “work”. Using the data presented by Levickij (1989: 61) and ordering them according to frequency, we obtain the data presented in Table 10.1.

Table 10.1
Functional environments of German words meaning “work”
(Ginka 1983; Levickij 1989)

Word	Frequencies	a	b	R^2
Arbeit	586, 475, 206, 184, 151, 93, 88, 37, 26, 24, 20, 5	877.0065	0.3812	0.9712
Werk	226,137,89,67,65,63,42,18,10,1,1	300.3264	0.3488	0.9595
Mühe	227, 59, 25, 7, 5, 4, 4, 3, 2, 2, 1	814.4751	1.2848	0.9979
Geschäft	114, 79, 46, 33, 31, 15, 10, 7, 6, 3	169.4343	0.4044	0.9918
Dienst	74, 64, 15, 15, 15, 12, 5, 4, 2, 1	126.6203	0.4943	0.9135
Leistung	37, 14, 13, 12, 11, 6, 5, 4, 3, 1	48.2344	0.4188	0.8859
Tätigkeit	36, 29, 27, 16, 15, 7, 6, 4, 2, 1	50.3889	0.2977	0.9541

Evidently, this type of analysis has an infinite continuation. One can take any word, search for its synonyms or functional environments, perform the analysis for different texts, separate the text types, compare them, perform the analysis in several languages, etc.

As a matter of fact, this search can be compared with that of the search for valency. Now, almost each word has a valency. The deeper it lies in the hierarchy, the greater its valency. In logic, one considers it as predication. For nouns, the predicates of first order are verbs and adjectives. The predicates of verbs and adjectives are for the noun predicates of second order, etc. The situation can be presented by means of a graph. Theorists of dependency perform the analysis by means of dependency bows and measure the distance in terms of words lying between the two words.

Linguist should try to bring order into this discipline. It can influence syntax and semantics, text analysis, typology, etc. It is situated somewhere between Belza-chains, hrebs and other supra-sentence structures but it has no fixed place up to now.

11. The Study of Hrebs

Hreb is a suprasentence construct consisting of a set of sentences having something in common. It may be a word, a pronoun, a reference etc.. L. Hřebíček, their discoverer, called them originally "sentence aggregates" but later on they obtained the name "hreb" in his honor. The study of hrebs is already a well developed science, studying structures which are greater than a sentence. Hřebíček himself analyzed 11 Turkish texts and prepared tables containing the hreb length and the number of hrebs of this length. The titles of texts can be found in his book (Hřebíček 1992). He applied the Menzerathian function but we can show that a slightly simpler exponential function is sufficient. The results are presented in Table 11.1.

Table 11.1
Fitting the exponential function to hrebs in Turkish texts (Hřebíček 1992)

T 1			T 2			T 3		
Length	Number	Exp	Length	Number	Exp	Length	Number	Exp
1	230	228.96	1	296	295.58	1	264	263.83
2	60	65.85	2	48	52.55	2	71	71.78
3	23	19.45	3	21	10.02	3	19	20.06
4	18	6.25	4	10	2.58	4	11	6.13
5	12	2.49	5	2	1.28	5	6	2.38
6	4	1.42	6	3	1.05	6	2	1.37
7	2	1.12	7	1	1.01	12	1	1.00
8	4	1.03	9	1	1.00	13	1	1.00
9	1	1.01	10	1	1.00	19	2	1.00
10	1	1.00	12	2	1.00	24	1	1.00
12	1	1.00				27	1	1.00
13	1	1.00						
14	1	1.00						
15	1	1.00						
16	1	1.00						
18	1	1.00						
20	1	1.00						
21	1	1.00						
22	1	1.00						
27	1	1.00						
a =801.2809, b = 1.2570. R^2 = 0.9942			a = 1683.4060, b = 1.7430, R^2 = 0.9973			a = 975.8826. b = 1.3118, R^2 = 0.9994		

T 4			T 5			T 6		
Length	Number	Exp	Length	Number	Exp	Length	Number	Exp
1	207	206.47	1	378	377.75	1	245	244.48
2	55	59.00	2	53	56.35	2	52	56.36
3	26	17.37	3	20	9.13	3	22	13.59
4	2	5.62	4	4	2.19	4	6	3.86
5	1	2.30	5	4	1.18	5	5	1.65
6	5	1.37	6	2	1.03	6	4	1.15
7	1	1.10	7	1	1.00	7	2	1.03
8	1	1.03	8	1	1.00	9	2	1.00
11	1	1.00	12	1	1.00	12	1	1.00
16	1	1.00				15	1	1.00
17	1	1.00				16	2	1.00
18	1	1.00						
a = 727.9081, b = 1.2649, R^2 = 0.9970			a = 2564.4871, b = 1.9179, R^2 = 0.9988			a = 1070.9150, b = 1.4812, R^2 =0.9978		

T 7			T 8			T 9		
Length	Number	Exp	Length	Number	Exp	Length	Number	Exp
1	114	114.01	1	92	92.08	1	238	237.25
2	21	20.88	2	23	22.34	2	70	74.55
3	4	4.50	3	5	6.00	3	31	23.90
4	2	1.62	4	1	2.17	4	5	8.13
5	2	1.11	7	1	1.02	5	10	3.22
15	1	1.00				6	5	1.69
17	1	1.00				7	3	1.22
						8	5	1.07
						9	2	1.02
						11	1	1.00
						12	2	1.00
						14	2	1.00
						20	1	1.00
						23	2	1.00
						26	1	1.00
						45	1	1.00
a = 642.3960, b = 1.7377, R^2 = 0.9999			a =388.7951, b = 1.4513, R^2 = 0.9995			a = 758.9253, b = 1.1670, R^2 = 0.9970		

T 10			T 11		
Length	Number	Exp	Length	Number	Exp
1	263	261.30	1	254	253.62
2	80	88.47	2	47	50.40
3	37	30.39	3	16	10.66
4	18	10.88	4	14	2.89

5	14	4.32	5	2	1.37
6	2	2.12	6	3	1.07
7	7	1.37	7	1	1.01
8	1	1.13	8	1	1.00
9	3	1.04	10	2	1.00
12	1	1.00	16	1	1.00
13	1	1.00	18	1	1.00
14	2	1.00	25	1	1.00
15	1	1.00			
18	1	1.00			
21	1	1.00			
22	2	1.00			
29	1	1.00			
32	1	1.00			
43	1	1.00			
a = 774.6581, b = 1.0906, R^2 = 0.9956			a = 1291.9851, b = 1.6320, R^2 = 0.9971		

Hřebíček also presented tables of sign aggregations but the frequencies are very similar and we can omit them. Later on, Hřebíček (1995) presented an analysis of further Turkish texts. Since the aggregate lengths are always different, we present the second table in the same way as the first. The results are presented in Table 11.2. Hřebíček shows that hrebs can be specialized, e.g. taking into account only sentences having a special character. The last two texts T 11 and T 12 are separately analyzed texts in Hřebíček (1990), namely T 11: Gökalp, Z. (1970). *Türkçülügün esaslari*. Istanbul: Millî Eğitim, and T 12: Yunus Emre, a poem published in A. Gölpinarli (1965), *Yunus Emre. Risâlat al-nushiyya va dîvân*. Istanbul: Sulhi Garan.

Table 11.2
Hreb lengths and numbers of hrebs of this length in 10 Turkish texts
(Hřebíček 1995)

T 1			T 2			T 3		
Length	Number	Exp	Length	Number	Exp	Length	Number	Exp
1	273	271.34	1	364	361.33	1	192	
2	83	91.41	2	116	128.57	2	69	
3	39	31.23	3	55	46.16	3	31	
4	14	11.11	4	29	16.99	4	14	
5	17	4.38	5	12	6.66	5	13	
6	5	2.13	6	10	3.00	6	7	
7	7	1.38	7	10	1.71	7	3	
8	4	1.13	8	1	1.25	8	6	
9	2	1.04	9	4	1.09	9	1	
10	2	1.01	10	4	1.03	10	2	
12	1	1.00	11	3	1.01	11	2	
13	1	1.00	12	1	1.00	15	1	

17	2	1.00	13	2	1.00	31	1	
19	1	1.00	15	1	1.00	32	1	
			18	2	1.00			
			22	2	1.00			
a = 808.4257, b = 1.0954, R^2 = 0.9949			a = 10177945, b = 1.0384, R^2 = 0.9956			a = 493.5500, b = 0.9600, R^2 = 0.9959		

T 4			T 5			T 6		
Length	Number	Exp	Length	Number	Exp	Length	Number	Exp
1	150	149.28	1	260	257.01	1	323	321.24
2	44	47.40	2	76	90.18	2	99	107.76
3	16	15.52	3	41	32.07	3	44	36.59
4	13	5.54	4	29	11.82	4	17	12.86
5	8	2.42	5	12	4.77	5	16	4.96
6	5	1.44	6	7	2.31	6	12	2.32
7	3	1.14	7	5	1.46	7	7	1.44
8	4	1.04	8	2	1.16	8	6	1.15
9	2	1.01	9	4	1.06	9	6	1.05
10	1	1.00	10	2	1.02	10	4	1.02
11	2	1.00	11	1	1.01	11	2	1.01
12	2	1.00	12	1	1.00	13	1	1.00
14	1	1.00	13	3	1.00	14	1	1.00
15	1	1.00	14	1	1.00	15	1	1.00
18	1	1.00	15	2	1.00	17	1	1.00
19	2	1.00	18	1	1.00	19	1	1.00
23	1	1.00	19	3	1.00	21	1	1.00
			21	1	1.00	46	1	1.00
			28	1	1.00	49	1	1.00
			29	1	1.00	54	1	1.00
a = 473.8391, b = 1.1618, R^2 = 0.9940			a = 734.9079, b = 1.0545, R^2 = 0.9895			a = 960.6189, b = 1.0985, R^2 = 0.9955		

T 7			T 8			T 9		
Length	Number	Exp	Length	Number	Exp	Length	Number	Exp
1	171	170.01	1	148	147.38	1	190	189.55
2	57	61.93	2	40	43.68	2	39	42.54
3	28	22.97	3	17	13.45	3	15	10.15
4	11	8.92	4	11	4.63	4	11	3.02
5	5	3.86	5	1	2.06	5	7	1.44
6	5	2.03	6	3	1.31	6	1	1.10
8	3	1.13	7	2	1.09	7	2	1.02
9	1	1.05	8	4	1.03	8	1	1.00
11	1	1.01	9	4	1.01	9	1	1.00
13	2	1.00	10	1	1.00	10	1	1.00

30	1	1.00	14	1	1.00	11	1	1.00
			15	1	1.00	12	1	1.00
			16	1	1.00	14	1	1.00
a = 468.7910, b = 1.0202, R^2 = 0.9973			a = 502.0334 b = 1.2324, R^2 = 0.9955			a = 855.8995, b = 1.5128, R^2 = 0.9959		

T 10			T 11			T 12		
Length	Number	Exp	Length	Number	Exp	Length	Number	Exp
1	287	285.51	2	48	47.78	2	17	17.00
2	65	74.99	3	10	11.69	3	5	5.00
3	32	20.24	4	6	3.44	6	1	1.06
4	20	6.00	5	4	1.56	7	1	1.02
5	14	2.30	6	4	1.13	8	2	1.00
6	9	1.34	9	1	1.00	9	1	1.00
7	4	1.09	11	1	1.00	14	1	1.00
8	6	1.02	13	1	1.00			
9	3	1.01	15	1	1.00			
10	2	1.00	16	2	1.00			
12	1	1.00	22	1	1.00			
14	1	1.00	23	1	1.00			
19	1	1.00	31	1	1.00			
23	1	1.00	45	1	1.00			
33	1	1.00						
a = 1094.0428, b = 1.3469, R^2 = 0.9911			a = 896.3949, b = 1.4764, R^2 = 0.9877			a = 255.7721, b = 1.3859, R^2 = 0.9953		

For some other results, see also C. Schwartz (1995).

Köhler and Naumann (2007) studied the phrase-hrebs, defining them quite differently, and using the Text *Der Erdstern,* they obtained a long sequence of hreb-sizes which were ranked. The authors applied the negative hypergeometric and the Zipf-Mandelbrot distributions with great success. The data are presented in Table 11.3 where we used the above mentioned exponential function. The result shows that no definition is fixed for ever.

Table 11.3
Rank-frequency of phrase hrebs in German (Köhler, Naumann 2007)

Rank	Number	Exp	Rank	Number	Exp	Rank	Number	Exp
1	22	21.44	21	1	1.00	41	1	1.00
2	9	10.71	22	1	1.00	42	1	1.00
3	6	5.62	23	1	1.00	43	1	1.00
4	4	3.20	24	1	1.00	44	1	1.00
5	3	2.04	25	1	1.00	45	1	1.00
6	2	1.50	26	1	1.00	46	1	1.00

7	2	1.24	27	1	1.00	47	1	1.00
8	2	1.11	28	1	1.00	48	1	1.00
9	2	1.05	29	1	1.00	49	1	1.00
10	2	1.03	30	1	1.00	50	1	1.00
11	2	1.01	31	1	1.00	51	1	1.00
12	2	1.01	32	1	1.00	52	1	1.00
13	1	1.00	33	1	1.00	53	1	1.00
14	1	1.00	34	1	1.00	54	1	1.00
15	1	1.00	35	1	1.00	55	1	1.00
16	1	1.00	36	1	1.00	56	1	1.00
17	1	1.00	37	1	1.00	57	1	1.00
18	1	1.00	38	1	1.00	58	1	1.00
19	1	1.00	39	1	1.00			
20	1	1.00	40	1	1.00			
a = 42.9941, b = 0.7437, R^2 = 0.9797								

As can be seen, the fit is very satisfactory. The length of hrebs shows the text concentration which can be expressed by means of an indicator. The longer they are, the more the phrases or sentences concern a particular object. The study of hrebs has been done inonly a few languages, and hence it is not quite clear what role the given language plays.

12. Belza-Chains

Belza-chains (Belza 1971; cf. also Skorochod`ko 1981) differ from hrebs only by the fact that the respective sign or a reference must occur in the subsequent sentence. A chain is then a unit containing the same sign, and its length can easily be computed, if one considers only its basic form. The name "Belza-chain" was proposed later on. The length of chains shows the text concentration. Chen and Altmann (2015) analyzed a number of texts and studied the length of Belza-chains. In the original article they applied the Zipf-Alekseev function, but here we shall use the simpler exponential. The results are presented in Table 12.1. Not all data can be captured by the exponential function because in some of them (especially the Chinese ones) the sequences are bell-shaped. These texts are usually very short, otherwise we expect that they would exhibit the same exponential property. Here L = length of chain, N = number of chains of length L. The texts are:

German: Goethe J.W.v,: *Der Erlkönig***.** **Slovak:** Bachletová, E. (2014). Leto v nás. In: *Riadky bytia*. Bratislava, and Svoráková, S. (2003). Čakanie na Štraussa. Rev. of: Tomáš Štrauss: Metamorfózy umenia XX. storočia. Bratislava: Kalligram, 2001. *D art - Revue súčasného výtvarného umenia* 10, 37. **Czech:** Havel, V**.** (1990, 1991). *End-of-Year Speeches*. **Italian:** Napolitano, G. (2013). *End-of-Year speech.* **Hungarian:** Petöfi, S. (1847). *Szeptember végén* (poem).

Though the results are satisfactory (for the given texts), it can be re-commended not to take very short text for analysis. Text length is a subsidiary condition which may be the condition for the rejection of a model. Some of the data sets have a bell-shape and cannot be captured by the exponential function.

Table 12.1
Length of Belza-chains in texts of 5 languages (Chen, Altmann 2015)

German (Goethe)			Slovak (Bachletová)			Slovak (Svoráková)			Hungarian (Petöfi)		
L	N	Exp	L	N	Exp	L	N	Exp	L	N	Exp
1	7	6.54	1	15	14.05	1	7	7.92	1	6	6.33
2	3	4.45	2	7	7.16	2	8	5.72	2	5	3.97
3	4	3.13	5	2	1.53	3	3	4.21	3	2	2.66
4	3	2.34	8	1	1.05	6	2	2.02	4	2	1.93
6	1	1.52				7	1	1.69	5	1	1.52
a = 8.9061, b = 0.4744, R^2 = 0.8049			a = 31.5824, b = 0.8173, R^2 = 0.9980			a = 10.1558, b = 0.3835, R^2 = 0.7934			a = 9.5519, b = 0.5834, R^2 = 0.9006		

Italian (Napolitano)			Czech (Havel 1990)			Czech (Havel 1991)		
L	N	Exp	L	N	Exp	L	N	Exp
1	25	26.20	1	45	47.19	1	21	22.52
2	17	13.49	2	31	22.32	2	18	15.18
3	7	7.19	3	4	10.84	3	10	10.34
4	1	4.07	4	2	5.54	4	9	7.15
5	1	2.52	5	3	3.10	5	2	5.05
6	1	1.75	7	1	1.45	6	3	3.67
7	1	1.37	8	1	1.21	8	1	2.16
8	1	1,19				16	1	1.04
a = 50.8284, b = 0.7017, R^2 = 0.9565			a = 100.0909, b = 0.7733, R^2 = 0.9278			a = 32.6711, b = 0.4174, R^2 = 0.9424		

It must be noted that Belza-chains can be set up also if one considers also the referents, synonyms etc. of the given concept. What is more, a Belza-chain can be defined also in poetry, e.g. in hexameter, in the form of the same sequence of feet. It can be conjectured that these sequences will be short but there are also poems in which it is prescribed to use a certain rhythm. The possibilities are enormous, not only considering text and languages but also entities which were not even touched upon, up to now.

13. Word Frequencies

The stating of word frequencies in a text is, as a matter of fact, a study of the diversification of words. There are texts (Dada) in which some few words are repeated steadily but in „normal“ texts, one diversifies the vocabulary. Sometimes one computes the relative richness of the text referring to the diversity of the vocabulary.

The ordering of word frequencies is, perhaps, the most privileged theme for those who like to perform computations using a computer. One needs not care for anything: a word begins after a pause and ends in front of a pause. One takes not only enormous corpora but one frequently separates individual books, articles and presents the results which are usually fascinating. There is only a tiny problem: all these data are mixtures. Even the full text of a stage play is a mixture, a long novel is a mixture, etc. One considers them usually as populations but as Orlov (1982) said, there are no populations in language or text. If a writer makes a pause in order to drink coffee or to go to bed, the text is not any more the same text. While writing, some neurons are stimulated, one develops a special tendency, but as soon as a pause is inserted, the stimulation dies out or changes and we have to do with another text. Of course, some texts are created in one go and nothing is changed any more (neither by the author nor by the editors) – in that case it is a representative sample of author’s production. In the majority of cases one cannot ask the authors, not only because the authors lived 100 years ago but usually one does not know them and if so, the author does not remember. Thus the only „correct“ approach is to consider separately individual chapters in a book, individual persons in the stage play, distinguish direct and indirect speech in novels, etc. that is, to make any sample as homogeneous as possible.

In word counting, there are two perspectives: ranking and spectrum. For ranking, Zipf presented his power distribution/function which was later modified in various ways but in most cases it holds true. The spectrum takes into account the number of words occurring once, twice, etc. The ranking and the spectrum can be mutually transformed but one obtains various distributions/functions. The examination is in full course, and one considers this problem as a simple task for the computer. But up to now one does not know whether one should count types or individual forms or tokens, that is, the problem has three aspects. If one considers only types, eliminating any morphological changes (forms), one makes a synthetic language into an analytic one. Usually, the researchers do not say how they performed the computation; the computer counts each form of a word as a separate word.

Here we shall restrict ourselves to Turkish data presented by Hřebíček (1997), in which he separated at least the individual chapters of D. Özlü’s (1990) text. In this way we obtain counts for eight chapters. The fit is presented in Table 13.1. Needless to say, this type of investigation was the most popular at the beginning because one could do everything by means of a computer. Readymade programs are available also on the Internet. The numbers in the table are to be read as follows: there are exactly 377 words occurring exactly once in the first chapter, etc. We work here with the spectrum, not with the ranking.

Table 13.1
Word frequency spectrum in Turkish text by Özlü (Hřebíček 1997)

Chapter 1			Chapter 2			Chapter 3			Chapter 4		
x	f_x	Exp	x	f_x	Exp	x	f_x	Exp	x	f_x	Exp
1	377	375.15	1	426	422.39	1	345	342.28	1	292	290.17
2	124	132.08	2	122	140.82	2	102	115.61	2	65	77.04
3	50	46.92	3	65	47.39	3	50	39.49	3	34	21.00
4	27	17.09	4	31	16.39	4	28	13.92	4	28	6.26
5	15	6.64	5	14	6.11	5	12	5.34	5	11	2.38
6	10	2.97	6	16	2.69	6	12	2.46	6	3	1.36
7	10	1.69	7	12	1.56	7	4	1.49	7	4	1.10
8	3	1.24	8	9	1.19	8	8	1.16	8	3	1.03
9	4	1.08	9	9	1.06	9	9	1.06	9	2	1.01
10	2	1.03	10	1	1.02	10	4	1.02	10	2	1.00
11	2	1.01	11	3	1.01	12	1	1.00	11	1	1.00
12	1	1.00	12	9	1.00	13	1	1.00	19	1	1.00
13	2	1.00	13	4	1.00	14	2	1.00			
14	2	1.00	16	4	1.00	17	1	1.00			
18	1	1.00	18	2	1.00	28	1	1.00			
19	1	1.00	26	2	1.00	29	1	1.00			
21	1	1.00	34	1	1.00						
a = 1067.9968 b = 1.0489 R^2 = 0.9972			a = 1269.9869 b = 1.1032 R^2 = 0.9915			a = 1016.2943 b = 1.0912 R^2 = 0.9932			a = 1099.6546 b = 1.3357 R^2 = 0.9883		

Chapter 5			Chapter 6			Chapter 7			Chapter 8		
x	f_x	Exp	x	f_x	Exp	x	f_x	Exp	x	f_x	Exp
1	314	311.95	1	462	457.60	1	507	502.68	1	363	362.24
2	99	108.47	2	126	149.04	2	142	163.73	2	106	108.89
3	43	38.14	3	69	49.00	3	68	53.79	3	28	33.22
4	26	13.84	4	37	16.56	4	47	18.12	4	30	10.62
5	14	5.44	5	24	6.05	5	21	6.55	5	10	3.87
6	3	2.53	6	19	2.64	6	20	2.80	6	4	1.86
7	10	1.53	7	10	1.53	7	18	1.58	7	7	1.26
8	5	1.18	8	10	1.17	8	10	1.19	8	2	1.08
9	3	1.06	9	6	1.06	9	15	1.06	9	3	1.02
10	3	1.02	10	6	1.02	10	4	1.02	12	2	1.00
11	1	1.01	11	8	1.01	11	5	1.01	16	1	1.00
12	1	1.00	12	3	1.00	12	5	1.00	17	1	1.00
18	1	1.00	13	3	1.00	13	2	1.00	23	1	1.00
19	1	1.00	14	3	1.00	14	1	1.00			
			17	1	1.00	15	2	1.00			

			18	1	1.00	16	1	1.00			
			19	3	1.00	17	3	1.00			
			22	1	1.00	20	1	1.00			
			26	1	1.00	24	3	1.00			
			29	1	1.00	25	1	1.00			
						26	1	1.00			
						29	2	1.00			
a = 899.6844			a = 1408.2478			a = 1546.5781			a = 1209.5159		
b = 1.0624			b = 1.1263			b = 1.1258			b = 1.2084		
R^2 = 0.9953			R^2 = 0.9892			R^2 = 0.9895			R^2 = 0.9959		

Taking a more complex formula, it would, perhaps, be possible to increase the significance of the determination coefficient but we strive for simplicity and unification, not for the best of all fit that were tried.

Though we want to omit the counting of word frequencies, nevertheless we want to mention at least Frumkina's law (cf. Frumkina 1962). This is nothing else than stating the frequencies of some entities, but in passages of an exactly determined number of words, e.g. 1000. But even in that case one can differentiate between special semantic or grammatical functions of the given entity. Thus, in most cases either the negative hypergeometric or the negative binomial distributions are adequate as has been shown by Altmann and Burdinski (1982) but Brainerd (1972) used mostly the Poisson distribution. In some cases one can also use the exponential function, as can be shown below. Taking 200 passages from the German book *Deutschlandstunde* by S. Lenz, the authors found the distribution of the word „das" as presented in Table 13.2. Here passages are possible containing no occurrence of the given word. Frumkina's law shows the specific diversification of text passages.

Table 13.2
Numbers of „das" in a text by S. Lenz (Altmann, Burdinski 1982)

Occurrence	No of passages	Exp
0	95	97.64
1	58	51.33
2	28	27.22
3	11	14.65
4	3	8.11
5	2	4.70
6	1	2.93
7	1	2.00
8	1	1.52
a = 96.6374, b = 0.6523, R^2 = 0.9883		

Several sets of data from the same article can be shown but the functions holds true only if the data are decreasing. Frumkina's passages have the advantage of correct sampling, direct comparability of results and do not cause any difficulties because the computer programs can determine the passage length mechanically.

14. Music

Ranking can be performed also in music. One defines some entities, e.g. individual tones, motifs, beats, chords and counts their repetitions. It can be supposed that even individual composers, having their own style produce different samples of whatever. It can be conjectured that different kinds of music and different times of creation will display different rankings. In order to check whether the above model could be applied also here, we take the distribution of tones in Beethoven's Sonata Op. 27, No. 2 as presented by Wimmer and Altmann (2001) and Sonata No. 6 (Martináková et al. 2008) and perform the computation of parameters and the determination coefficient as shown in Table 14.1. The above authors applied the displaced negative hypergeometric distribution; here the exponential function is fitted.

Table 14.1
Frequencies of individual tones in Beethoven's
Sonata Op. 27, No. 2 and Sonata No. 6 (Martináková et al. 2008)

Frequencies	**a**	**b**	**R^2**
106,89,84,79,68,66,58,50,44,42,42,34,33,28, 25,21,20,19,18,17,17,16,14,14,14,14,13,11,11, 11,10,10,5,5,5,5,4,4,3,3,2,2,2,2,1,1,1,1,1,1,1,1,1	111.0136	0.0963	0.9935
404,316,303,302,281,278,275,265,247,227,214, 208,200,192,187,185,181,167,156,150,146,138, 129,127,122,113,110,94,89,87,85,83,79,78,77, 72,70,69,64,59,57,54,53,53,48,46,42,36,31,31, 29,15,13, 11,6,6,5,3,3	375.5736	0.0480	0.9853

The pitches in a composition by Palestrina have been examined in Popescu, Martináková-Rendeková, Altmann (2012). The results are presented in Table 14.2.

Table 14.2
Pitches in a composition by Palestrina
(Popescu, Martináková-Rendeková, Altmann 2012)

Composition	**Frequencies**	**a**	**b**	**R^2**
Pitches in Palestrina's *Missa Ascendo ad Patrem, 5. Sanctus*	70,69,64,56,45,42,34,33,30,26, 22,20,18,16,11,10,9,9,4,2,2,2,1	87.0726	0.1322	0.9806

As can be seen, the function holds true also for this kind of musical computation. It would be interesting to study in a musical piece the frequency of tones ac cording to their distance to the basic tone of the given piece. Here, no ranking but scaling would be involved. There are only 11 possible distances (counted in half tones),because afterwards the basic tone comes again. Music is full of problems but to be a musicologist means mostly to exclude mathematics. Not only the development of individual composers but also the development of the music as a whole in one nation could be captured in this way.

15. Conclusions

As can be seen, the above simple model is sufficient for capturing a great number of data from various linguistic disciplines, even from music. Automatically the question arises whether one can simplify some procedures also in other sciences. To this end it would be necessary to process all data ever written about the Zipf's law, about classification, etc. The task is superhuman and can be fulfilled only with a team of collaborators. It is sure that in the course of such a work, ever new subsidiary conditions would appear and the simple model proposed above would become more and more complex. But it would, perhaps, be possible to discover the source of deviations and perform a classification of texts and other phenomena based on the character of these deviations.

In general, if there is already a theory from which something can be derived but the subsidiary conditions are not yet known, one should always begin with the simplest approach and take the more complex ones only if the applied one does not yield a satisfactory result. Using a function, one can rely on the determination coefficient or its „improvements", but the value it should reach is a personal decision. In social sciences one usually sets $R^2 > 0.80$ but this is nothing that is based on data, it is a trace of subjectivity in our domain. Further, it plays no role whether one uses a probability distribution or a simple function and the terms „continuous" and „discrete" are merely our views of reality. We should forget that we attain some kind of truth, we merely adapt the reality to our possibilities.

The author hopes that many data will be evaluated anew whereby the applied models will be simplified. This does not concern merely diversifications but any phenomena appearing in texts.

References

Altmann, G. (1991). Word class diversification of Arabic verbal roots. In: Rothe, U. (ed.), *Diversification Processes in Language: Grammar: 57-59.* Hagen: Rottmann.

Altmann, G. (1992). Two models of word association data. Rieger, B.B.(ed.), *Glottometrika 13: 105-120*. Brockmeyer: Bochum

Altmann, G. (2005). Diversification processes. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.), *Quantitative Linguistics. An International Handbook: 646-658*. Berlin: de Gruyter.

Altmann, G. (2014). On morphological complexity in Indonesian. *Glottometrics 29, 59-69.*

Altmann, G., Best, K.-H., Kind, B. (1987). Eine Verallgemeinerung des Gesetzes der semantischen Diversifikation. In: Fickermann, I. (ed.), *Glottometrika 9*: *130-139.* Bochum: Brockmeyer.

Altmann, G., Burdinski, V. (1982). Towards a law of word repetitions in text-blocks. In: W. Lehfeldt, U. Strauss (eds.), *Glottometrika 4: 147-167.* Bochum: Brockmeyer.

Altmann, G., Kind, B. (1983). Ein semantisches Gesetz. In: Köhler, R., Boy, J. (eds.), *Glottometrika 5: 1-7*. Bochum: Brockmeyer.

Altmann, G., Köhler, R. (2015). *Forms and Degrees of Repetitions in Texts*. Berlin/Boston: de Gruyter.

Andreev, S., Popescu, I.-I., Altmann, G. (2016). On Russian adnominals. *Glottometrics 35, 64-84.*

Andreev, S., Popescu, I.-I., l Altmann, G. (2017). Some properties of denominals in Russian texts. *Glottometrics 38, 2017, 77-106*

Baayen, H.R., Piepenbrock, R., Rijn, H. van (1993). *The CELEX lexical database (CD-ROM).* Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bagheri, D. (2002). Definitionsfolgen und Lexemnetze. In: Altmann, G., Bagheri, D., Goebl,, H., Köhler, R., Prün, C. (eds.), *Einführung in die quantitative Lexikologie: 94-133.* Göttingen: Peust & Gutschmidt.

Becker, H. (1995). *Die Wirtschaft in der deutschsprachigen Presse*. Frankfurt: Lang.

Belza, M.I. (1971). K voprosu o nekotorych osobennostjach semantičeskoj struktury svjaznych textov. In: *Semantičeskie problemy avtomatizacii i informacionnogo potoka: 58-73*. Kiev.

Beöthy, E., Altmann, G. (1984a). Semantic diversification of Hungarian verbal prefixes III. “föl-“, “el-“, “be-“. In: Rothe, U. (ed.), *Glottometrika 7: 45-56.* Bochum: Brockmeyer.

Beöthy, E., Altmann, G. (1984b). Semantic diversification of Hungarian verbal prefixes II. ki-. *Finnisch-ugrische Mitteilungen 8, 29-37.*

Beöthy, E., Altmann, G. (1991). The diversification of meaning of Hungarian verbal prefixes I. “meg-“. In: Rothe, U**.** (ed.) (1991). *Diversification Processes in Language: Grammar: 60-66*. Hagen: Rottmann.

Best, K.-H. (1990). Die semantische Diversifikation eines Wortbildungsmusters im Frühneuhochdeutschen. In: Hřebíček, L. (ed.), *Glottometrika 11: 107-110*. Bochum: Brockmeyer.

Best, K.-H. (1991). *Von*: Zur Diversifikation einer Partikel des Deutschen. In: Rothe, U. (ed.) (1991). *Diversification Processes in Language: Grammar: 94-104*. Hagen: Rottmann.

Best, K.-H. (1994). Word class frequencies in contemporary German short prose texts. *Journal of Quantitative Linguistics 1(2),144-147.*

Best, K.-H. (1997). Zur Wortartenhäufigkeit in Texten deutscher Kurzprosa der Gegenwart. In: Best, K.-H. (ed.), *Glottometrika 16, 276-285.* Trier: Wissenschaftlicher Verlag.

Best, K.-H. (2000). Verteilung der Wortarten in Anzeigen. *Göttinger Beiträge zur Sprachwissenschaft 4, 37-51.*

Best, K.-H. (2001). Probability distributions of language entities. *Journal of Quantitative Linguistics 8(1) 1-11.*

Best, K.-H. (2001a). Zur Gesetzmäßigkeit der Wortartenverteilungen in deutschen Pressetexten. *Glottometrics 1, 1-26.*

Best, K.-H. (2006). *Quantitative Linguistik. Eine Annäherung. Göttingen:* Peust & Gutschmidt.

Best, K.-H. (2007). Diversifikation bei Eigennmen. In: P. Grzybek, R. Köhler (eds.), *Exact methods in the study of language and text: 21-31.*Berlin/New York: Mouton de Gruyter.

Best, K.-H. (2008a). Zur Diversifikation deutscher Hexameter. *Naukovyj Vysnik Černivec'koho Universitetu. Herman'ska filolohia. Vypusk 41, 172-180.*

Best, K.-H. (2008b). Zur Diversifikation lateinischer und griechischer Hexameter. *Glottometrics 17, 45-53.*

Best, K.-H. (2008c). Das Fremdwortspektrum im Türkischen. *Glottometrics 17, 8-11.*

Best, K.-H. (2010). Zum Fremdwortspektrum im Japanischen. *Glottotheory 3(1), 5-8.*

Best, K-.H. (2012). Diversifikation der starken Verben im Deutschen. *Glottometrics 24, 1-4.*

Best, K.-H., Altmann, G. (2015). On the frequency of simple attributes in German. *Mathematical Linguistics 1(1), 1-8.*

Bichsel, P. (1996). *Stockwerke*. Stuttgart: Reclam.

Bobrowski, J. (1987). *Lipmanns Leib*. Stuttgart: Reclam.

Böll, H. (1958). *Und sagte kein einziges Wort. Erzählungen*. Moskau. Izd. Inostarnnoj literatury.

Brainerd, B. (1972). Article use as an indicator of style among English-language authors. In: Jäger, S. (ed.), *Linguistik und Statistik: 11-32.* Braunschweig: Vieweg.

Brüers, N., **Heeren, A.** (2004). Pluralallomorphe in Briefen Heinrich von Kleists. *Glottometrics 7, 85-90.*

Bruyn, G. de (1999). *Babylon*. Frankfurt: Fischer Taschenbuch Verlag.

Busse, U. (1999). "Prithee now, say you will, and go about it." *Prithee vs.pray you as Discourse Markers in the Shakespeare Corpus.* In: Neumann, F.-W., Schülting, S. (eds.), *Anglistentag 1998 Erfurt. Proceedings: 485-500.* Trier: Wissenschaftlicher Verlag.

Chen, R., Altmann, G. (2015). Conceptual inertia in texts. *Glottometrics 30, 73-88.*

Drobisch, M.V. (1866). Ein statistischer Versuch über die Formen des lateinischen Hexameters. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig. Philologisch-historische Classe 18, 73-139.*

Drobisch, M.V. (1868a). Weitere Untersuchungen über die Formen des Hexameters der Vergil, Horaz und Homer. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig. Philologisch-historische Classe 20, 16-53.*

Drobisch, M.V. (1868b). Über die Formen des deutschen Hexameters bei Klopstock, Voss und Goethe. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Lepzig. Philologisch-historische Klasse 20, 138-160.*

Drobisch, M.W. (1872). Statistische Untersuchungen des Distichon (von Hrn. Dr. Hultgren). *Berichte über die Verhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Philologisch-Historische Classe 24, 1-33.*

Drobisch, M.W. (1875). Ueber die Gesetzmässigkeit in Goethe's und Schiller's Distichen. In: *Königlich-Sächsische Gesellschaft der Wissenschaften zu Leipzig, Philologisch-Historische Classe, Berichte über die Verhandlungen 27, 8-34-146.*

Dshurjuk, T.V., Levickij, V.V. (2003). Satztypen und Satzlängen im Funktional- und Autorenstil. *Glottometrics 6, 2003, 40-51.*

Echols, J.E., Shadily, H. (1963). *An Indonesian-English Dictionary.* Ithaca: Cornell University Press.

Engel, U. (1988). *Deutsche Grammatik*. Heidelberg: Groos.

Engwall, G. (1984). *Vocabulaire du roman francais (1962-1968). Dictionnaire des fréquence.* Stockholm: Almquist & Wiksell International.

Fan, F., Altmann, G. (2007). Measuring the of compounds. In: Kaliusčenko, V., Köhler, R., Levickij, V. (eds.), *Problemy tipologičnoï ta kvantitativnoï leksikologiï. Problems of Typological and Quantitative Lexicology: 190-209.* Černivcy: Ruta.

Fan, F., Altmann, G. (2008). On meaning diversification in English. *Glottometrics 17,66-78.*

Finkenstaedt,Th., Wolff, D. (1973). *Ordered Profusion. Studies in Dictionaries and the English Lexicon* with Contributions by H. Joachim Neuhaus and Winfried Herget. Heidelberg: Winter.

Frumkina, R-M. (1962). O zakonach raspredelenija slov i klassov slov. In: Mološnaja,T.N. (ed.), *Strukturno-toplogičeskie issledovanija: 123-133.* Moskva: ANSSSR.

Fuchs, R. (1991). Semantische Diversifikation der deutschen Präposition *auf*. Rothe, U. (ed.): *Diversification Processes in Language: Grammar: 105-115*. Hagen: Rottmann.

Ginka, B.I. (1983). Leksiko-semantičeskaja gruppa suščestvitel'nych so značeniem "trud, rabota" v sovremennom nemeckom jakyze. Odessa: Diss.

Gnatchuk, H. (2015a). A statistical analysis of English compounds in the newspaper style. *Mathematical Linguistics 1(1), 81-90.*

Gnatchuk, H. (2015b). A quantitative analysis of English compounds in scientific texts. *Glottometrics 33, 1-7.*

Gnatchuk, H. (2015c). A quantitative investigation of English compounds in prose texts. *Glottometrics 32, 1-8.*

Grzybek, P. (2005). Zur Vorgeschichte quantitative Ansätze in der russischen Sprach- und Literaturwissenschft. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.), *Quantitative Linguistics. An International Handbook: 23-64.* Berlin: de Gruyter.

Habermann, M., Müller, P.O. (1987). Zur Wortbildung bei Albrecht Dürer. *Zeitschrift für deutsche Philologie 106, Sonderheft: Frühneuhochdeutsch 117.137.*

Hammerl, R. (1987). Untersuchungen zur mathematischen Beschreibung des Martingesetzes der Abstarktionsebenen. In: Fickermann, I. (ed.), *Glottometrika 8, 113-129*. Bochum: Brockmeyer.

Hammerl, R. (1989). Untersuchung struktureller Eigenschaften von Begriffsnetzen. In: Hammerl, R. (ed.), *Glottometrika 10: 141-154.* Bochum: Brockmeyer.

Hammerl, R., Sambor, J. (1991). Untersuchungen zur Verteilung der Bedeutungen der polyfunktionalen polnischen Präposition *w* in Text. In: Rothe, U. (ed.): *Diversification Processes in Language: Grammar: 127-137*. Hagen: Rottmann.

Hammerl, R., Sambor, J. (1993). *O statystycznach prawach językowych.* Warszawa: Polskie Towarzystwo Semiotyczne.

Handwörterbuch der deutschen Gegenwartssprache. In zwei Bänden. (1984). Berlin: Akademie Verlag.

Helbig, G., Schenkel, W. (^{8}1991). *Wörterbuch zur Valenz und Distribution deutscher Verben*. 8., durchgesehene Auflage. Tübingen: Max Niemeyer Verlag.

Hennern, A. (1991). Zur semantischen Diversifikation von *in* im Englischen. In: Rothe, U. (ed.): *Diversification Processes in Language: Grammar: 116-126*. Hagen: Rottmann.

Hirsch-Wierzbicka, L. (1971). *Funktionale Belastung und Phonemkombination am Beispiel einsilbiger Wörter der deutschen Gegenwartssprache.* Hamburg: Buske.

Hřebíček, L. (1990). Menzerath-Altmann's law on the semantic level. In: Hřebíček, L. (ed.), *Glottometrika 11: 37-56.* Bochum: Brockmeyer.

Hřebíček, L. (1992). *Text in Communication: Supra-Sentence Structures.* Bochum: Brockmeyer.

Hřebíček, L. (1995). *Text levels. Language constructs, constituents and the Menzerath-Altmann law*. Trier: Wissenschaftlicher Verlag.

Hřebíček, L. (1997). *Lectures on Text Theory*. Prague: Oriental Institute.

Hug, M. (2000).. Partial disambiguation of very ambiguous grammatical words. *Lecture August 25, 2000, QUALICO IV. Prague.*

Hug, M. (2001). Partial disambiguation of very ambiguous grammatical words. *Journal of Quantitative Linguistics 7, 217-226.*

Ishutin, D., Gnatchuk, H. (2017). Ukrainian compounds in the texts of computer science. *Glottometrics 39, 87-91.*

Job, U. (1981). Annotated bibliography on the statistical study of hexameter verse. In: Grotjahn, R. (ed.), *Hexameter Studies: 226.262.* Bochum: Brockmeyer.

Juilland, A., **Brodin, D., Davidovitch, C.** (1971). *Frequency Dictionary of French Words*. The Hague:Mouton.

Kaliuščenko, V.D. (1988). *Deutsche denominale Verben*. Tübingen: Narr.

Kant, H. (1987). *Bronzezeit.* Berlin: Aufbau-Verlag.

Kant, H. (1989). *Schöne Elise.* Leipzig.

Kempgen, S. (1995). *Russische Sprachstatistik*. München: Sagner

Kisro-Völker, S. (1984). On the measurement of abstractness in lexicon. In: Boy, J., Köhler, R. (eds.), *Glottometrika 6, 139-151.* Bochum: Brockmeyer.

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

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

Köhler, R. (2005a). Quantitative Untersuchungen zur Valenz deutscher Verben. *Glottometrics 9, 2005, 13-20*

Köhler, R. (2012). *Quantitative syntax analysis*. Berlin/Boston: de Gruyter.

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

Köhler, R., Naumann, S. (2007). Quantitative analysis of co-reference structures in texts. In: Grzybek, P., Köhler, R. (eds.), *Exact Methods in the Study of Language and Text: 317-329.* Berlin/New York: Mouton de Gruyter.

Köpcke, K.-M. (1982). *Untersuchungen zum Genussystem der deutschen Gegenwartssprache*. Tübingen: Niemeyer.

Körner, H. (2004). Zur Entwicklung des deutsches (Lehn-)Wortschatzes. *Glottometrics 7, 25-49.*

Krylov, Ju.K. (1982). Ob odnoj paradigme lingvističeskich raspredelenij. *Tartu Riikliku Ülikooli Toimetised 628, 80–97.*

Kubát, M. (2016). *Kvantitativní analýza žánrů.* Ostrava: Ostravská univerzita.

Kuby, E. (19897). *Der Spiegel im Spiegel*. München: Heyne.

Kunert, G. (1990). *Der Hai.* Stuttgart: Reclam 1990

Kunze, K. (2004). Perspektiven neuer Familiennamen-Geographie. *Sprachreport 20(4), 21-27.*

Kurcz, I., Levicki, A., Sambor, J., Woronczak, J. (1976). *Słownictwo współczesnego języka polskiego. Listy frekwencyjne, Vol. VI.* Warszawa.

Kuße, H. (1991). *A* und *no* in N.M. Karamzins Pis´ma Russkogo Putešestvennika. In: Rothe, U. (ed.), *Diversification Processes in Language: Grammar: 173-182*. Hagen: Rottmann.

Laufer, J., Nemcová, E. (2009). Diversifikation deutscher morphologischer Klassen in SMS. *Glottometrics 18, 13-25.*

Levickij, V.V. (1989*). Statističeskoe izučenie leksemičeskoj semantiki.* Kiev: UMK BO.

Levickij, V.V., Drebet, V.V., Kiiko, S,V. (1999). Some quantitative characteristics of polysemy of verbs, nound and adjectives in the German language. *Journal of Quantitative Linguistics 6(2), 172-187.*

Levickij, V., Hikov, L. (2004). Gebrauch der Wortarten im Autorenstil. *Glottometrics 8, 12-22.*

Levickij, V., **Lučak, M.** (2005). Category of tense and verb semantics in the English language. *Journal of Quantitative Linguistics 12(2-3), 212-238*.

Levickij, V.V., Pavlyčko, O.O., Semenyuk, T.G. (2001). Sentence length and sentence structure as statistical characteristics of style in prose. In: Uhlířová, L. et al. (eds.), *Text as a linguistic paradigm: levels, constituents, constructs: 177-186*. Trier: Wissenschaftlicher Verlag.

Levickij, V.V., Romanova, T.A. (1997). Use of tenses of verbs and adverbs in the English language: a statistical study. *J. of Quantitative Linguistics 4, 135-138.*

Levin, B. (1998). English verb classes and alternations. Chicago: University of Chicago Press.

Liu, H. (2009). Probability distributions of dependencies based on a Chinese dependency Treebank. *Journal of Quantitative Linguistics 16(3), 256-273*.

Liu, H. (2012). Probability distribution of semantic roles in a Chinese treebank annotated with semantic roles. In: Naumann, S., Grzybek, P., Vulanović, R. Altmann, G. (eds.) (2012). *Synergetic Linguistics. Text and Language as Dynamic Systems: 101-107*. Vienna: Praesens.

Lučak, M. (2003). Selektivnist` spolučuvannosti leksičeskich i gramatičnich značen` u funkcionalnych sliljach sučasnoï anglijskoï movi. In: *Naukovij Visnik Černiveckogo Universitetu, 165/166, 297-307.* Černivci: Ruta.

Mackuliak, J. (2007). Vživanija nimec'kich skladnich imennikiv u riznich funkcional`nych stiliach. In: Kaliuscenko, V., Köhler, R., Levickij, V. (eds.), *Problemy tipologičnoï ta kvantitativnoï leksikologiï. Problems of Typological and Quantitative Lexicology: 210-216.* Černivcy: Ruta.

Mann, Th. (1959). *Buddenbrooks: Verfall einer Familie*. Moscow: Izd. Inostrannoj literatury.

Mann, Th. (1973*). Lotte in Weimar*. Berlin und Weimar.

Martin, R. (1974). Syntax de la définition lexicographique: étude quantitative des définissants dans le „Dictionnaire fondamental de la langue français“, In: David, J., Martin, R. (eds.), *Statistique et linguistique: 61-71*. Paris: Klincksieck.

Martináková, Z., Mačutek, J., Popescu, I.-I., Altmann, G. (2008). Some problems of musical texts . *Glottometrics 16, 2008, 80-110.*

Meuser,K., Schütte, J.M., Stremme, S. (2008). Pluralallomorphe in den Kurzgeschichten von Wolfdietrich Schnurre. *Glottometrics 17, 12-17.*

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

Nagórko-Kufel, A. (1984). Die Anwendung des Häufigkeitskriteriums bei der Wortbildung. In: Boy, J., Köhler, R. (eds.), *Glottometrika 6, 48-64.* Bochum: Brockmeyer.

Nemcová, E. (1991). Semantic diversification of Slovak verbal prefixes. In: Rothe, U. (ed.) *Diversification Processes in Language: Grammar: 67-74*. Hagen: Rottmann

Nemcová, E. (2007). Zur Diversifikation des Bedeutungsfeldes slowakischer verbaler Präfixe. In: Grzybek, P., Köhler, R. (eds.), *Exact methods in the study of language and,text:499-508.* Berlin/New York: Mouton de Gruyter.

Nemcová, E. (2009). Nominal suffixes in German press texts**.** *Glottometrics 18, 70-77.*

Orlov, J.K., Boroda, M.G., Nadarejšvili, I.Š. (1982). *Sprache, Text, Kunst. Quantitastive Analysen.* Bochum: Brockmeyer

Overbeck, A., Tuzzi, A., Popescu, I.-I., Altmann, G. (2010). Analysis of Italian word classes. *Glottometrics 20, 12-28.*

Özlü, D. (1990). *Bir Yaz Mevsimi Roamsì.* Istambul: Ada.

Palermo, D.D., Jenkins, J.J. (1964). *Word association norms. Grade school through college.* Minneapolis: Universiyt of Minnesota Press.

Pan, X., Liu, H. (2014). Adnominal constructions in Modern Chinese and their distribution properties. *Glottometrics 29, 1-30*.

Pawlowski, A. (1999). The quantitative approach in cultural anthropology: Application of linguistic corpora in the analysis of basic colour terms. *Journal of Quantitative Linguistics 6(3), 222-234***.**

Pelegrinová,K., Altmann, G. (2017). The study of adverbials in Czech. *Glottometrics 17, 34-53*

Pestalozzi, J.H. (1992). *Fabeln*. Zürich: Manesse.

Popescu, I.-I., Altmann, G. (2008). On the regularity of diversification in language. *Glottometrics 17, 94-109.*

Popescu, I.-I., Altmann, G., Köhler, R. (2009). Zipf's law – another view. *Quality and Quantity 44, 713-731* (doi: 10.1007/s11135-009-9234-y).

Popescu, I.-I., Best, K.-H., Altmann, G. (2014). *Unified modeling of length in language.* Lüdenscheidei: RAM-Verlag.

Popescu, I.-I., Kelih, E., Best, K.-H., Altmann,G. (2009). Diversification of the case. *Glottometrics 18, 32-39.*

Popescu, I.-I., Martináková-Rendeková, Z., Altmann, G. (2012). Stratification in musical texts based on rank-frequency distribution of tone pitches. *Glottometrics 24, 25-40.*

Prün, C. (1995). *Die linguistischen Hypothesen von G.K. Zipf aus systemtheoretischer Sicht.* Magisterarbeit: Universität Trier

Prün, C., **Steiner, P.** (2005). Quantitative Morphologie: Eigenschaften der morphologischen Einheiten und Systeme. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.), *Quantitative Linguistics. An International Handbook: 227-242.*. Berlin/New York: de Gruyter.

Rather, A., **Rothe, U.** (1991). Diversifikation der deutschen Komposita: \`Substantiv plus Substantiv\` In: Rothe, U. (ed.): *Diversification Processes in Language: Grammar: 85-91*. Hagen: Rottmann.

Remarque, E.-M. (1975). *Drei Kameraden*. Moskau: Izd. inostrannoj literatury.

Remarque, E.-M. (1981). *Im Westen nichts Neues*. Moscow: Izd. inostrannoj literatury.

Rensinghoff, S. (1988). *Unteruchungen zu Wortlänge und Polysemie am Beispiel des Französischen.* Trier: Universität (Seminararbeit)

Roelcke, Th., Altmann, G. (2015). Measuring morphological complexity of the word. *Glottotheory 6(1), 93-111*.

Roos, U. (1991). Zur Diversifikation der japanische Postposition *ni*. In: Rothe, U. (ed.) (1991). *Diversification Processes in Language: Grammar: 75-82*. Hagen: Rottmann.

Rothe, U. (1986). *Die Semantik des textuellen* et. Frankfurt: Lang.,

Rothe, U. (1990). Semantische Motivation der Genuszuweiseung. In: Hřebíček, L. (ed.), *Glottometrika 11, 95-106*. Bochum: Brockmeyer.

Rothe, U. (1990a). Verteilung der Suffixe denominaler Verben nach ihrem semantischen Wortbildungsmustern. In. Hammerl, R. (ed.*), Glottometrika 12: 107-114.* Bochum: Brockmeyer.

Rothe, U. (ed.) (1991). *Diversification Processes in Language: Grammar*. Hagen: Rottmann.

Rothe, U. (1991a). The diversification of the case: genitive. In: Rothe, U. (ed.) (1991). *Diversification Processes in Language: Grammar: 140-156*. Hagen: Rottmann.

Rothe, U. (1991b). Diversification processes in grammar: An introduction. In: Rothe, U. (ed.) (1991). *Diversification Processes in Language: Grammar: 3-32*. Hagen: Rottmann.

Sambor, J. (1989). Polnische Version des Projekts "Sprachliche Synergetik. Teil I. Quantitative Lexikologie. *Glottometrika 10, 171-197.*

Sambor, J. (2005). Lexical networks. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.), *Quantitative Linguistics - An Inernational handbook: 447-458.*. Berlin: de Gruyter.

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

Sanada, H. (2007). Synergetic linguistics and Japanese lexicology. In: Kaliusčenko, V., Köhler, R., Levickij, V. (eds.), *Problemy tipologičnoï ta kvantitativnoï leksikologiï. Problems of Typological and Quantitative Lexicology: 61-82*. Černivcy: Ruta.

Sanada,H., Altmann, G. (2009). Diversification of postpositions in Japanese. *Glottometrics 19, 70-79.*

Schierholz, S. (1989). Kritische Aspekte zum Martinschen Gesetz. In: Hammerl, R. (ed.), *Glottometrika 10,: 108-128*. Bochum: Brockmeyer.

Schierholz, S. (1991). *Lexikologische Analysen zur Abstraktheit, Häufigkeit und Polysemie detscher Substantive*. Tübingen: Niemeyer.

Schindelin, C. (2005). Die quantitative Erforschung der chinesischen Sprache und Schrift. In: Köhler, R., Altmann, G., Piotrowski, R.G. (eds.), *Quantitative Linguistics - An Inernational Handbook: 947-970*. Berlin: de Gruyter.

Schnurre, W. (1997). *Als Vater sich den Bart abnahm. Erzählungen.* München: Piper Verlag.

Schwartz, C. (1995). The distribution of aggregates in text. *ZET - Zeitschft für Empirische Textforschung 2, 62-66.*

Schweers, A., **Zhu, J**. (1991). Wortartenklassifizierung im Lateinischen, Deutschen und Chinesischen. In: Rothe, U. (ed.) (1991). *Diversification Processes in Language: Grammar: 157-165*. Hagen: Rottmann.

Seibicke, W. (1982). *Die Personennamen im Deutschen*. Berlin/New York: de Gruyter.

Shmidt, E., Gnatchuk, H. (2016). German compounds in the texts of technical science. *Glottometrics 35, 1-5.*

Silnickij, V.V., Drebet, V.V., Kiiko, S.V. (1999). Some quantitative characteristics of polysemy of verbs, nouns and adjectives in the German language. *Journal of Quantitative Linguistics 6(2), 172-187*

Skalmowski, W. (1964) A note on the distribution of Arabic verbal roots. *Folia Orientalia 6, 97-100.*

Skorochod´ko, E.F. (1981). *Semantische Relationen in der Lexik und in Texten*. Bochum: Brockmeyer.

Sowinski, B. (1979). *Werbeanzeigen und Werbesendungen.* Stuttgart: Metzler.

Steiner, P., **Prün C.** (2007).The effects of diversification and unification on the inflectional paradigm of German nouns. In: Grzybek, P., Köhler, R. (eds.), *Exact methods in the study of language and text: 623-632*. Berlin/New York: Mouton de Gruyter.

Strauss, U., Sappok, Ch., Diller, H.J., Altmann, G. (1984). Zur Theorie der Klumpung von Textentitäten. In: Rothe, U. (ed.), *Glottometrika 7, 73-100.* Bochum: Brockmeyer.

Tomaševskij, B.V. (1929). *O stiche*. Leningrad (München: Fink 1970).

Tuzzi, A., Popescu, I.-I., Altmann, G. (2010). *Quantitative Analysis of Italian Texts.* Lüdenscheid: RAM-Verlag.

Uhlířová, L. (2002). The case of Czech possessive adjectives and their head nouns: some distributional properties. *Glottometrics 2, 1-10.*

Uhlířová, L. (2007). Using Altmann-fitter for text analysis: An example from Czech. In: Grzybek, P., Köhler, R. (eds.), *Exact methods in the study of language and text: 659-664.* Berlin/New York: Mouton de Gruyter.

Väyrynen, P.A., Noponen, K., Seppänen, T. (2008). Preliminaries to Finnish word prediction. *Glottotheory 1, 65-73.*

Wang, L. (2016**).** Parts-of-speech studies in Chinese. *Journal of Quantitative Linguistics 23(3), 235-255.*

Wimmer, G., **Altmann, G.** (1995). A model of morphological productivity. *Journal of Quantitative Linguistics 2(3), 212-216.*

Wimmer, G., **Altmann, G.** (1999). Verteilung der Polysemie in Maori. In: Genzor, J., Ondrejovič, S. (eds.), *Pange Lingua: 17-25.* Bratislava: Veda.

Wimmer, G., **Altmann, G.** (2001). Models of rank-frequency distributions in language and music. In: Uhlířová, L., Wimmer, G., Altmann, G., Köhler, R. (eds.), *Text as a linguistic paradigm: Levels, constituents, constructs: 283-294.*Trier: WVT.

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

Wolff, D. (1969). *Statistische Untersuchungen zum Wortschatz englisher Zeitungen.* Diss. Saarbrücken,.

Yesypenko, N. (2009). An integral qualitative-quantitative approach to the study of concept realization in the text. In: Kelih, E., Levickij, V., Altmann, G. (eds.), *Methods of Text Analysis: 308-327*. Chernivcy: CNU.

Yesypenko, N. (2008). Writer's voice in the texts of „Peace and War" themes. *Glottometrics 16, 17-26.*

Yin, B. (1986). Hanyu cilei de dingliang yanjiu. *Zhongguo yuwen 6, 428-436.*

Yue, M., **Liu, H.** (2011). Probability distribution of discourse relations based on a Chinese RST-annotated corpus. *Journal of Quantitative Linguistics 18 (2), 107-121.*

Zhu, J., Best, K.-H. (1992). Zum Wort im modernen Chinesisch. *Oriens Extremus 35, 45-60.*

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., Wimmer, G., Altmann, G., Köhler, R. (eds.), *Text as a Linguistic Paradigm: Levels, Constituents, Constructs. Festschrift in Honour of Ludek Hřebíček: 295-312*. Trier: WVT.

Index

The RAM-Verlag Publishing House edits since 2001 also the journal ***Glottometrics* –** up to now 39 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 (39, 2017) is as follows:

Herausgeber – Editors of Glottometrics