Glottometrics 31
2015

RAM-Verlag

ISSN 2625-8226

Glottometrics

Glottometrics ist eine unregelmäßig erscheinende Zeitdchrift (2-3 Ausgaben pro Jahr) für die quantitative Erforschung von Sprache und Text.
Beiträge in Deutsch oder Englisch sollten an einen der Herausgeber in einem gängigen Textverarbeitungssystem (vorrangig WORD) geschickt werden.
Glottometrics kann aus dem **Internet** heruntergeladen werden (**Open Access**), auf **CD-ROM** (PDF-Format) oder als **Druckversion** bestellt werden.

Glottometrics is a scientific journal for the quantitative research on language and text published at irregular intervals (2-3 times a year).
Contributions in English or German written with a common text processing system (preferably WORD) should be sent to one of the editors.
Glottometrics can be downloaded from the **Internet (Open Access)**, obtained on **CD-ROM** (as PDF-file) or in form of **printed copies.**

Herausgeber – Editors

Bestellungen der CD-ROM oder der gedruckten Form sind zu richten an

Orders for CD-ROM or printed copies to RAM-Verlag RAM-Verlag@t-online.de

Herunterladen/ Downloading: https://www.ram-verlag.eu/journals-e-journals/glottometrics/

Die Deutsche Bibliothek – CIP-Einheitsaufnahme
Glottometrics. 31 (2015), Lüdenscheid: RAM-Verlag, 2015. Erscheint unregelmäßig.
Diese elektronische Ressource ist im Internet (Open Access) unter der Adresse
https://www.ram-verlag.eu/journals-e-journals/glottometrics/ verfügbar.
Bibliographische Deskription nach 31 (2015) **ISSN 2625-8226**

Contents Glottometrics 31

Glottometruícs 31, 2015, 1-30

Sound symbolism: Myths and reality

Hanna Gnatchuk

Abstract. The given article is devoted to V.V. Levitskij's book "Sound symbolism: myths and reality" (2009). In this book the linguist 1) distinguishes the most topical problems in this area which require further investigations; 2) he gives a thorough look at the theoretical studies of sound symbolism in the USA, Canada, Europe and the countries of the previous Soviet Union; 3) he represents the results of the experiments conducted at Chernivtsi National Juri Fedjkovich university and compares them with the outcomes by the other researchers; 4) he pays a careful attention to the methodological demands in his experiments. The researcher takes the view that the *authentic* and objective results can only be obtained when one follows correct methodological principles. This survey is a kind of homage to our teacher.

Keywords: sound (phonetic) symbolism, iconicity, phonosemantics, subjective and objective sound symbolism, motivation.

1. What is symbol? Sign and its types

Before studying such a linguistic phenomenon as sound symbolism, V. Levitskij gives some information about signs. Sign is regarded in semiotics as a material object which points to the subject of the outer reality. On the whole, the sign has two *features*: a) it is material; b) it points to something. By way of illustration, the author gives the example of "smoke". On the one hand, "smoke" may be a sign of wood fire; on the other hand, it may signal that enemies are coming. In such a way, these features are both perceived by people and point to the subject of outer reality (regardless themselves). Nevertheless, they are connected with the reality in a different way:

1. The connection that has cause-effect *character* ("smoke" is a result of wood fire);
2. The connection that has conventional *character*.

The signs of the first type are called "natural" (Schaff, 1963:183), the others – artificial. Levitskij states that the *character* (the connection of sign with the outer reality) plays more significant role than the features ("natural" feature – created by nature; "artificial" – created by people). Here the researcher gives the following example: if it is sunny in the morning, then the participants of the excursion may decide to gather at the bus-station at that time. Under the other conditions, the excursion will not take place. In this case, "sunny or cloudy" morning will take no part in the human mind. It will be a conventional signal, not a symptom (as a smoke of fire, cough when one is cold). Therefore, Levitskij considers that it will be more correct to make the first division of signs according to "symptom-nonsymptom" (by analogy with Schaff), The researcher declares that signs-symptoms are connected with our environment by cause-effect relations.

The other signs are divided by him into three basic groups: indexes, icons and symbols. Indexes are connected with the environment by conventional relations (all linguistic signs are indexes); icons (photos, pictures, sculpture, etc) are copies or reflections of the

reality. Special attention should be paid here to the emblems or the color of flags. In this case, the author believes that symbol takes an intermediate place between indexes and icons. Levitskij draws a parallel with E. Lerch (1939) who noticed that symbol keeps only relation accuracy ("relationstreu"), not material one. In such a way, the distinctive feature of symbol is its connection not with material resemblance but with structural.

Finally, the author summarizes the above-mentioned opinions:

All signs have the following common properties: 1) they are material (perceived by people); 2) they point to any subject of the outer reality (they have "pointing" function); 3) they point to the subject regardless themselves (they have "substituting" function); 4) the meanings of the signs are ascribed by certain social groups. The most important property which makes a distinction between the signs is the type of the relations of signs with the outer reality: cause-effect connection, material resemblance, structural resemblance or the absence of resemblance. In accordance with these, all signs are divided into symptoms, icons, symbols and indexes:

SIGNS

↓	↓	↓	↓
Symptoms	**Icons**	**Symbols**	**Indexes**
↓	↓	↓	↓
Cause-effect relation	Material resemblance	Structural resemblance	the absence of resemblance

Supporting Pierce's theory, Wescott (1980) divides all signs into two categories – signals and indexes. Wescott interprets symbols as arbitrary signs whereas icons – the signs which are in the accordance with the designated subjects. Levitskij suggests comparing his scheme with R. Wescott's one.

The author considers that sound symbolism is based on the structural resemblance between a sound and a sense (not material). Therefore, it would be incorrect to identify the terms iconicity and symbolism (the term "iconicity" has been widely used in the USA recently). The author considers that the term "iconicity" can be interpreted too wide: it combines both iconicity and sound symbolism. Or it can be regarded too narrow: only as onomatopoeia in so far as iconic signs are connected with subjects by material resemblance. It is clear that there cannot be a full material resemblance between the iconic language sign (sound imitative word) and the subject which is signified by this sign (sound imitative words of different languages which mean identical subjects of the outer reality). In this situation, it remains unclear for Levitskij why the words "flit", "float", "fly" refer in the work by Magnus (2001) to the examples of the real iconicity. The author asks who has proved that *fl-* indicates a hesitating movement. For example, L. Bloomfield (1968) connects *fl-* with the idea of a quick and a light movement. Nevertheless, there is a variety of examples with *fl-* in his book in which the combination of these sounds has quite different meanings. Here Levitskij refers the above-mentioned examples to the symbolic types of signs (not iconic ones).

The objection to the existence of sound symbolism was usually formulated in the following way: if there were non-arbitrary connection between certain signs and notions, then the existence of various names of the same subject would be impossible as well as the change of the word's sounds in the process of historical development. In particular, the notion "big" is expressed in Greek as "megas", in Gothic – as "mikis", in Latin – "magnus", etc. Here the availability of these vowels is caused by phonetic and morphological signs (but not by sound symbolism).

In order to overcome the above mentioned problems, Levitskij suggests doing the following tasks: 1) to corroborate sound symbolism convincingly; 2) to explain the possibility of the existence of sound symbolism in synchrony (in different languages) and in diachrony (in the historical process of language development).

2. Phonosemantics, sound imitation and sound symbolism

The first part of the book sheds light upon theoretical backgrounds of phonosemantics. Levitskij takes the view that phonosemantics deals with sound imitation and sound symbolism. Nevertheless, it is obligatory not to mix up the above-mentioned notions in the process of theoretical and experimental studies. The reason for it is that both phenomena are quite specific and require their specific methods of study.

Levitskij gives a brisk overview of the typology of sound symbolism. He considers that the classification by the Taylors is the most appropriate in which subjective (the connection of sounds and senses is observed in human mind) and objective (the connection between sounds and senses are fixed in the words) types of sound symbolism are distinguished.

Furthermore, the author places special emphasis on three important problems discussed in the works by American and Canadian researchers in the 50-60s of the previous century:

a) The methodology of experimental study;
b) The character of sound symbolism;
c) The causes of evoking sound symbolism.

Describing the general history of the study of sound symbolism, the author focuses on the research of the given phenomenon in Western Europe, in the USA and Canada and in Eastern Europe (the countries of the former Soviet Union).

The study of Sound symbolism in Western Europe

Levitskij notices that sound symbolism was not considered to be a linguistic problem in Western Europe. This predetermined the character and methods of its studies as well as the selection and the analysis of the material which were performed by typical for that time linguistic methods (without applying mathematical procedures). As a result, the above mentioned problem was not solved by taking into account only linguistic procedures. The only known experimental research in Western Europe was carried out in 1935 by G. Müller (Germany). In this experiment the researcher presented the informants unknown words chosen from “exotic” languages supposing to find the appropriate meanings of the words. The data were grouped by Müller into the classes (Strukturen). In spite of the fact that the researcher came to the confirmation of sound symbolism, his experiment is characterized by a number of methodological errors. The experimental works published in American and Canadian journals in psychology and European linguistics remained unknown. In such a way, the problem was not solved in Europe. Nevertheless, the attempts of the search for theoretical fundamentals of sound symbolism have been undertaken by Werner (1932), Sieberer (1947), Wandruzhka (1952), Kronasser (1952), Ullmann (1964).

The study of Sound symbolism in the USA and Canada

The main focus of the research was on the experimental research of sound symbolism and on the development (improvement) of the methods for the experiments. The first

experimental research was carried out in the USA by E. Sapir (1929). Sapir gave English native speakers senseless words *mil* and *mal* which were to be correlated with a big or a small table. As a result, sound [i] was evaluated as something small, whereas [a] is used to mean a big table. S. Newman (1933) investigated different vowels and consonants using Sapir's methodology. The researcher explored a larger list of English words with the meanings of "big", "small" aiming to corroborate that the evaluation of sounds is not connected with language habit. S. Newman made the following conclusions: a) English vowels were arranged according to small-big scale in the following way: /i/, /e/, /ɛ/, /ae/, /a/; b) the consonants from small to big were ordered in the following way: t, d, p, b, k, g (voiceless are evaluated as small); c) the evaluation of sounds is connected with phonetic (acoustic-articulatory) characteristics.

The previous investigations of sound symbolism in the USA and Canada were devoted to the improvement of the methods for the research of sound symbolism (Brown/ Black/Horowitz 1955, Brown/Nuttal, 1959, Miron, 1961). Theoretical backgrounds of sound symbolism were studied in the second half of the twentieth century (Brown, 1958; Taylor/Taylor, 1962; Weiss, 1964; Ervin-Tripp/Slobin, 1966; O'Hala, 1994).

The study of Sound symbolism in the previous countries of the USSR

The study of sound symbolism was at first restricted to the study of sound imitative words. In this case, V. Levitskij outlines that sound imitative lexis are on the periphery of the language. In 1965 Gazov-Ginzberg published the monograph in which the researcher had gathered and analyzed the vast material on sound imitative lexis in the Semitic languages. At the end of the book, Gazov-Ginzberg did a statistical analysis which showed that 115 verbal roots in the Semitic languages out of 181 could be regarded as sound imitative by their origin. In such a way, it was the basis for a detailed study of sound symbolism.

The interest in the experimental study of sound symbolism was aroused in the soviet linguistics in the mid 60ies. In 1967 Leont'ev published the book *Psycholinguistics* which motivated to undertake experimental research of Sapir/Whorf's hypothesis about the correlation of the language, thinking and sound symbolism. In 1966 a series of material about the perception of the sounds under Panov's supervision was published. The results of the experiments by Stern and Levitskij were published in 1967.

In the book "Phonetical meaning" (1974), A. P. Zhuravlov introduces the notion "the content of a language form" (p. 15-16), determines the phonetic meanings in the structure of a word and publishes the results of the experimental research of Russian sounds according to 25 scales of Osgood's semantic differential. The researcher made an attempt in his book to measure "the content aspect of sound forms" of text. Here Zhuravlov proposed the formula for finding numbers of phonetic meaning of word, where special attention was paid to the position of stressed and unstressed, initial and non-initial sounds.

The book by Levitskij "Semantics and phonetics" (1973) deals with a number of problems: typology of language signs and the place of symbols among them, nature and the causes which evoke sound symbolism, the degree of universality of sound symbolic rules. The vast majority of experiments were undertaken during 1970-90 in Chernivtsi University where the focus of the research was on the study of symbolic properties of vowels and consonants in English (Komarnutska, 1985), German (Kushneryk, 1987), Ukrainian, Russian and Moldavian (Levitskij, 1973). Special attention was paid to the connection of phonetic meaning and connotative meaning of word, phonetic meaning and motivation, semantic and stylistic functions of the combinations of phonemes.

3. Experimental study of subjective sound symbolism

Levitskij holds the view that any experimental study of sound symbolism is based on the creation of such an artificial situation in which the informants compare the forms and content (sound and sense). According to the quantity of sounds and senses, it is possible to distinguish the variants "1-0" (experiment by G. Müller and Levitskij, 1935), 1-1 (Brackbill/Little, 1957), 2-1, 1-2, 2-2 and others.

Levitskij explains that the variant "1-0" means that the informants evaluate only one word, but there is no given meaning. The task of the informants is to guess the meaning of these words. The variant 2-2 means that the informants evaluate two words and two meanings.

Taking into account the amount of the material, Levitskij distinguishes the following procedures. One of these procedures is that informants evaluate the words (forms) of the unknown natural language (forms) and the words of the native language (content). This procedure is called "matching experiment". In spite of the wide application of the matching experiment in the USA, it has a lot of faults (a detailed analysis is given in the work by J. Peterfalvi, 1970). The most important fault is that "guessing" the meanings of foreign words may be affected by phonetic and structural resemblance, but not by phonetic symbolism.

Levitskij considers that the sounds given to the informants for evaluation can be divided into one, three or many phonemic. The procedure with one phoneme was used by A. P. Zhuravlov (1974), S. Ertel (1969). Nevertheless, the most popular model was represented by trigrams (consonant-vowel-consonant type). As far as the meanings are concerned, the meanings of the sounds can be represented a) with the help of the words of the native language; b) by colors. The meanings are usually given in the form of Osgood's semantic differential with 5-7 divisions (the poles of this scale are antonymic, as small-big, weak-strong). Levitskij states that no essential difference between the oral or written presentation of the material is revealed.

4. The causes and the forms of sound symbolism

Levitskij is of the opinion that the methodology of the investigation of sound symbolism is connected with two other problems:

a) What causes sound symbolism?
b) The character of sound symbolism.

At the first stage of the study, it was supposed to prove that the informants of different languages guess the same meanings of the sounds. This hypothesis about "international" character of sound symbolism was experimentally confirmed in the investigations by S. Tsuru and G. Vries (1953), R. Brown and his collaborators (1955-1959). They used the matching experiment (variant 2-2) applying the procedure with artificial words of the type (CVC). Nevertheless, the works by the Taylors cast doubt on the authenticity of the previous results. The Taylors predominantly criticized the errors of the procedures in all previous research with the help of matching experiment. In such a way, the universal character of sound symbolism was under doubt. In this case, the Taylors relied upon the findings of their own experimental research for English, Korean, Tamil and Japanese. The researchers gave the pupils to compare 144 artificial words (CVC) according to 4 scales of Osgood's semantic differential. These words were created by the Taylors with the help of the so-called "Latin square". The results were interpreted in the following way:

1. Sound symbolism has a national character. In particular, English consonant [t] was evaluated as small whereas the Korean [i]– as big;
2. Language habit makes up a basis of sound symbolism. For example, English phoneme [g] occurs at the beginning of the words with the meaning of something big (grand, great, grow, gain, gross). That's why the respondents evaluate CVC with initial [g] as big.

Here, the closer the genetic connection between the languages is, then the higher is the probability that sound symbolic properties may overlap (and vice versa). The Taylors called their conception "the theory of backward connection". This theory is regarded in psycholinguistics as a variant of "associative theory"

Ertel's findings were in favor of universal character. A series of his experiments (1969, 1966, 1965) were conducted in which Czech and German native speakers ascribed the same meanings to the artificial words (the type of words "okai – elini"). R. Tarte and L. Barrit (1971) used the same procedures (English and German received artificial words and pictures) and made conclusions that the universal character of sound symbolism required further experiments.

In such a way, Levitskij makes conclusions that the confirmation of either international or universal character required further exploration. In this situation it is possible to use two procedures:

(1) To determine phonosemantic properties of the same sounds in different languages;
(2) To compare these results in different languages.

5. Sound symbolism and sound-letters

Objecting to the Taylors and Levitskij's methodology, A. P. Zhuravlov (1974) wrote in his book:

Differential phonological features play a significant role in the sounds for the speakers of different languages. In particular, the feature "mild-hard" does not exist in the English language. In this case, the English native speakers do not react to it. As a result of a statistical procedure, the initial consonants in the English sound combinations "dip-dop" can be perceived as the same. Nevertheless, the Russian native speakers react to the change from mild-hard because of this differential feature. In such a way, the initial sounds in the Russian sound combinations "∂un-∂on" [dip-dop] will be perceived as different ones. This was not taken into consideration in Levitskij's experiment. Therefore, the outcomes for initial consonants cannot be compared with the results of the other languages.

Levitskij considers that Zhuravlev's arguments are against the methodology used in the experiment. In particular, Levitskij explains the conditions under which his experiment took place: a) Levitskij received the data about the symbolism of sounds in all positions (tig, tag, tog, tum) with the help of Latin square; b) in the process of treating the data, it is possible to obtain the following three types of grades: 1) for mixed types (i.e. without palatal and velar); 2) for velar consonants; 3) for palatal consonants (the grades for [t] in the words "тад", "тум" and "тид", "тюг" were separately analyzed). It is this procedure which was used by Levitskij 1973 in studying the symbolism of Ukrainian vowels and consonants.

Zhuravlev gave the letters to the informants which were pronounced by the experimenter and printed in the questionnaire. But it remains unclear in what way the experimenter presented the palatal sounds which were written in all his tables in the following way: н', л', м', с', etc. This procedure was not mentioned in Zhuravlov's work. Zhuravlov

was supposed to think that phonetic meaning might be represented by sound-letters fixed in our mind. In this case Levitski cast doubt upon the existence of the letters н', л', м', с' in the Russian language. In the given situation, Levitskij asks the following questions: How were the isolated palatal consonants pronounced by the experimenter? With what LETTERS are palatal consonants associated? With letter + comma? But there are no letters with commas in the Russian language.

The major advantage of Zhuravlov's methodology is its convenience and economy of time in order to prepare and treat the material…If there is an essential correlation between the data obtained with the help of CVC method and the isolated letters, then the preference will be given to Zhuravlov's methodology (due to its convenience). The CVC method will take advantage over the isolated consonants unless one may detect the correlation in question.

In this situation, Levitskij made an attempt to compare the data by Zhuravlov (1974:46-49) and by him (Levitski, 1973:25-26). The author dealt with the data according to two scales: the scale of size (small-big in Levitskij and big-small in Zhuravlov) and the scale of evaluation (pleasant-unpleasant by Levitskij and good-bad by Zhuravlov). The Russian vowels according to the scale of size were arranged in the following way (from big to small): и, а, у, э, о, ы; in Zhuravlov – а, у, и, э, ы, о; the range correlation was used so as to treat the results (Table 1).

Table 1
Range correlation for Russian vowels according to the scale of size

Vowels	Levitskij	Zhuravlev	The difference of ranges	d^2
И	1	1	0	0
А	2	3	1	1
У	3	2	1	1
Э	4	4	0	0
О	5	6	1	1
Ы	6	5	1	1
				$d^2 = 4$ r = +0.89 df = 4

It means that no essential divergences between Zhuravlov and Levitskij's data were found. Quite opposite results were obtained according to the scale of evaluation (Table 2):

Table 2
Range correlation for Russian vowels according to the scale of evaluation

Vowels	Levitskij	Zhuravlev	The difference of ranges	d^2
И	1	3	2	4
А	2	2	0	0
У	3	1	2	4
Э	4	5	1	1
О	5	4	1	1
Ы	6	6	0	0
				$d^2 = 10$ r = +0.71 df = 4

In spite of the high value of coefficient r, it is not significant. There is an essential divergence between the values of vowels according to the scale of evaluation. While comparing the values for Russian vowels, Levitskij has revealed r = +0.55 at the value in Table r = 0.58 according to the scale of size (the value of coefficient of correlation does not reach the significant mean); according to the scale of pleasant-unpleasant the coefficient of correlation is lower r = +0.47 at r = +0.58.

In such a way, Levitskij has noticed the resemblance with Zhuravlov's data according to the scale of size. In order to answer the question which methodology can lead us to more authentic and accurate results, it is necessary to conduct the further experiments. Here it is worth taking into account Kalyta's observations that some letters may evoke certain associations in the speaker. In particular, E is associated with energy, M – with satisfactory feelings, A – awakening (Kalyta, 2001:56). This observations cast doubt upon Zhuravlov's methodology. It is possible to suppose that Zhuravlov could have received quite different results by presenting the letters G and g according to the scale of "round-angular".

It is also supposed that the forms of Cyrillic Г and Latin g may influence the grades of the sounds according to the scale of form. The impact of graphemes was mentioned by Ch. Bally (1955:149-150). Moreover, the influence of the graphics on the result of the experiments was connected with the notion "black", discussed by Fischer-Jorgensen (1978:87). R. Wescott pointed to the iconicity of hieroglyphs (1980:6-7). This allows Levitskij to make the conclusion that the forms of the letters may affect the results of the experiment. It shows that it is necessary to use such methods in psycholinguistic experiments where sound-letters will have different surroundings. Therefore, Levitskij declares that the method of CVC (according to the Latin square) is more correct.

6. Subjective sound symbolism

The existence of subjective sound symbolism (the correspondence between phonemic and semantic units in the human mind) was initially corroborated and published by E. Sapir and S. Newman. In spite of it, Levitskij points out that it is not clear nowadays whether sound symbolism has national or international character. In order to demonstrate the international character of sound symbolism, it is necessary to prove convincingly that the correspondence between sound and meaning does not depend on the genetic relation of the languages. Or vice versa, the sound-meaning correspondence depends on the relatedness and the same sounds possess the same meanings in different languages. Unfortunately, the comparison of symbolic properties in different languages encounters serious difficulties. Firstly, this comparison presupposes conducting numerous experiments according to different scales on the material of different languages. Secondly, it is necessary to perform the research according to the same methodology. In Table 3 the data about the symbolic meanings of English, German, Russian, Ukrainian, Moldavian and Hungarian is given according to several scales. There is a source in the right corner of the table from which the data are taken (The data about Tamil, Japanese and Korean are mentioned in the article by I. and M. Taylors).

It is a well-known fact that the number of vowels and consonants are not the same in different languages. Moreover, the quantity of the analyzed data received by the Taylors (6 vowels and 12 consonants) and Levitskij (40 consonants) are not the same. In Table 4 Levitskij represents only the ranges of the sounds (not the values of symbolic meanings). It makes the further comparative analysis easier. The French vowels according to the scale of size were presented by two rows – front and back (like J.-M. Peterfalvi's experiment (1970)).

How is it possible to compare the given psycholinguistic data and subjective sound symbolism? First of all, Levitskij has changed some Tables where there were a maximal

number of languages in order to have the general phonemes for all languages (like Ertel's experiment).

In order to make the conclusions about symbolic properties of the sounds in different languages, it is necessary to reveal whether the distribution of sounds depends on the relatedness of the languages. The comparison of this distribution in both tables with the help of correlation analysis has revealed the correspondence between English and German (+0.15), between Russian and Ukrainian (+0.38), between English and Moldavian (+0.24), between Moldavian and Tamil (+0.25), between Russian and Hungarian (+0.28), between Ukrainian and Tamil (+0.33), between Moldavian and Ukrainian (+0.67). In this case, Levitskij makes the conclusion that the value of coefficient for related languages (English and German, Russian and Ukrainian), the value of coefficient between the unrelated languages (Ukrainian and Tamil, Russian – Hungarian) or between unrelated languages (Ukrainian and Moldavian). It is possible to conclude that sound symbolic similarity between the languages does not depend upon their genetic relatedness. This conclusion is confirmed by the data received by Levitskij (1973, 27-28) and Levitskij/Sternin (1989:170).

Table 3
Statistical significant distribution of vowels and consonants

Scale	Vowels	Consonants
Size	[i, e] – [a, o, u]	[p, k] – [r, b, d]
Strength	[ɛ, i] – [a, o, u]	[l] – [r, d]
Activity	[i, e] – [a, o, u]	[m, n] – [r, t, p]
Evaluation	-	[m, n, t, d] – [ts, tŝ, ŝ, s]
Temperature		[v, tŝ, z. j] [r, z, b]

Levitski noticed that the opposition of semantic features can be done by the opposition of sound [r] to sounds [l, m, n]. This opposition was also found between voiced [b, d] and voiceless [p, t, k]. The exception was a scale of evaluation where the pleasant [m, n, l] are opposite to the fricatives and affricates ts, tŝ, ŝ, z.

There are numerous oppositions. In particular, Ukrainian sound [d] was evaluated as unpleasant, Moldavian sound [a] – cold, although it is hot in German, Russian and Ukrainian. German sound [b] (according to S. Ertel's data) is slow. Kushneryk's (2004) data have shown that this sound is neutral. Nevertheless, sound [b] is evaluated in the other languages as fast. In this case, Levitskij considers that the comparison of symbolic meanings is more effective. Therefore, the researcher transformed the tables with symbolic features of sounds into the tables of features using chi-square procedure. As a result, this statistical procedure has helped to reveal the significant statistical connection for the alternative tables ($x^2 \geq 3.84$). In such a way, the notion of "big" is expressed by voiced, vibration, back row and low; the notion small – voiceless, lateral, front row, upper and mid; the notion strong – voice, fricative, plosive, vibrating, back row, labial; the notion weak – voiceless, sonority, lateral, front row; notion fast – plosive; notion "slow" – sonority, fricative.

The other features proved to be statistically insignificant. In this case it remains unclear whether the symbolism of scales of temperature, light, hardness has national or international character. In Ukrainian the scale of light is symbolized by the opposition of front row – back row, in the Moldavian upper – lower, but the symbolization of the scale of "temperature" is the same in Moldavian and Ukrainian: there are relevant oppositions [labial] – [back] (for consonants). In such a way, the received data by Levitskij (1973) can be supplemented by the data based on a) a larger number of languages; b) a larger number of

sounds; c) a larger number of scales. In this situation Levitskij concludes that subjective sound symbolism has an international character according to the scale of size. Moreover Levitski adds that international character of subjective sound symbolism is fixed in the other scales.

Levitskij outlines that it is better here to make additional conclusions. The experience has shown that the bearers of symbolic meanings can be three types of phonetic meaning: phoneme, phonetic feature and the units which can be designated as "sound symbolic complexes". The last notion is represented by a group of phonemes united by certain phonetic features which are in the opposition to the other symbolic complexes. In particular, semantic opposition [big] – [small] is symbolized with the help of the opposition of sounds [i, e] – [o, a] or [p, k] – [b, r, d]. The larger number of consonants has shown that the consonant complexes according to the scale of size include the following sounds [m, n, p, l, t, s] [r, d, b, g, dŝ]; the opposition for the scale of strength includes the following complexes [k, s, t, ts, p, r] : [l, v, m, n] (Levitskij/Sternin, 1989: 180). Therefore, Levitskij declares that the ultimate conclusion can be drawn after receiving the data about symbolic properties of vowels and consonants which belong to different language families (Finno-Ugric, Turkic, Semitic and others).

7. The comparative analysis of the data by the other authors

The majority of works on subjective sound symbolism were made with the help of matching experiment. According to Levitskij this does not allow us establishing more accurate semantic correspondence at the level of phoneme or phonetic feature. Nevertheless, separate data about the symbolic properties of sounds can be compared with Levitskij's. S. Newman 1933 arranges the vowels according to the scale of size in the following order [i, e, ɛ, ae, a]. The consonants were grouped in the following order [t, d, p, b, k, g]. Consequently, front, high vowels, voiceless, front labial consonants symbolize small size whereas low vowels, voiced and back consonants – big size. These results coincide with Levitskij's outcomes about semantic and phonetic correspondence according to the scale of strength at the level of phonetic features. According to Newman, sound [i] and [e] proved to be both small and light. In other words, there is a correlation between the scale of strength and light. That was also revealed in Ukrainian by Levitskij (r = +0.78 P = 0.01) (Levitskij, 1973:43).

Peterfalvi (1970) investigated symbolic meanings of French sounds. As a result, the sounds were distributed in the following order: [i], [y], [e], [o], [ɛ], [oe]. These results overlap with the outcomes for the other languages. For example, the scale "sharp-blunt" is symbolized by the French vowels in the following order [i, ɛ, e, y, oe, o, a, u, o, o], the scale of "light" – i, ɛ, e, a// y, ae, oe// u, o, o. These results coincide with the data about symbolic properties for Ukrainian and Moldavian [Levitskij, 1973, 1979]. The arrangement of Moldavian vowels according to the scale of form (from sharp to blunt) – и, е, а, о, э, у, ы; Ukrainian vowels (scale of light) – і, е, о, и, у, а. Therefore, sounds [i] and [e] symbolize light in all three languages – French, Moldavian and Ukrainian. Jakobson/Waugh (1979) established that [i] is associated with light and u – with darkness. Tarte/Barrit (1971) found that Czech and English referred the sounds [u] to elliptical figures and [i] – to angular.

Furthermore, the above-mentioned experiments have shown that sound [a] symbolizes a big size, [i] – small. Polukarova (1989) has found that front English vowels symbolize something "tender", "secure"; back – "terrible", "masculine"; labial – "round"; non-labial – "angular". This helped the researcher to state that the scale of form (sharp – blunt) belonged to the part of sensory continuum in which there is a correspondence between meaningful

forms and sounds. In spite of it, the symbolic meanings of labial vowels were also revealed in the majority of the experiments. Lihomanova has found that [i] is to be found in the verbs of movements, it is short and abrupt whereas [ae], [u] – slow and heavy.

In the previous works on sound symbolism (in particular J. O'Hala) there is a correspondence between the features of vowels and consonants and their acoustic frequencies: high tone, the vowels with the high second formants (i.e. [i]) and the consonants with the high acoustic frequencies are associated with small size; low tones, the vowels with low second formant (i.e. [u]) and low-frequent consonants are associated with a big size, mildness and a slow movement (Hinton et al, 1994:10). This coincides with Ultan's findings (1978): the size is symbolized by front high vowels in 90% of languages (the same data was received by Levitskij, 1969). S. Ertel investigated the symbolic properties of 22 German consonants and 17 vowels according to 18 scales. Before comparing Ertel's results, Levitskij advised to take into account two important circumstances:

a) After statistical processing and conducting a factor analysis, the results (18 scales) were brought into three scales (dimensions): strength, movement and evaluation;

b) In order to demonstrate the difference to the informants between short and long vowels, S. Ertel presented existing German words. According to Levitskij, this is a serious methodological error.

In this case, Levitskij makes a comparison between the results by Kushneryk and Ertel (21 consonants and 13 vowels have been compared). The comparative analysis has shown that the meanings of the words could affect the informants' guessing who evaluated the sounding forms of these words according to the scales of semantic differential. Moreover, Ertel criticized Canadian researchers Taylors for this error. Finally, S. Ertel includes in his list of the words the sound [ks] which is not a German phoneme but he did not include the phoneme [pf].

Therefore, a significant statistical correlation is observed in all cases of consonants. There is no significant correlation (except the movement) between vowels. Levitskij explains that the cause for it is Ertel's procedure of the experiment. This conclusion is grounded on 1) the coincidence of consonants' values by Ertel and Kushneryk; 2) the coincidence of the values for vowels in the experiments conducted at Chernivtsi University (the informants were Ukrainian and Russian) and at the University of Voronezh (the respondents were Germans).

Levitskij pays attention to the language with phonological oppositions of long and short vowels (German – English) whereas this opposition is used in order to symbolize the whole row of notions. In particular, the notion "slow" is symbolized in English and German by long vowels; "fast" – short ones. This observation was also confirmed by T. A. Polikarov.

The Dutch researcher E. Fischer-Jorgensen (1978) connected the notion "thick" with u, o, ə, whereas "thin" – e, ɛ, a. Here Levitskij notices that her experiment was conducted incorrectly: the informants received existing Dutch words designating notions thick-thin, mild-hard, weak-strong, etc which were to be correlated with one of the above-mentioned notions. At first Fischer-Jorgensen denied the influence of the words on the informants' choice. But then she admitted the possible influence (p.83).

8. Sound symbolism in four unrelated languages

Levitskij observes that the participants of the discussion predominantly check the results of the research received by their opponents. In particular, I. Taylor (1963) did a statistical analysis of the results of sound symbolism received with the help of matching experiment. She has come to the conclusion that these outcomes are not quite authentic in order to state the fact about the universal character of sound symbolism. In its turn, S. Ertel analyzed the

data which was presented by I. and M. Taylors to the native speakers of 4 languages (English, Japanese, Korean, Tamil). He revealed that 47% of senseless three phonemic trigrams CVC really had meanings in the languages in question. Therefore, the Taylors' results could be caused by the influence of the existing words' semantics (not by sound symbolic factor). The above-mentioned made Levitskij perform the further analyses of the Taylors' data in their article "Phonetic symbolism in 4 unrelated languages". In this case, Levitskij changed the table in such a way that vowels and consonants were arranged in the decreasing order of their average values. In order to apply accurate statistical criteria for detecting the similarity or difference between the languages, Levitskij took the following steps: 1) he made the tables of the frequency distributions of sounds on the poles of the scales; 2) he formed sound symbolic complexes and processed the data with the help of chi-square; 3) he wrote out all the sounds where the interrelation of frequencies for vowels is 2:1, 2:0, 3:1, 3:0, 4:0; for consonants – 3:1; 3:0, 4:0; The opposition of the frequencies of sound symbolic complexes is represented in the form of alternative tables.

Table 4
Alternative distribution of frequencies of sound complexes according to the scale of activity

Poles	[t], [d]	[m], [n]	Total
Left	6	0	6
Right	2	8	10
Total	8	8	16

$X^2 = 9.8$ P = 0.01 K = 0.77

In order to compare the received data with the outcomes of the other researchers, it is necessary to clarify and represent symbolic properties of sounds in 4 unrelated languages using phonetic features (as shown Table below):

Table 5
Phonosemantic regularities in 4 unrelated languages

Scale	Sound symbolic opposition (sounds)	X^2	Sound symbolic opposition (phonetic features)	X^2	Other languages
Size	[ə, u] : [ɛ, u]	5.3	Voiced - voiceless	3.97	Ukrainian German
Evaluation	[j, t] : [k, g] [ə, o] : [u]	10.6 9	Upper non-upper	6.4	Ukrainian Moldavian German
Activity	[b, t] : [m, n] [i]:[ü, U]	9.6 4.7	Obstruent – sonority; Upper - non-upper Labial – Non-labial	12 4.3 4	Ukrainian German Moldavian Ukrainian
Temperature	[r] : [t, k] [ə, o] : [ɛ, ü]	9 6.5	Voice-voiceless Vibration – voiceless Labial - velar	9 6.5	Ukrainian German Ukrainian Moldavian

In such a way, the data received by the Taylors does not correspond to the conclusion made by the researchers (first of all, the fact about the national character of sound symbolism). In this case, Levitskij thinks that it would be relevant to scrutinize symbolic properties of sound complexes (not separate sounds). Furthermore, the data of the experiments could be distorted by the procedures: instructions, the choice of the stimuli, the forms of the material, the informants, the way of processing the received data.

In spite of it, Levitskij considers that it would be incorrect to refute "the theory of backward connection" (Taylors' hypothesis). The previous observations have shown that the values of the sounds may depend on the associative connection of these sounds with certain meanings.

9. The study of sound-color associations

At this stage Levitskij dwells upon Zhuravlov's (1974) findings concerning the "colors" of Russian vowels. In such a way, Zhuravlov has revealed that Russian [a] is associated with red, [u] – with blue, [o] – yellow. In this case, Levitskij decided to conduct the experiment in order to reveal sound-color associations of Ukrainian and Moldavian vowels (1973:86-89). The aim of his research was not to detect sound-color correspondence, but to corroborate the following hypothesis: the informants are intended to use the mechanism of language habit in the process of referring a certain color to a vowel.

As a result, red color is associated in the Ukrainian language with [o] and [a] (the author suggests comparing Ukrainian word for red – *chervonuj*), yellow [o] (*zhovtuj*), blue – with [и] (*synij*), green – [e] (*zelenyj*), but the connection with a violet color and sound [y] was influenced by the word "lilac" (*byzkovuj*), brown with [u] (*korychnevuj*). The analogical types were found in Moldavian: red (*roshu*) with [o], blue (*albastru*) – [e], [a], yellow (*talben*) – [a].

Levitskij explains it in the following way: "It is obvious that there are such dimensions in which the transposition of feelings occurs easily (we refer to it the scale of size, strength and hardness) and the dimensions in which the transposition does not take place easily (the scale of color, tastes and temperature). When the informants deal with easier scales, then the mechanism of synesthesia occurs. When the task is more difficult, then the search for solving it leads to switching on "the mechanism of language habit". Sometimes it happens that both mechanisms act by turn. In this case it is a cause for receiving different results even by one researcher. Changing the instruction (giving the informants the mechanism of transposition or language habit), we receive different results. As a result, the transposition of one type of feelings into another creates the basis of sound symbolism. In spite of it, three factors may influence the results of tests: instruction, the scale of size and the whole phonological system of the language. (Levitskij, 1973:88-89).

The full and detailed research in this area was undertaken by Prokofjeva (2007). The researcher introduced in her study a considerable number of works unknown to the readers. Nevertheless, Levitskij takes a keen interest in two questions concerning this research: (1) How does the researcher explain the causes of sound symbolism (synesthesia or language habit); (2) At what degree do her data coincide with Zhuravlov's.

It is possible here to find the following explanations to the first question. The same values for English and Russian sounds mean that color names have a significant impact on the associations of graphemes. In such a way, Ferdinand S. de Mendoza's observation (Levitskij finds this observation in the work by Prokofjeva) was confirmed on the basis of the experiments among non-synesthians speakers (Russian and English as native language). This remark is also in favor of universal color associations. Mendoza writes that "the syllables and words are colored in correspondence with the vowels and some influential consonants. The meanings (even in the names of flowers) do not influence the evoked image. Two synonyms

are colored in a different way (Prokofjeva, 2007). Here Prokofjeva writes that this notice is extremely important in so far as it demonstrates that the dependence of the association on the names of the color is not proven enough".

In this situation Levitskij does not understand Mendoza's explanation concerning the correlation of sound and color. What does it mean that the words are colored in accordance with their vowels and consonants? There is no exact correspondence between a sound and its color? Sound [a] may mean black and red; [e] designates two colors – light-brown and white, etc. The reader is unaware of what the informants evaluate – letters or sounds. Judging from Mendoza's data é/e with different types of stress, they are letters.

The second Mendoza's phrase denotes that the meanings, even in the flowers' names, do not affect the evoked image. Does it mean that the sounds in the names of flowers do not influence the color of words? So we may suppose that two different in sounds words designating the color of the same flower must be colored in the same way. But Mendoza gives quite an opposite comment: two synonyms are colored differently. In this case it remains unclear to what Mendoza's utterances Prokofjeva refers the comment on a lack of proofs concerning the dependence of the associations on the name of flowers? How to understand "universal color association"? Does it concern the similar sound associative correspondences in different languages? In this situation Levitskij suggests conducting a qualitative and a quantitative analysis of Prokofjeva's data.

Quantitative analysis. Table 6 is made in which the answers to the following questions will be given: a) Can the first vowel or consonant of the color names be in the structure of the graphemes which are symbolized by a certain corresponding color according to Prokofjeva's data; b) Can the stressed vowels in the names of the colors be in the graphemes which are symbolized by a corresponding color? c) Which graphemes take the first place according to their frequencies symbolized by a certain color?

Table 6
Results of qualitative analysis by L. P. Prokofjeva

The name of color	The first word's sound	Stressed vowel	The most frequent consonant	The most frequent vowel
1	2	3	4	5
Krasnij (red)	+	+	К+	А+
Sinij (blue)	+	+	С+	И+
Zelenij (green)	+	+	З+	Ё+
Zheltij (yellow)	+	+	Ж+	О+
Fioletovij (violet)	+	-	Ф+	Ю+
Oranzhevij (orange)	+	-	О+	Р+
Chernij (black)	+	-	Ч+	Ы+
Belij (white)	+	-	Б+	О+
Korichnevij (brown)	+	-	К/г+	Ы+
Red	+	-	R+	a-
Blue	+	+	B+	u+
Green	+	+	G+	e+
Yellow	+	+	Y+	e+
Orange	+	-	E+?	o+
Black	-	-	X+	o-
Brown	+	-	d-	u-
Violet	+	-	v+	i+
White	+	+	w+	i+

Red color is symbolized by [к] and [а]. Here [к] is the first consonant in the Russian word *krasnj* (red) and [а] is stressed vowels in this word. The same is observed for the letters [с] and [и] which are symbolized by a blue color and can be found in the structure of this color: the first consonant and the stressed vowels. These letters take the first place according to their frequencies among the respondents' answers.

Results: If "+" and "-" mean positive and negative answers, then 4 pluses are found in the Russian words "krasnij" (red), "sinij" (blue), "zelenij" (green), zholtuj (yellow); 2 pluses have "fioletovij"(violet) and "belij" (white), three pluses have "oranzhevj" (orange), "chernij" (black) and "korichnevij" (brown). A little different results were obtained for English color-names. Only four pluses can be found for blue, green, yellow and white; three pluses – violet, two pluses – orange and red, 1 plus – brown, no positive sign is given in black.

Table 7
The results of qualitative analysis of color symbolization

Russian	The number "+"	English	The number of "+"
Krasnij (red)	4	Red	2
Sinij (blue)	4	Blue	4
Zelenij (green)	4	Green	4
Zheltij (yellow)	3	Yellow	4
Fioletovij (violet)	2	Orange	2
Oranzhevij (orange)	3	Black	0
Chernij (black)	3	Brown	1
Belij (white)	2	Violet	3
Korichnevij (brown)	3	White	4

In such a way, 7 Russian words out of nine (77% out of 100%) are positive; English words – from 56 to 77% are positive (from 5 to 7 words out of 9). The difference between writing and pronunciation of English letters and words are more obvious than in Russian. It is worth mentioning that Russian native speakers filled in the questionnaire under the experimenter's supervision; English native speakers received the questionnaires per Internet. The conditions of the experiments were not the same. In this case, it is relevant to take into account the time for "guessing" the answer by the informants.

Judging from Prokofjeva's data (2007), the Russian native speakers refer the meaning "black" (Russian "chernij") not only to the grapheme /ч/ (F = 374 – 1st place) but to /ш/ (F = 0258), /п/ (F = 232), /т/ (F = 220), /х/ (F = 187), /д/ (F = 167), /щ/ (F = 158). In this row there is a dominance of whistling sounds /ч/, /ш/, /щ/) which (together with /х/) are evaluated as bad and unpleasant (Zhuravlov, 1974:46-47). Here Levitskij concludes that the mechanism of association is "switched on" in the informants while evaluating the represented word (/ч/ черный - chernij). The word's association of the words may also depend on the paradigmatic and syntagmatic relation of this word. In the above-given example, the connotations of "black" (chernuj) are bad, unpleasant. The paradigmatic associations of the word *black* (chernuj) is *dark, tragic, corpse, bad*; the syntagmatic association – "black hole" (chernaja dira).

The above conducted qualitative analysis casts doubt upon the fact that the mechanism of association (not synesthesia) makes up the basis of sound symbolic correspondence in Russian and English. In order to give an exact and accurate answer to the question (at what degree do the graphemes (which are able to symbolize certain colors) coincide with the graphemes (which are in the structure of names), it is necessary to do a quantitative analysis.

According to Levitskij, the most adequate methods in this case are chi-squared test and coefficient K.

Table 8
The distribution of frequencies of graphemes (k, r, a, s, n, I, j) in the symbolization of colors

Graphemes	Red	Other colors	Total
K, r, a, s, n, I, j (красный)	1439 a	c 3745	5184
Other graphemes	2343 b	d 15838	18180
Total	3781	19583	23364

$X^2 = 658$ K = 0.168

Table 9
The values of x^2 and K for sound symbolic association in Russian and English
(on the basis of Prokofjeva's data, 2007)

	X^2	K	Color	X^2	K
Sinij (blue)	1878	0.283	white	425	0.138
Zelenij (green)	1028	0.21	violet	421	0.137
Krasnij (red)	658	0.168	yellow	4.03	0.134
Chernij (black)	232	0.10	blue	168	0.087
Belij (white)	145	0.08	green	146.5	0.081
Zheltij (yellow)	142.6	0.08	red	94.3	0.065
Fioletovij (violet)	141.5	0.08	orange	84.8	0.062
Korichnevij (brown)	44.6	0.04	brown	4.8	0.015
Oranzhevij (orange)	38.8	0.04	black	68.4	-0.055

$X^2 = 0.12$, K = 0.074

1) All sums of chi-square have the necessary statistical significance, the lowest sum of chi-square was revealed for English "brown";

2) All coefficients are positive (except English *black*);

3) In any case (except black) the color in English and Russian are symbolized by sound-letters which are in the structure of the word. This peculiarity can be measured with the help of coefficient of contingency;

4) The colors in Russian and English can be divided into three subgroups relying on the degree of coefficient K: a) the value of K higher than 0.10; b) K between 0.05 – 0.10; c) K lesser than 0.05. Consequently, we refer to the strong symbolization such Russian words a) sinij (blue), green (zelenij), red (krasnij); b) average – black (chernij), white (belij), zheltij (yellow), fioletovij (violet); 3) weak – brown (korichnevij), orange (oranzevij). The results for English colors are as follows: a) strong – white, violet, yellow; b) average – blue, green, red and orange; c) the weak – brown.

10. Objective sound symbolism

According to Levitskij, the linguistic methods of the study of objective sound symbolism have not been changed for the last 100-150 years (they did not even differ in the Antique time). As usual, these procedures consist of the following stages: a) at first we set out a hypothesis about the connection of certain sounds with their meanings; b) After setting out the

hypothesis, a researcher gives a list of lexemes which share the common phonetic and semantic features. Two faults are characteristic of this procedure:

1) The words for the list are chosen by the researcher at random. In particular, aiming to corroborate the connection between "an abrupt change of the direction" and the German consonant [k], K. Fenz (1940) made a list of lexemes *corner, chin, hook, knee, knuckle, curve.* In this situation, the list includes only the words with a certain phoneme [k]. Nevertheless it remains unclear whether there are the words with the same meaning but different sounds. For example, Jespersen mentions English words with the meaning *little, tiny, wee* [i]**,** but he failed to mention more frequent (in comparison with *wee, tiny*) word *small* [o].

This problem was not solved by Wescott either. In the preface of the book "Sound and Sense" Bolinger calls Wescott a "poet" who has a lyrical perception of the reality (Wescott, 1980 : xii) in so far as Wescott illustrates his thoughts and hypotheses by choosing appropriate examples. In particular, R. Wescott explains that the marked forms (such as plural, past tense) are usually longer than non-marked (p.13). It shows the existence of iconicity in grammar. Even the word "longer" is longer than "long". Here Levitskij notices that the word "shorter" is also longer than short. Levitskij considers that it is absurd to talk about the iconicity in the given case (additional grammatical information must be expressed by the additional formal means). Therefore, the productive forms should be structurally more complex).

In case when the language facts contradict to Wescott's utterances, the author found the other examples in favour of iconicity. In particular, Latin *pis* (nominative case) with long *i* is opposed by R. Wescott by the forms *–ped* in the oblique case (where *i* is considered to be more intensive than *e*) which contradicts to the above-mentioned examples.

Dwelling on the frequencies of labial *m* in the words of relatedness, Wescott connects it with the movement of the toddler's lips. He considers that the frequencies of this sound in the English words *murmur, mouth, milk*, Estonian *mokk (lips), musi (kiss)* have an iconic (symbolic) character. Nevertheless, Levitskij declares that the enhanced frequencies of this sound in the words of relatedness are corroborated with the help of statistical research in the majority of languages. Therefore, Wescott's given examples witness about the subjective character of the methods for revealing the iconicity in the places where it cannot be found: it is enough to have a look at the translations of the given words with the sound *m*. The exception is the word *milk* (Russian *moloko)* which can be explained by the common origin of this word in Germanic and Slavic languages but even Latin word for *milk* is represented by another root.

According to Levitskij, Wescott's poetic perception of the language reality should be added by strict rationalism (best of all, by the usage of statistical methods). Levitskij concentrates his attention on Wescott's work in detail in order to show that the selection of language facts (which shows the functions of sound symbolism) has extremely subjective character even in the contemporary research.

2) The selection of languages under research is quite subjective. In particular, O. Jespersen (1933) gives the examples from Eskimo and Japanese, but there is no example of Russian (or other Slavic languages);

3) One list of languages has more words for analysis (10-15 words), the other – lesser (1-2 words). In some cases the researcher analyzed rarely frequent words, in the others – the high frequent ones.

In the second half of the twentieth century (predominantly in psychology), the researchers made numerous attempts to investigate sound symbolism experimentally. S. Ertel was the first to undertake it. Being psychologist, Ertel made a variety of methodological errors. That cast doubt upon the results of his experiment. He selected and grouped the languages under analysis not according to the linguistic criteria, but according to geographical

ones (the languages of Asia, Africa). As a result, his list included 4 Slavic (2 languages – close related), three Indo-Iranian (two close related – Hindi and Urdu), two Finn-Hungarian and only one Turkic. The same can be found in the work by Thorndike (1945).

Levitskij notices that it is impossible in the statistical experiments of this kind to divide the groups of related languages equally. Otherwise, the proportions of the frequencies of the same root are violated (accordingly, the same group of sounds). Levitskij made an attempt to solve this problem (1969) by (1) equaling the number of Indo-European and Non-Indo-European languages (14-14); (2) selecting from each group of related languages 3 languages.

Nevertheless, the first step (1) of selecting the languages proved to be unnecessary in so far as the degree of sound symbolic resemblance does not depend upon the degree of genetic relatedness; the second procedure (2) cannot be performed because in some cases the whole 'families' are represented by one two languages, the others – 10-15 languages.

That's why the best solution to the problem was to average the results of statistical data according to each group of the related languages. The main results of the experimental and statistical study of objective sound symbolism on the basis of 53 languages can be summarized in the following way:

(1)The highest symbolic potential in the natural languages are characteristic of vowels (i) and (a) and consonants l, r, t, m, p.

(2)The highest symbolic activity is characteristic of the scale of hardness (mild-hard), smoothness (smooth-rough), activity (slow-fast), light (light-dark), form (sharp – angular), size (little-big); the lowest - the scale of temperature and evaluation.

(3)The role of vowels and consonants in the symbolization of different scales is not the same: the scales differ in their vocalic and consonant activity. In particular, the vowels symbolize activity, form and size whereas consonants – hardness and smoothness.

(4)There is a correlation between the scales. This means that similar semantic units are characterized by the same phonetic units in different languages:

Table 10
The correlation between the scales

Scale	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Size		-20	-18	-05	+51	+26	+13	+09	-17	-15	+01	-01
Strength			+47	+50	+03	-32	-05	-26	-37	-29	+09	+15
Hardness				+10	+28	-36	+32	+09	-12	-27	-29	-16
Activity					+06	-03	-17	-34	-28	-64	+33	-18
Weight						+11	+49	+10	-28	-54	+18	+02
Evaluation							-05	-24	-01	+04	+04	+24
Light								+19	+03	+51	+58	-22
Temperature									+31	+16	+05	+22
Form										+18	-03	+18
Smoothness											-56	+17
Humidity												-08

(5) The distribution of sounds in the words denoting similar notions in non-related languages is not equal. This allows formulating statistical sound symbolic rules (regularities). For example, the notion of size is symbolized by the symbolic opposition upper-lower for vowels and voice-voiceless for consonants.

(6) Different features do not have the same symbolic potential. The features for vowels are represented in the decreasing order: upper, lower, mid, labial, back row, front row. The features for consonants are represented by vibration, front velar, voiceless, obstruent, labial, fricative, voiced, palatalization, back velar, affricates.

(7) The poles of the same scale have different symbolic activity. The highest symbolic activity is characteristic of small, strong, mild, light. The poles of the scales of activity (fast-slow) and form (round-angular) have the same symbolic activity.

11. Phonosemantic correlation between the languages

The problem of the correlation between the languages is very important for phonosemantics in so far as the choice of the theories depends upon it which may explain the cause and the existence of sound symbolism. If sound symbolic rules depend upon the genetic relatedness of the languages, then the associative theory makes up the basis of sound symbolism. Otherwise (symbolic rules do not depend on the genetic relatedness) the transposition of one type of feelings into the other does not depend on the associative connection between phonetic and semantic units.

According to Levitskij, the presence of the connection between the languages may be caused by two factors: a) by the similarity of the lexemes in the language structure (Russ. слабый, Polish *slaby*, Roum. *slab*), b) the similarity of phonetic units which are in the structure of the corresponding (not related etymologically) lexemes, caused by certain historical conditions of the development of lexico-semantic and phonetic system for each language (Russ. *мупой*, Germ. Stumpf, Indones. *Tumpel*, Chinese. *dundy*).

Levitskij does not exclude the influence of the first factor although he considers that its influence on the degree of the correlation between the languages is quite limited. The average coefficient of correlation for each group of Indo-European languages with the other Indo-European ones is lesser than the average coefficient of the correlation of these languages with the languages of the other groups.

12. Psycholinguistic methods of the research of sound combinations

The detailed investigations of the properties and functions of initial consonants in Engish (Lvova, 2005) and German (Levitskij/Najdesh 2011) have been conducted at Chernivtsi University for the last 10-15 years. Thirty English consonants *bl, br, dr, kw, sk, sn, skw, str, thr, tw,* etc. have been investigated in Lvova's research. With the help of semantic differential, the initial phonemes have been investigated according to the scale of strength (weak-strong), evaluation, activity, cruelty, roughness and size. In such a way, she received the following results: pleasant phonemes are *fl, pl, sm*; unpleasant – *kr, gr, kw, st, skw, sv, ts, fr*; strong – *br, kr, dr, fr, gr, pr, sk, spr, st, skw, sv, ts, fr*; weak – *fl, pl, sf, ts*. The further research was done in order to answer some theoretical questions. First of all, the data was processed with the help of correlation analysis with the aim of learning whether there is a correlation between the scales. The results are given in Table 6.

Table 11
The correlation between the scales

Scales	**1**	**2**	**3**	**4**	**5**	**6**
1 strength		-0.14	+0.61	-0.21	-0.33	+0.8
2 evaluation			+0.41	+0.9	+0.94	-0.2
3 activity				+0.25	+0.21	+0.4
4 roughness					+0.87	-0.28
5 cruelty						-0.29
6 size						

It is possible to see that there is a correspondence between the scale of evaluation and cruelty (0.94); evaluation and roughness (0.9), roughness and cruelty (0.87), strength and size (0.8), strength and activity (0.61). Therefore, the scales of evaluation and cruelty are the variants of one scale. The scale of roughness is close to them. The closer are the scales of strength, activity and size. These results are in favour of synesthetic theory of sound symbolism.

If we widen the table for finding the coefficient of correlation at 90', then it is possible to receive the coefficient of correlation for each combination of phonemes. Having done the necessary procedure, Lvova got the following rows of initial consonants connected with one another according to their symbolic functions:

Bl-Gl
Br-Kr-Fr-Gr-Dr-Sk-**Sp**-Spr-St-**Str**-Sv-**Tr**
Kl-Fl-Sm-Sw
Kr-Dr-Fr-Gr-Sp-**Spr**-Skw-**Str**-Sv-Thr-**Tr**
Dr-Gr-Spr-Thr-**Tr**
Fl-Gl-**Pl**-Sl-Sm-**Sw**
Fr-Gr-Sp-**Spr**-Skw-**Str-Sv**-Thr-Tr
Gl-Pl-Sm-Sw
Gr-Shr-Sp-**Spr**-Skw-**Str**-Sv-Thr-Tr
Pl-Sm-Sw
Sk-Sp-**St**-Tr
Shr-Spl-Spr-**Skw**-Thr
Sl-Sm-Sn-**Sw**
Sm-Sn-**Sw**

The fatted combinations have the highest coefficient of correlation. The above-mentioned row shows that the majority of initial consonants are connected according to the second element [r] and [l]. In such a way, it is possible to make the conclusion that the most important component is the second one in the combinations br, kr, fl, etc. The similar data were received by Najdesh (2011) in German.

13. The meanings of the combination of phonemes according to the data by the other researchers

English and German initial consonants were studied at Chernivtsi University in the second half of the twentieth century by Levitskij (1983), Levitskij/Zhernovej (1988). In particular,

Levitskij processed the roots with initial two and three phonemic combinations (using the dictionary by Pokorny, 1959). As a result, Levitskij found that German initial combination br is connected with the meaning of *fracture, crumble, roar*; English br- *quarrel, break, courageous, lively, clear, merry, clean.* In German protolanguage bhr- (Indo-European bhr- Germ. br) has the following meanings: *to destroy, movement, to shine, to swell.* On the whole, it is possible to consider the semantic components in these three experiments (English, German and Protolanguage) coincide. The common components are "to shake", "to swell", "to shine".

The initial combination bhl- possesses the following meanings in Protolanguage: "to shine", "to swell". English bl- means *to roar, pale, to chatter, spot.* German bl- denotes *to shine.* The initial phonemes sl is fixed in Indo-European with the meaning *to release, to relax, to smear.* English sl has the following semantic components: *sleep, untidy, saliva, slow, sleepy, idle, swamp.* This combination has always been the central focus of attention for numerous researchers. Indo-European sl- belongs to the semantic sphere "to press", "resistance, quietness". The meaning "sound" refers in German to str- and the meaning "to reap" belongs to English str-. The combination of phonemes st- is also connected with the idea of sound in Indo-European.

The phonemes st and the dental t have been investigated in the phonosemantic research. In particular, Meye characterizes t and st as sounding units connected with "stubborn resistance" (1990:254). Levitskij confirms that st is connected with the hypersemes [labour – press] and t – with the hypersemes [pull]. This is confirmed on the basis of the data in both phonosemantics and etymology. Meye notices that st- occurs in the row of words denoting "to resist", "to pull" (1939:195). It is interesting to notice that Meye's examples correspond to the connection st and t observed by Levitskij (the author gives the example of Lat. studeth "to be aimed", Greek "steibo" (step, stamp), ag. Stoc (stick)).

The phonemes gr- is connected in Indo-European with such notional spheres as "destroy", "sounding", "to press". The English combination kr (Germ. kr = ie. Gr) is characteristic of the following meanings – to press, to crunch, to murmur, crack, fracture, bang; in German – to scratch, quarrel, shriek. The similar meanings were observed by Smithers (1954). R. Luehr (1988:184-187) studied the expressiveness of the analyzed forms in the Germanic languages. He was not engaged with the investigation of the initial consonants, but with final phonemes with a certain expressive meaning. To them he referred – amp, -anf, -ing, -ang, -ank, -ind, -unk and the others occurring in Germanic lexical word stock which were divided in such semantic word classes: sound symbolic words with the meaning "glisten", "movement", "shining" and the words with the meaning "thin, narrow, calm, quarrelsome".

Part 2
The functions of sound symbolism in language and speech

Levitskij notices that the place, the role and the functions of sound symbolism in the language and in the speech (here it is relevant to dwell upon the other external and internal factors in the development of word-stock) fall out of scope in phonosemantic research. Taking into account the above-mentioned, the author tries to shed light upon the connection between the phonetic motivation of a word and phonetic meaning; between connotative, stylistic and morphological status of a word and the types of motivation; and finally their interconnection in the development of the word-stock.

Phonetic motivation in the lexico-grammatical status of a word

14. The types of motivation in the language

Morphological motivation is the simplest type of structuring the material side of a word. Segmenting the words includes the division of lexical and grammatical meanings (i.e. Be-deut-ung-s-lehr-e). If the boundaries of segmenting between the morphemes disappear due to some reasons (the development of the language), the motivation evokes the secondary divisions of words (i.e. Russian word 'zont-ik' /umbrella/).

Semantic motivation is not expressed in the word to the high extent, but it takes place. This motivation includes the procedure in which the name of one subject comprises the name of the other one on the basis of resemblance or contiguity (i.e. Germ. *Stab*).

Phonetic motivation is based on the fact that the sounding organization of the name is identical to the structure of the designated subject or phenomenon. In particular, the semantics of the German *Zickzack* (according to the explanation of this word in the Dictionary Klappenburg, 1978) includes such components as *scharf, Knick* and *hin-her*. That corresponds to the phonetic structure of this word in which there is opposition of high *i* and the lower *a* which can be found in the surrounding with the meaning of *abrupt, broken line*. Structural resemblance (not the material) makes up the basis of phonetic motivation. Sound imitative words witness about it. In particular, the sounding of one word may be transferred by a similar (but not identical) selection of phonemes (Germ. Kikeriki, Eng. Cock-a-doo-dle-doo).

According to Levitskij, there are three types of relations a) the organization of sounds is in correspondence with the meaning (the structure of both aspects are identical); b) the structure of sounds objects to its meaning; c) sounds and meanings are in neutral relation. The first type is positive, the second – negative, the third is neutral. Levitskij considers that *phonetic motivation* of a word is a positive relation between sounds and their meanings.

Phonetically motivated words include sounds imitative (the denotatum of a word is the meaning expressed by animate or inanimate creatures) and sound symbolic (the denotatum of a word is a subject incapable of producing sounds). These phenomena are interconnected and sometimes it is difficult to distinguish them.

Levitskij notices that there is growing interest in the study of sound symbolism during the last 20-30 years. Nevertheless, the problem of functioning phonetic motivation of a word and its methodology remains unclear. It is important to reveal what types of lexis and at what degree phonetic motivation of a word is available as well as the factors on which phonetic motivation of a word depends. Here we deal with the connection of phonetic motivation and denotative meaning of a word, morphological and stylistic status of a word, its frequencies.

15. Phonetic motivation and denotative meaning of a word

Sound imitative words denote the phenomena which appear as a result of the inter action of two or more objects (rubbing, click) or the sounds produced by the humans or animals, physiological processes (breath, chewing, sneezing). The classification of sound imitative lexis is fully given by S.V. Voronin (1982:43). As far as the classification of sound symbolic lexis is concerned, it is considered to be one of the most difficult problem in phonosemantics. In particular E. A. Hurdzujeva (1873:2) proposed 16 basic thematic groups of words possessing symbolic properties (wind, natural phenomenon, running and walking, movement, gaiety and love, grief and malice, etc). The other researchers refer to this category the designation of something round, smooth, near-far, little-big (Kojbaeva, 1987; Mazanajev, 1985). Levitskij considers that the detection of the dependence between a certain type of

denotative meanings and phonetic motivation of a word can be established with the help of statistical methods.

In this case, it is relevant to mention the research by J.-M. Peterfalvi (1970) who did the analogical analysis on the basis of the French language. He divided the proposed 73 words (adjectives, nouns and verbs) into 5 semantic classes. The informant evaluated the degree of the correspondence between the sounds and meanings of the proposed words. As a result, the highest phonetic motivation proved to be in the words with the meaning of perceptual experience, movement and sounding, the lowest motivated words – the words with abstract and concrete meanings.

Komarnytska (1985) under Levitskij's supervision has found that the highest phonetical motivation is characteristic of the words referring to the following semantic classes: sounding, movement, perceptual experience, light, emotional state, size and form. The negative or neutral relation is observed in the classes of intellectual activity, moral qualities, colour, and the parts of space. The results for the German language (Kushneryk, 1987) are as follows: the semantic classes with the highest phonetic motivation – sounding, movement, size, distance, positive features and natural phenomena. The lowest motivation belongs to the words of light, colour, the human state. In such a way, the results of English, German and French coincide. This allows the author distinguishing the main classes of words having the highest phonetic motivation – sounding, movement, sensory continuum.

16. Phonetic motivation, frequency and stylistic status of a word

On the basis of the previous research, Levitskij declares that the motivation of a word is connected with its expressiveness. The author supports Kanungo and Lambert's opinion that long-lasting repetition of the word (semantic and verbal satiation) leads to the loss of meaning (1963:421). In this case, Levitskij concludes (by taking into account the investigations by Kushneryk (1987) (German) and Komarnytska (English) (1985)) that the frequency of the word's usage is connected with motivation and expressiveness. In the process of studying the frequencies (here the frequency dictionaries were used) and phonetic motivation of the words in the English and German languages, the author has revealed the following features: the high-frequented words are motivated in the English language more than low-frequented. In other words, phonetic motivation of a word is more characteristic of the high frequency. In contrast, the highest motivated German words are both frequent and rare lexemes. Here Levitskij makes the conclusion that the wider the sphere of the word's usage, the closer its content is. Having reached the peak of its frequency, the word loses its expressiveness and comes into lesser frequent, and its place is taken by more motivated lexeme. As far as the stylistic word's status is concerned, then it was found in English and German that the effect of sound symbolism is found in the sphere of illiterate lexis (colloquial, vulgar, rough and dialect words).

17. Phonetic motivation and connotative meaning of a word

The connection of phonetic motivation (PM) and the connotative meaning of a word was investigated by M. Wertheimer (1958) and L. Weiss (1968). In particular, Wertheimer detected that PM of a word (the words which meanings correspond to their sounds) possess the connotative meaning to the higher extent in comparison with the words which do not correspond to their meanings. In spite of it, J. Weiss changed the procedures which were in Wertheimer, cast doubt upon the previous conclusions and formulated his hypothesis: the correspondence between meaning and sounds of a word depends on its belonging to certain

grammatical classes. Furthermore, N.A. Pavluk devotes the study to the research on the reasons which influence the change of a word's connotative meaning. The researcher has found that the speed of this change depends on two factors: on the distance between connotative and phonetic meaning and on the word's belonging to certain semantic group (1976:72). Levitskij concludes that no problems in the above-mentioned works were solved due to a small quantity of the research devoted to a) the connection of connotation and sound symbolism, b) the contradiction of the results by the authors and finally c) the errors in the research procedures. Criticizing Pavluk's research, Levitskij notices a small number of measurements do not enable the author of the research to state about the influence of the first factor (the distance between connotative and phonetic meaning). Dwelling on M. Wertheimer (1958) and J. Weiss' (1968) investigations, Levitskij observed that the researchers did not give the informants the notions, but the words in their experiment. In this case, it was unclear to Levitskij what the informants evaluated – the connotation or the symbolic properties of words.

Taking into account these difficulties, Levitskij decided to measure with the help of semantic differential the sounding of the words which meanings were unknown to the respondents. And then he was to compare the results with the values of connotative meaning (CM) of these words. It was possible to do this task by using the famous work of American researchers Jenkins (1958) about the connotative meanings of more than 300 English words and the results of the experiments conducted at Chernivtsi University. In such a way, the experimenter had two rows of grades: 1) the values which measure the degree of correspondence of 286 English words and polar notions of three scales – activity, strength and evaluation; 2) the values which measure CM of 286 words. The statistical analysis has shown that the values of coefficient of correlation between PM and CM of words for all three scales have statistical significance: 0.33 (the scale of activity), 0.41 (the scale of evaluation), 0.37 (the scale of strength). Here Levitskij outlines that the values of CM (which were considered in the works of American researchers) proved to be the sum of two vectors – connotation and phonetic motivation. The author of the book concludes that the informants in his experiment were aware of neither connotative nor denotative meaning of the proposed sounding complexes. In this case, significant statistical coincidence of CM and PM of words cannot be explained by the effect of connotative meaning. In other words, CM does not coincide fully with PM. At this stage Levitskij formulates the following research question: On what factors does the degree of correspondence of CM and PM depend? In the form of hypotheses he sets out the following three groups: a) the frequency of a word; b) the word's belonging to a certain grammatical class; c) the word's belonging to a certain semantic class. Unfortunately, the effect of b)'s factor was impossible to study due to the absence of the grammatical indexes in the list of English lexemes. The other two factors were studied with the help of statistical and psychometric methods.

As a result, Levitskij has found that the more frequent the word is used, the higher degree is that PM corresponds to CM. This conclusion coincides with the above-mentioned conclusion. As far as the third factor is concerned, then it was measured with two statistical methods: the coefficient of correlation between PM and CM and the average value of violation of the value CM from PM. The slightest difference between these two types of values is observed between the words of movement, sensory notions, the highest is observed for the words designating abstract notions. If we take into account that PM is approaching to the connotative meanings of the group of words where the degree of phonetic motivation is established to the higher extent, then it is possible to conclude that the connotative meaning is created due to the phonetic motivation of a word.

At this point Levitskij summarizes that phonetic motivation and connotative meaning is closely connected with each other, but they are not identical. It seems to the author that connotative meaning is a result of the interaction of values and the impressions from the word's sounds and notions.

18. Phonetic motivation and phonetic meaning

Theoretical and experimental study of sound symbolism requires distinguishing such notions as "phonetic motivation" and "phonetic meaning". Under phonetic motivation, we understand the correspondence of the sounds of a word and its meaning. The degree of its correspondence is measured with the help of the psycholinguistic methods. Under phonetic meaning (it was introduced by Zhuravlov (1974:60)), we understand the ultimate evaluation of symbolic meanings of sounds which are in the structure of the word. Symbolic meanings of sounds are measured with the help of semantic differential whereas phonetic meaning of a word is to be measured according to a special formula with a special emphasis on the position of sounds in the word.

In such a way, Zhuravlov measured phonetic meaning of thousands of Russian words. Nevertheless, two statements in Zhuravlov's work remain unclear. a) the selection of the analyzed material. Zhuravlov treated such Russian lexemes statistically in which he expected to reveal phonetic meaning beforehand. As a result all the lexemes proved to have symbolic meanings. Furthermore, it remains unclear which types of lexis and what degree PM is available; b) the interpretations of the grades are quite subjective. Here the author does not understand which grades received according to different scales should be regarded as phonetic meaning of a word? In what cases can we consider that the word has phonetic meaning, in which it does not have? Which of three grades can be accepted? If we take all three grades, could we consider that these grades correspond to the denotative meaning of a word? In this case, the methods and interpretations of phonetic meaning require further studies.

The procedure of calculation and interpretation of phonetic meaning is as follows: 514 nonaffixal German words were investigated (157 nouns, 177 adjectives and 180 verbs) (the experiment was conducted by Kushneryk (1986) under the author's supervision). Here the experimenters used Zhuravlov's formula and the data about phonetic meanings of German sounds according to 7 scales, conducted by Kushneryk/Levitskij, 1986. As a result they received the values of phonetic meaning of 514 words with signs "+" or "-". For example, the German word Schub is expressed by +1,34 (strong), -1.12 (slow), -0.13 (cold), etc. Levitskij proposed two basic procedures for objective interpretation of the received values. Firstly, it is necessary to establish the boundaries in the gradation of the degree of a certain feature (i.e. 0,5 and lesser – weal degree, from 0,5 to 1 – average). Secondly, it is possible to find the average value of phonetic meaning and consider that all cases, which enhance this average one, witness the presence of the feature. The calculation has shown that the average value of phonetic meaning is 0,5. These data were received with the help of the psycholinguistic methods (the informants evaluated the degree of correspondence between sounds and the meaning of words with the help of the scale with 5 divisions.

Levitskij considers that there are the following types of the relations between two features characteristic of the dependence between the sounds and meanings – (phonetic meaning and phonetic motivation): 1) Phonetic meaning and phonetic motivation are statistically significant in the lexemes; 2) the word possesses phonetic meaning which is statistically found, but it does not have a positive phonetic meaning (the sounding form does not correspond to the meaning from the informants' point of view); 3) the word has a positive phonetic motivation, but it does not possess statistically significant phonetic meaning (the

values are near to "0"); 4) the word is devoid of significant phonetic meaning and phonetic motivation.

a) The availability of statistically significant PhM and PhM in the word. Forty analyzed words belong to this type whereas adjectives and nouns, denoting movements, actions, properties and qualities and having the lowest frequencies, build up the majority. (The frequencies are determined according to the dictionary Kaednig, 1897).

b) The presence of PhM and the absence of PhM. More than sixty words belong to this type of relation. The have the low frequencies and denote predominantly concrete and abstract notions;

c)The presence of PhM by the absence of PhM. This type includes 140 words whereas the nouns, verbs and adjectives are distributed equally. These words predominantly denote actions, qualities and properties, sounding. Unlike the two previous types, both the most frequent and the least frequent words belong to this category. It is worth noting that a significant number of words are represented by the low stylistic lexis.

d) The cases, in which there is no correspondence of PhM and PhM, are of great interest to the understanding of the mechanism of action and functions of sound symbolism. Here the author asks the question how to interpret the cases when the phonetic meaning is observed (the informants did not "observe" the evident correspondence between the meaning of the word and its sounds. The availability of sounds in the word proved to have a symbolic meaning which is unnecessary for the realization in the speaker's psyche. Consequently, the symbolic meanings are not characteristic of the sounds by nature, but they are ascribed by the speakers according to certain rules. These rules are grounded on the physical properties of sounds (acoustic and articulatory). The effect of synesthesia (which makes up the basis of sound symbolism) has a potential character and can be found under certain conditions: when the sounding form of the word corresponds to its meaning as a result of the influence of internal and external factors of the language development.

19. Phonetic and semantic motivation of words

It is a well-known fact that E. Jespersen (1933) and J. Orr (1944) put the question about the influence of sound symbolism upon the word's life and the anomalies in the action of phonetic laws. J. Orr noticed that Lat. parvus (which includes the sound [a]) contradicts to the meaning small and was replaced in the Romanic languages by the other words, including the sound [i]: Roumanian *mic*, Italian piccolo, Fr. petit. The contrary results were obtained in the experimental psychology. In particular, S. Ertel (1969) compared the "old" and "new" forms (337 pairs) which designated the same notion in five ancient and seven modern languages. This was based on the results of the psycholinguistic experiment (the respondents evaluated the correspondence between sounding form and its meaning). As a result, Ertel found that the "new" words, which were altered by phonetic changes, have lower symbolic properties than earlier forms. Here Levitskij wants to outline that the status of word in the system of language is altered under the influence of two types of motivations – semantic and morphological as well as the other language and non-language factors. The correspondence of phonetic and semantic motivation (the research on the development of the words designated the notion of size in German and English) can be summarized as follows:

In the process of interacting old and new words possessing different types of motivation, it is possible to observe the following relations: 1. One word is devoid of any motivation, the other has only one type of motivation: a) semantic (this relation is observed between the old and new designation of "head" in German *houbit* and *Kopf* (Sperber,

1923:38)); b) phonetic (this relation is made between Lat. parvus and new words in the Romanic languages (Orr, 1944:1-8). 2. One word possesses only phonetic motivation, the other – only semantic (this relation is between Ancient English *lytel* and *smael*. In this case, the competitive struggle depends upon the additional circumstances; 3. One of the two units has one type of motivation, the second – the other type of motivation. Here we distinguish two varieties. The first one happens when the word has semantic motivation, the second – semantic and phonetic (ancient German *small* and *kleini*). In this situation semantically motivated word does not disappear from the language, but goes into the other conceptual sphere. The second variety is observed when one word has phonetic, the other one has phonetic and semantic motivation (Ancient Germ. *luzzil* and *kleini*). In this case, phonetic motivation of the word *luzzil* did not endure the pressure of lexical competitor which had two types of motivation (*kleini*).

The general peculiarities of the development of the word stock (which were established by Sperber and developed by the other linguists) consist in the fact that old, vague and devoid of expressiveness words, starting the competitive struggle with new emotionally expressive words, are replaced by them (by new words). The new units according to the law about the backward correlation and the sign's content turned into demotivated signs. In this situation it is difficult to expect that the number of sound symbolic words in the language will decrease (as in Ertel's experiment). Ertel's results may be caused by the fact the chosen words were standard, normal which have reaches their peak in their development, neglecting their expressive competitors (aa the synonyms to the German word klein: *fipsig, fieselig* or the synonyms to the word Kopf: *Schädel, Birne, Kürbis, Kuller*, etc). Ertel used the famous dictionary by Buck, 1949. If the tendencies towards the decrease of sound symbolic words in the modern languages will be corroborated, then the expressiveness of the lexis at the modern stage of the language development and thinking is caused by the predominance of semantic imagination. In other words, the imagination, which appeared on the basis of notional associations, proved to be more vivid than the imagination caused by symbolic associations.

20. Sound symbolism and some problems of Indo-European studies

The problem of the role of sound symbolism in the lexical word-stock is a part of the Indo-European studies. This problem was considered by A. M. Gazov-Ginzberg (1965), S. V. Voronin (1982), R. Wescott (who concentrated on the correlation between iconicity and origin of the language (1980:3-16; 38-39)). Wescott believes that initially the language consisted of the iconic signs: gradually this iconicity was destroyed by introducing symbols (1980:16). In this case it is difficult to Levitskij to agree on the theory of "lexical poly-genesis" (38-49). Sound symbolism can influence the origin and the semantic development of the word, but the comparison of English dig – dog (the dog < digging a hole) can hardly be taken serious for professional etymologists.

R. Lühr (1988), who dealt with the germination of Germanic consonants, came to the conclusion that the germination in the Germanic languages was caused by different factors and did not have an expressive character. Nevertheless, the nasalization as one of the source of germination serves as "der unmittelbaren und mittelbaren (durch den "Kontext" bedingten) Lautsymbolik" (1988:378). This is illustrated by lexical units of Ancient and modern Germanic languages. R. Lühr grouped the material into certain lexico-semantic subclasses to which she referred sound symbolic words (schwab. *Brienken*, sw. *brienggen* (to weep), but schw. *Briegen*, sw. *brieggen* (to weep) without n); sound imitative words (*monken* "murmur", ahg. *Munken* "to speak quietly); the words meaning movement (Germ. *schunkelen* "to roll" but *schuckelen* has the same meaning); the words with the meaning "to glisten"

(*spranke* "spark" but Il "*sprake*"). The last two examples confirm Levitskij hypothesis that the words with the meaning " to glisten" belong to the roots with the primary meaning "intermittent motion" which resulted from the meaning "to cut". Moreover, Levitskij agrees that the germination pp and bb symbolize something thick, round, massive in the Germanic languages (Lühr, 1988: 276:277).

Bibliography

Brackbill, Y., Little, K. B. (1957). Factors determining the guessing of meaning of foreign words. *Journal of Abnormal Social Psychology 54*, 312-318.

Brown, R. (1958). *Words and Things.* Glencoe, Illinoise: Free Press.

Brown, R., Black, H. A., Horovitz, E. A. (1956). Phonetic Symbolism in Natural Languages. *Journal of Abnormal Social Psychology 50*, 388-389.

Brown, R. Nuttal, R. (1959). Methods in phonetic symbolism experiments. *Journal of Abnormal Social Psychology 59*, 441-445.

Bloomfield, L. (1968). *Jazyk* [Language]. Moscow: Progress (in Russian).

Ertel, S. (1969). *Psycholinguistik. Untersuchung über Lautsymbolik und Motivation.* – Göttingen: Verlag für Psychologie C. J. Hogrefe.

Ertel, S. (1966). Auf dem Glatteis der Methodik in Experimenten zur Lautsymbolik Erwiderung zur Kritik Taylors und Kritik an Taylor. *Zeitschrift für experimentelle und angewandte Psychologie 13*, 110-132.

Ertel, S. (1965). Der Lautcharakter künstlicher Lautgebilde. *Psychologische Forschung 28*, 491-518.

Ertel, S. (1966). Ist Lautcharakter sprachspezifisch oder universell? *Studia Psychologica 8:3*, 216-230.

Ervin-Tripp, S., Slobin, D. (1966). Review of psycholinguistics". *Annual Review of Psychology 17*, 435 – 474.

Fenz, E. (1940). *Laut, Wort, Sprache und Ihre Deutung. Grundlegung einer Lautbedeutungslehre.* Wien: Deuticke.

Fischer-Jorgensen, E. (1978). On the universal character of phonetic symbolism with special reference to vowels, *Studia Linguistica 32: 1-2*, 80-90.

Gazov-Ginzberg, A. M. (1965). *Byl li jaz'k izobraziteln'j v svoih istokah?* [Was the language symbolic in its origins] (in Russian), M.: Nauka.

Hurdzujeva, E. (1973). Elementarnij zvykovoj simvolizm (statisticheskoje issledovanije) [Elementary sound symbolism (statistical research)]. Autoreferat diss. 10.0204, Moscow, 31.

Jakobson, R., Waugh, R. Linda. (2002). *The Sound Shape of Language*. Berlin: Walter de Gruyter..

Jespersen, O. (1933). Symbolic value of the sound /i/", *Linguistica*. Copenhagen, 283-303.

Kanungo, R. Lambert, W.E. (1963). Semantic satiation and meaningfulness. *American Journal of Psychology*, 76:3.

Kojbajeva T. Ch. (1987). Zvykosimvolicheskaja leksika anglijskogo i osetinskogo [Sound symbolic lexis of the English and Ocetian languages]: AKD (in Russian)

Komarnytska, L.A. (1985). Subjektivnij i objektivnij zvykovoj simvolizm v anglijskom jaz'ke [Subjective and objective sound symbolism in the English language], Odessa: AKD,

Kronasser, H. (1952). *Handbuch der Semasiologie*. Heidelberg: J. Groos Verlag.

Kushneryk, V.I. (1987). Foneticheskoje znachenije i foneticheskaja motivirovannost v sovremennom nemetskom jazyke [Phonetic meaning and phonetic motivation in the modern German language], Odessa: AKD.

Kushneryk, V.I, Levitskij, V.V. (1986). Symvolichni znachenna holosnyh I pryholosnyh suchasnoji nimetskoji movu" [Symbolic meanings of vowels and consonants of the modern German language]. *Movoznavstvo,* 3 – 21-26 (in Ukrainian).

Leont'ev, A.A. (1967). *Psicholingvistika* [Psycholinguistics], Leningrad: Nauka.

Levitskij, V.V. (2009). *Zvukovoj simvolizm: Mify i real'nost* [Sound symbolism: myths and reality], Chernivtsi National University, 263 (in Russian).

Levickij, V.V. (1973). Semantika i fonetika. Posobije, podgotovlennoje na mateiale eksperimentalnch issledovanij [Semantics and phonetics. The book is based on the material of the experimental research], Chernovtsi: Izd-vo Chernov. University (in Russian).

Levickij, V.V. Sternin, I. A. (1989). Eksperimental'nje metod' v semasiologii [Experimental methods in semasiology]. Voronezh: Izd-vo Voronezh. Universiteta (in Russian).

Levitskii V., Naidesh, O. (2011). The Phonosemantic Properties of a Poetic Text. In: B.P. Sherr, J. Bailey, E.V. Kazartsev (eds.). *Formal Methods in Poetics. A Collection of Scholary Works Dedicated to the Memory of Professor M.A. Krasnoperova:* 227-235. Lüdenscheid: RAM-Verlag.

Lühr, R. (1988). *Expressivität und Lautgesetz im Germanischen Heidelberg.* Monographien zur Sprachwissenschaft.

Lvova, N. (2005). Semantic functions of English initial clusters, *Glottometrics* 9, 21-27.

Mazanajev, I.A. (1985). *Osnovnije grupp' zvukosimvolicheskih slov: fonosemanticheskij analiz (na material anglijskogo i lezginskogo jaz'kov)* [Basic groups of sound symbolic words: phonosemantic analysis (on the basis of English and Lesgian languages], Leningrad.

Magnus, M. (2001). *What's in A Word? Studies in Phonosemantics*. Norwegian University of Science and Technology.

Miron, M. A. (1961). A cross-linguistic investigation of phonetic symbolism, *Journal of Abnormal Social Psychology,* 62 : 3, 623 – 630.

Müller, H. (1935). *Experimentelle Beitrage zur Analyse des Verhaltnisses von Laut und Sinn*. Potsdam: Berlin: Muller I. Kiepenheuer Verlag.

Newman, S. (1933). Further experiments in phonetic symbolism, *American Journal of Psychololgy. 45*, 53 – 75.

Ohala John J., Hinton L., Nichols J. (1994). *Sound Symbolism*. Cambridge: Cambridge University Press.

Orr, J. (1944). On some sound values in English, *British Journal of Psychology,* 35 : 1, 1 – 8.

Pavliuk, N.A. (1976). Sopostavlenije leksicheskogo i foneticheskogo znachenija. Problem' motivirovannosti jaz'kovogo znaka [The comparison of lexical and phonetic meaning. The problem of motivation of the language sign]. Kaliningrad, 66-73 (in Russian).

Peterfalvi, J-M. (1970). *Recherches experimentales sur le symbolism phonetique. Monographies francaise de psychologie 21*. Paris: Centre National de la Recherche Scientifique.

Pokorny, J. (1959-1969). *Indogermanisches etymologisches Wörterbuch*. Bd 1-2. ß Bern: München: Francke Verlag.

Polukarova, T.A. (1989). *Foneticheskaja znachimost anglijskih monoftongov: Problem fonosemantiki*. Moscow 1989 – 54-55 (in Russian).

Sapir, E. (1929). A Study in Phonetic Symbolism, *Journal of Experimental Psychology 12:3*, 225 – 239.

Schaff, A. (1963). *Vvedenije v semantiku*. [Introduction to semantics]. Izd-vo inostrannoj literatyry, 376. (in Russian)

Sieberer, A. (1947). Primäre oder sekundäre Lautbedeutungssamkeit? *Anzeiger der Wissenschaften, philosophisch-historische Klasse* 9, 35-52.

Smithers, G.V. (1954). Some English ideophones, *Archivum linguisticum* 6:2, 73 – 111.

Sperber, H. (1923). *Einführung in die Bedeutungslehre*. Schröder Verlag.

Tambovtsev, J.A. (2007). Prokofjeva. Zvukovo-tsvetovaja assotsiativnost: universalnoje, natsionalnoje, individyalnoje. [Prokofjeva. Sound-colour associations: universal, national, individual]. Saratov: Izd-vo Saratovskogo universiteta (in Russian).

Tarte, R., Barrit, L. (1971). Phonetic symbolism in adult native speakers of English, Three Studies, *Language and Speech* 76 : 2, 231 – 239.

Taylor, I.K. Taylor, M.M. (1962). Phonetic symbolism in four unrelated languages, *Canadian Journal of Psychology 16, 2*. – 344-356.

Taylor, I.K. (1963). Phonetic symbolism Re-examined. *Psychological Bulletin 60:2*, 200-209.

Thorndike, E. (1945). On Orr's hypotheses concerning the front and back vowels. *British Journal of Psychology 36:1*, 10-13.

Tsuru, S. Fries, H. (1953). A problem in meaning. *Journal of General Psychology* 8:281-284.

Ullmann, Stephen. (1964). *Semantics: An introduction to the science of meaning.* Oxford: Basil Brackwell.

Ultan, R. (1978) Size sound symbolism. In: Joseph Greenberg, (ed). *Universals of Human Language 2: Phonology: 527-568.* Stanford University Press,.

Wandruzhka, M. (1952). Der Streit um die Deutung der Sprachlaute. Tübingen: Niemeyer, 214-227.

Weiss, J. (1968). Phonetic symbolism and perception of connotative meaning. *Journal of verbal learning and verbal behavior 7*, 574-576.

Werner, H. (1932). *Grundlagen der Sprachphysiognomik*. Leipzig: Verlag von Velhagen u. Klasing.

Wescott, R. (1980). *Sound and Sense: linguistic essays on phonosemic subjects*. Lake Bluff, IL: Jupiter.

Voronin, S. V. (1982). *Osnovy fonosemantiki*. [Fundamentals of phonosemantics] L.: Izd-vo Leningrad. Universiteta, 242 (in Russian).

Wertheimer, M. (1958). The relation between the sound of a word and its meaning. *American Journal of Psychology* 71, 412 – 415.

Zhernovej, A.N., Levitskij V.V. (1988). Nachal'nje sochetnija fonem v nemetskom jaz'ke. Psycholingvisticheskie issledovanija znachenij slova i ponimanija texta [„Initial combinations of phonemes in German]. In: Psiholingvisticheskie issledovanija znachenija slova I ponimanija texta . [*Psycholinguistic research of the meanings of words and the comprehension of the text]*, 124-132. Kalinin (in Russian).

Zhuravlov, A. P. (1974). *Foneticheskoje znachenije*. [Phonetic meaning] Leningrad University Press (in Russian).

Glottometrics 31, 2015, 31-37

A continuous model for polysemy

Emmerich Kelih
Gabriel Altmann

Abstract. In the article it will be shown that modeling polysemy in language can be performed in a rather unified way.

Keywords: diversification, polysemy, Zipf-Alekseev function, exponential function

Polysemy is a special case of diversification which exists in all domains of language and is one of the steady evolution processes of language. According to the given domain and type of language, one can model this phenomenon in quite different ways (cf. e.g. Altmann 1985, Altmann, Best, Kind 1987, Rothe 1991 and newer developments). There are such a number of boundary conditions that different models may be attained by simply changing the condition in the underlying difference or differential equations. However, the aim of the research is to restrict the analysis to a special domain and strive for a unified model – if possible.

It is usual to model the distribution of polysemy by means of a discrete approach, e.g. by means of a Poissonian birth-and-death process, because one may suppose that the individual meanings are discrete, at least in the dictionary; they are born, live and die. The fact that each meaning is rather a multidimensional space with a number of degrees, or still better, a continuous spectrum capturing a sector of the reality, may be ignored in order to show that there is some law behind polysemy. Hence one may consider polysemy also a continuous variable and the number of meanings of the word (x = 1,2,3,…) as averages or lower/upper bounds of an interval. In that case one can propose a continuous function without normalization. Such a step does not change anything in our demands for "truth" because no mathematical model expresses the truth, it is merely a means for capturing a phenomenon in such a way that we understand it, can insert it in a background theory and link it with other phenomena. Besides, there are no discrete or continuous phenomena in the reality and what more, discrete models can be transformed in continuous ones and vice versa (cf. Mačutek, Altmann 2007).

Here we shall restrict ourselves to semantic diversification or polysemy and use only published data. We lean against the unified theory (cf. Wimmer, Altmann 2005), i.e. we start with the simple differential equation

$$\frac{dy}{y} = \frac{g(x)}{h(x)}dx, \tag{1}$$

where $g(x)$ represents a function associated with the topical norm in language and the influence of the speaker, $h(x)$ is the control of the community/hearer striving for equilibrium, y is the number of cases with polysemy x. Since we search for a continuous model, we may omit all polysemies whose number is zero and adapt (1) to

$$\frac{dy}{y-1} = \frac{g(x)}{h(x)}dx. \tag{2}$$

The right hand side of the differential equation can consist of $g(x) = A + B*logX$ where A is the constant of the language, B is the force of the speaker influencing not directly X but only its logarithm, and $h(x) = CX$ is the equilibrating force of the speaker or the community. Solving the differential equation and reparametrizing it we obtain the well know Zipf-Alekseev function used already in different domains of linguistics (cf. e.g. unit length in Popescu, Best, Altmann 2014). The logarithm is probably associated with the human perception. The resulting function is

$$(3) \qquad y = cx^{a + b \log x}$$

or, omitting classes having zero occurrence and striving for theoretical values greater than or equal to 1,

$$(4) \qquad y = 1 + cx^{a + b \log x},$$

which yields always values great than or equal to 1. Function (3) admits computed values smaller than 1 but mostly the fitting is satisfactory. The results of fitting the above functions to classes or the whole dictionary in 6 languages are presented in Table 1. The data may be found in the quoted references.

Table 1a
Semantic diversification of word classes

Source	**a**	**b**	**c**	**R^2**
German verbs (Levickij, Kiiko, Spolnicka (1996) (4)	-2.1665	-1.4556	9878.1011	0.9998
German verbs: Texts (Levickij, Drebet, Kiiko 1999) (4)	1.0655	-1.2826	146.1013	0.9948
German nouns (Levickij, Drebet, Kiiko 1999) (4)	-0.0636	-1.0846	1572.1306	0.9982
German nouns; Texts (Levickij, Drebet, Kiiko 1999) (4)	1.0894	-1.0973	160.0230	0.9874
German nouns (Schierholz 1991) (3)	-0.8306	-0.3526	2599.7945	0.9997
German adjectives (Levickij, Drebet, Kiiko 1999) (4)	-0.1812	-0.8249	212.1173	0.9993
German adjectives: Texts (Levickij, Drebet, Kiiko 1999) (4)	1.9881	-1.3448	20.4406	0.9883
Russian 11-14th (Andreevskaja 1990) (3)	-2.2456	0.0073	1681.2993	0.9996
Russian 18th (Andreevskaja 1990) (4)	-1.8601	-1.1123	2813.0288	1.0000
Russian 19th (Andreevskaja 1990) (4)	-2.1491	-0.7644	3383.9869	1.0000
Russian 20th (Andreevskaja 1990) (4)	-1.8760	-0.9490	3396.9807	1.0000
Russian all (Andreevskaja 1990) (4)	-2.1535	-0.5548	13597.9953	1.0000
English adjectives (Višnjakova 1976) (4)	-0.2816	-0.6839	2982.4149	0.9999
English adverbs (Višnjakova 1976) (3)	-0.6651	-1.6055	204.0278	0.9987
English verbs (Višnjakova 1976) (3)	0.0627	-0.5414	746.3902	0.9937
Russian verbs SO (Krylov, Jakubovskaja 1977, Krylov 1982) (4)	-0.8109	-1.1528	7041.4818	1.0000
Russian nouns SO (Krylov, Jakubovskaja	-1.2201	-1.1964	12521.3205	1.0000

1977, Krylov 1982) (4)				
Polish adjectives (Hammerl 1991) (4)	-2.3388	-0.5356	7206.9255	1.0000
Polish adverbs (Hammerl 1991) (4)	-1.9981	-0.7359	1112.9775	1.0000
Polish nouns (Hammerl 1991) (4)	-0.8879	-1.0597	8800.3470	1.0000
Polish verbs (Hammerl 1991) (4)	-0.2091	-1.0624	3130.4044	0.9999
Polish total (Hammerl 1991) (3)	-1.2500	-0.8025	20353.6346	1.0000
Hungarian all (Papp 1976) (3)	-0.2164	-1.0185	30220.1644	1.0000
Maori words (Wimmer, Altmann 1999) (3)	-1.1965	-0.7329	5109.3108	0.9999

Some data can be better captured by the exponential function in whose differential equation there is no interaction of forces [g(x)/h(x)] but merely a proportionality constant. Setting $dy/y = -bdx$ we obtain

$$(5) \qquad y = a\exp(-bx)$$

or with omitting the zeroes and values smaller than 1

$$(6) \qquad y = 1 + a\exp(-bx).$$

Some of the data could be better captured by (5) or (6). They are presented in Table 1b. This may depend on the source, on severe self-regulation, on the way of preparing the data, on definitions, etc. As a matter of fact, it is preliminarily impossible to give reasons for the choice of the simple or the logarithmic influence.

Table 1b
Exponential fitting

Source	**a**	**b**	**R^2**
Russian 15-17th (Andreevskaja 1990) (6)	12063.6572	-1.6494	0.9994
English nouns (Višnjakova 1976) (6)	21111.8336	-0.8962	0.9961
Japanese postpositions T1 (Sanada, Altmann 2009) (5)	3361.1378	-0.2181	0.9852
Japanese postpositions T2 (Sanada, Altmann 2009) (6)	5354.7916	-0.1670	0.9785
Japanese postpositions T3 (Sanada, Altmann 2009) (6)	4642.9967	-0.1890	0.9592
English all (HO) (Polikarpov 1987) (6)	259405.8320	-1.9698	0.9984
English all (SHO) (Polikarpov 1987) (6)	116312.1870	-0.9364	0.9972

In Russian the complete dictionary has been analyzed by Andreevskaja for individual centuries. Only the data from 15-17th century displayed simple exponential relation. SO: „Slovar Ožegova“; MAS „Slovar' russkogo jazyka" ed. by A.P. Evgen'eva.

In Japanese 3 different texts were used.

English: HO: „Hornby: Oxford Advanced Learner's Dictionary of Current English“; SHO: „Shorter Oxford English Dictionary“.

As can be seen, both texts and dictionaries can be analyzed. In dictionaries one relies on the completeness of the source, in texts one takes only the locally given meanings and

sums them. But one can use a special class of units to distinguish text sorts or to study the development of a writer.

Individual words and morphemes

The second aspect is the polysemy analysis of an individual unit, e.g. a special word or morpheme. It is well known that synsemantics have many meanings while autosemantics are semantically rather restricted. If one considers the occurrence of individual meanings of a unit in texts, one obtains the usual rank-frequency distribution. This can be characterized in many different ways both synchronically and historically, in individual texts and in corpus, in individual writers, text-sorts and languages. Here we have to do with the auxiliary variable represented by ranks which allows us to observe some regularity. The results of the analysis are displayed in Table 2.

In Polish, the data have been taken from two dictionaries: SPP = Słownik poprawnej polszczyny ed. by W. Doroszewski, H. Kurkowska. Warszawa 1973, and MS = Mały słownik języka polskiego ed. by S. Skorupka, H. Auderska, Z. Łempicka. Warszawa 1968.

Table 2
Rank-frequency dependence in the polysemy of individual units

Source	**a**	**b**	**c**	**R^2**
German “von” as particle (Best 1991) (4)	-0.5042	-0.1129	52.4838	0.9794
German “von” as preposition with adverbial meaning (Best 1991) (3)	-0.2850	-0.2449	9.7582	0.9542
German “von” as preposition in prepositional object (Best 1991) (3)	-0.6487	-0.2694	37.8958	0.9866
German “von” as preposition in attributive phrase (Best 1991) (3)	0.0560	-0.3324	21.1295	0.9836
German “auf” in Text 1 (Fuchs 1991) (4)	-0.6643	-0.2037	22.6680	0.9756
German „auf“ as preposition in attribute Text 1 (Fuchs 1991) (3)	-0.5787	-0.2214	23.6084	0.9644
German “auf” as preposition in Attribute Text 2 (Fuchs 1991) (3)	-0.9508	-0.2561	29.8151	0.9834
German “auf” in fixed phrases T2 (Fuchs 1991) (4)	0.5507	-0.8377	24.9964	0.9824
Polish „w“ SPP (Hammerl, Sambor 1991) (3)	-0.8759	-0.2234	199.3127	0.9967
Polish “w” MS (Hammerl, Sambor 1991) (3)	-2.3687	0.3164	298.1104	0.9913
English “in” (Hennern 1991) (4)	-0.1864	-0.4540	52.1876	0.9304
Slovak “do-“ (Nemcová 1991) (3)	1.4539	-2.3136	22.0245	0.9943
Slovak “na-“ (Nemcová 1991) (3)	-0.1080	-0.3380	29.8423	0.9738
Slovak “o-“ (Nemcová 1991) (3)	0.4163	-0.6163	16.1645	0.8528
Slovak “od-“ (Nemcová 1991) (3)	-2.0774	0.5776	25.0148	0.9996
Slovak “po-“ (Nemcová 1991) (3)	0.2424	-0.6729	59.5296	0.9651
Slovak “pre-“ (Nemcová 1991) (3)	-0.9338	-0.0503	32.9225	0.9848
Slovak “roz-“ (Nemcová 1991) (3)	0.1549	-0.1905	25.9347	0.9502

Slovak "s/z-" (Nemcová 1991) (3)	-1.2505	-0.0358	70.8579	0.9937
Slovak "u-" (Nemcová 1991) (3)	0.3750	-1.0889	55.4833	0.8596
Slovak "vy-" (Nemcová 1991) (3)	-0.2845	-0.1828	61.1131	0.9781
Slovak "za-" (Nemcová 1991) (3)	-0.7437	-0.5129	76.9220	0.9951
Japanese "ni" (Roos 1991) (3)	0.3600	-0.5806	38.4638	0.9565
Hungarian "föl-" (Beöthy, Altmann 1984) (3)	-0.3716	-0.0950	10.8001	0.9602
Hungarian "el-" (Beöthy, Altmann 1984) (4)	-3.4680	0.1526	82.0003	0.9999
Hungarian "be-" (Beöthy, Altmann 1984) (3)	-0.4796	-0.2510	19.7244	0.9822
Hungarian "ki-" (Beöthy, Altmann 1984a) (3)	0.0749	-0.3637	11.8491	0.9760
Hungarian "meg-" (Beöthy, Altmann 1991) (3)	-4.0127	1.0718	106.9028	0.9956

Table 2b
Exponential y = a*exp(b*x)

Source	**a**	**b**	**R^2**
German „auf" in Text 2 (Fuchs 1991) (6)	497.9016	-0.5134	0.9493
French „et" (Le Petit Prince) (Rothe 1986) (6)	18.4498	-0.1515	0.9455

As can be seen, one obtains good fittings in all cases. Now, we may conjecture that self-regulation which is present in the model of the Zipf-Alekseev formula leads to some dependence between the parameters *a* and *b*. There are two possibilities of finding it: (1) One unites all data or (2) one considers languages/units separately. Though we conjecture that the greater is *a* representing the norms in the language, the smaller should be *b*, hence $b = f(a)$ must be a decreasing function, either linear or non-linear. Unfortunately, we do not have enough cases for all aspects, for example, not all parts of speech, not enough texts for a certain unit (e.g. a preposition), etc., hence we may merely observe that Slovak or Hungarian prefixes, Polish parts of speech, etc. have the above mentioned decreasing tendency. Uniting all results, we obtain a decreasing tendency but the oscillation is too great. Many further systematic analyses of individual phenomena must be performed to find the $b = f(a)$.

In cases where there are more *homogeneous* data, one can easily state that $b = f(a)$ is a function of the type $y = c + d*exp(x)$ but this can be shown only if there will be sufficient amount of data.

The problem whether other diversification phenomena can be modeled in the above mentioned way remains a task for the future.

References

Altmann, G. (1985). Semantische Diversifikation. *Folia Linguistica 19, 177-200.*

Altmann, G., Best, K.-H., Kind, B. (1987). Eine Verallgemeinerung des Gesetzes der semantischen Diversifikation. In: Fickermann, I. (ed.), *Glottometrika 8, 130-139.* Bochum: Brockmeyer.

Altmann, G. (1991). Modelling diversification phenomena in language. In: Rothe, U. (ed.), *Diversification processes in language: Grammar: 33-46.* Hagen: Rottmann.

Andreevskaja, A.V. (1990). Kvantitativnoe issledovanie polisemii kornevych slov russkogo jazyka XI-XX vekov. In: *Kvantitativnaja lingvistika i avtomatičeskij analiz tekstov 6,* 3-11. [= Učenye zapiski tartuskogo gosudarstvennogo univer siteta, 912]

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

Beöthy, E., Altmann, G. (1984a). The diversification of meaning 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 1. "meg-". In: Rothe, U. (ed.), *Diversification processes in language: Grammar: 60-66.* Hagen: Rottmann.

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

Fuchs, A. (1991). Diversifikation der Präposition *auf*. In: Rothe, U. (ed.), *Diversification processes in language: Grammar: 105-115.* Hagen: Rottmann.

Hammerl, R. (1991). *Untersuchungen zur Struktur der Lexik: Aufbau eines lexikalischen Basismodells.* Trier: Wissenschaftlicher Verlag Trier.

Hammerl, R., Sambor, J. (1991). Untersuchungen zum Beöthy-Gesetz im Polnischen. In: Rothe, U. (ed.), *Diversification processes in language: Grammar: 127-137.* Hagen: Rottmann.

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

Krylov, Ju.K. (1982). Ob odnoj paradigme lingvističeskich raspredelenij. In: *Trudy po lingvostatistike 8: Lingvostatistika i vyčislitеč'naja lingvistika,* 80-102. [= Učenye zapiski Tartuskogo Gosudarstvennogo Universiteta, 628]

Krylov, Ju.K., Jakubovskaja, M.D. (1977). Statističeskij analiz polisemii kak jazykovoj universalii i problema semantičeskogo toždestva slova. *Naučno-techničeskaja informacija, Serija 2,* 3, 1-6.

Levickij, 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.*

Levickij, V.V., Kiiko, J.J., Spolnicka, S.V. (1996). Quantitative analysis of verb polysemy in modern German. *Journal of Quantitative Linguistics 3(2), 132- 135.*

Mačutek, J., Altmann, G. (2007). Discrete and continuous modelling in quantitative linguistics. *Journal of Quantitative Linguistics 14(1), 81-94.*

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

Papp, F. (1967). O nektorych količestvennych charakteristikach slovarnogo sostava jazyka, *Slavica 7, 51-58.*

Polikarpov, A.A. (1987). Polisemija: sistemno-kvantitativnye aspekty. *Kvantita tivnaja lingvistika i avtomatičeskij analiz tekstov 3, 135-154.* [= Učenye zapiski tartuskogo gosudarstvennogo universiteta, 774]

Popescu, I.-I., Altmann, G. (2008). On the regularity of diversification in language. *Glottometrics 17, 94-108.*

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

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

Rothe, U. (ed.) (1991). *Diversification processes in language: Grammar*. Hagen: Rottmann.

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

Schierholz, S. (1991). *Abstraktheit, Häufigkeit und Polysemie deutscher Substantive*. Tübingen: Niemeyer.

Višnjakova, S.M. (1976). Opyt statističeskogo issledovanija mnogoznačnosti slov v anglijskom jazyke. In: Guseva, E.K., Andrjuščenko, V.M., Revzin, I.I. (eds.), *Vyčislitel'naja lingvistika: 168-178.* Moskva: Nauka.

Wimmer, G., Altmann, G. (1999). Rozdelenie polysémie v maorijčine. In: Genzor, J., Ondrejovič, S. (eds.), *Pange lingua. Zborník na počesť Viktora Krupu: 17-24.* Bratislava: Veda.

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

Glottometrics 31, 2015, 38-49

Anglicisms in the Austrian Newspaper *KLEINE ZEITUNG*

Hanna Gnatchuk

Abstract. The given article deals with the study of anglicisms in the Carinthian newspaper *Kleine Zeitung*. The purpose of this research is twofold: a) we are intended to observe the development of anglicisms in the analyzed newspaper from 1995 till 2015, b) we are aimed at analyzing the structure and cohesion of English-German as well as German-English hybrid compounds. In order to do it, one issue of *Kleine Zeitung* for each year (1995 till 2015) was analyzed. In such a way, 20 newspapers make up the corpus of our study. The results are processed with the help of such statistical methods as *Piotrowski law*, power function to the ranking of compounds, rank frequency distribution, Zipf-Alekseev function.

Key words: borrowings, anglicisms, americanisms, language contact, hybrid (mixed) compounds.

1. Introduction

The process of borrowing the words is characteristic of all languages. The borrowing of lexemes from one language into another is a result of language contacts in its history. The enrichment of the vocabulary is an inevitable process in the history of any language. Its main task is to satisfy the demands of the society in the communication at each historical stage. Focusing on the German word-stock, it is necessary to mention that the enrichment of its vocabulary was initially represented by Latin and French words. Nevertheless, the English borrowings gave way into German in the 19th century and their impact is quite evident even nowadays.

In order to study the borrowings in a proper way, it is relevant to draw a distinction between their linguistic and extralinguistic causes. On the whole, linguistic causes can be described in the following ways: 1) the absence of the equivalent word in the native language in order to designate a new object or notion; 2) the tendency to the economy of language means (in particular, the tendency to the usage of one borrowed word instead of the whole phrase or a word combination); 3) the intention to preserve and enhance the communicative accuracy (which helps to elucidate polysemy or homonymy in the recipient language); 4) the tendency to expressiveness which leads to the appearance of foreign synonyms; 5) the absence in the native language the possibility of forming derivatives (whereas it is possible to form them from the borrowed synonyms).

The extra-linguistic causes are represented by the following ones: 1) the cultural impact of one nation on the other; 2) the presence of oral or written contacts of the countries with different languages; 3) the increasing interest in learning the language; 4) the prestige of the language which sometimes results in the appearance of internationalisms; 5) the admiration by certain social strata of the culture of a foreign country.

As far as the English borrowings are concerned, it is relevant to differentiate the following variants of English: British, Canadian, American, Australian, New Zealand, South African). It is worth mentioning that the American English started its way into German in the second half the twentieth century. Therefore, it is also possible to come across such terms as anglicisms or americanisms. The second half of the twentieth century is famous for spreading the American variant of the English language. This provided a favourable background for anglo-americanisms in the German language.

The number of anglicisms in all spheres is constantly increasing (computer technology, sport, economy, etc). One of the first linguists was H. Zindler (1959, 1975) who gave the most appropriate definition of anglicisms. From his point of view, the anglicisms are not only the borrowings from the British or American English. They are also represented by some changes of meaning of German words or their contextual usage according to American or British example. In such a way, the term „anglicism" is predominantly used by the contemporary researchers in order to designate the borrowings from English.

Focusing on the increase of English borrowings in German, K. Heller (1965, 1966) found that 148 anglicisms existed in 1842; 392 English lexical units – in 1899; 900 – in 1909. B. Carstensen (1965, 1980) did the research in three German newspapers with the following results: 1910 – 19 anglo-americanisms, 1930 – 14 English lexemes.

K-H. Best (2002) was engaged with the research of the suffix –ical. He has found a systematic increase of word forms with the suffix *–ical* (as in the word *musical)*. The researcher also dealt with the Arabian borrowings and their increase since the 14th century (Best, 2004). Special attention was also paid to the investigation of Turkic words where a certain increase of Turkic borrowings was observed in the German language (Best, 2005). As far as Iranian and Hebrew words are concerned, it is possible to find ca. 150 borrowings in German (Best, 2013).

According to H. Paul (1920), it is relevant to take into account the proponents and opponents' opinion concerning English borrowings. The proponents take the view that the anglicisms have the following advantages:

a) A structural advantage deals with the compact form of anglicisms in comparison with the proposed German substitutes;
b) Economic advantage helps to simplify (to make it easy) the communication in all the spheres of the outer world;
c) Cultural advantage consists in the fact that English gives the opportunity to participate in the world communicative culture (Internet, the "youth" language).

The opponents of anglicisms consider that:

a) The German language loses its previous power;
b) It stops to be prestigious;
c) The Germans do not appreciate the values of the German language.

In such a way, the anglicisms are more active in advertisements, computer and innovation technology, media, economy, leisure, fitness, fashion, the youth culture (including pop-culture). These spheres have a profound impact on the social consciousness. The lexis and the terms of the language are formed in these spheres by expanding oral speech.

2. The investigation of anglicisms in a Carinthian newspaper *KLEINE ZEITUNG* (1995-2015)

In order to study the anglicisms, it would be relevant to explain what a *word* is. On the whole, the word is a sequence of morphemes united by grammatical rules of a certain language. According to V. Levitskij (2012) the word (as a central and a basic language unit) is connected with the units of different levels: phonemes, morphemes, on the one hand, and word combinations and sentences, on the other hand. As far as the connection of a word or a lexeme is concerned, this aspect is predominantly studied in the quantitative linguistics with a special emphasis on the correlation of a word's length and its semantics or the research of sound symbolism. Dwelling upon the relation of a lexeme and morphology, it is relevant to

mention that affixes are intended to express the relation between the subjects and the phenomena of the outer reality (which are designated by the words). Nevertheless, some attempts were undertaken by J. D. Apresjan (1962) in order to establish the isomorphic relation between syntactic constructions and lexico-semantic variants of a word. But the relation between lexemes and syntax is not studied enough (Levitskij, 2012:342).

The purpose of the research is to observe how the anglicisms developed in the Carinthian newspapers during 1995 till 2015 using Piotrowski law. In such a way, we are intended to investigate the development of English borrowings in one of Austrian lands – Carinthia.

The material of the research consists of 20 Carinthian newspapers (*Kleine Zeitung*) from 1995 till 2015. We have analyzed one issue per year (in our case, the newspaper issue for January) by writing out all new anglicisms appearing in each newspaper issue. Moreover, we have analyzed 50 pages of the analyzed newspaper in so far as the number of pages varied during that period to the high extent. The analyzed rubrics were *Politics, Economy, Sport, Culture, TV and People*. Nevertheless, it should be marked here that under new anglicisms the author of the article understands the English borrowed words which did not occur in the previous analysed newspapers (for example, the English word, the Anglicism "fitness" was found in issue 1996 (January) whereas the analysed word was not found in the previous issue (January, 1995). In such a way, the word "fitness" is considered new anglicisms in 1996).

The discussion of the results. 1214 anglicisms have been found in a Carinthian newspaper *Kleine Zeitung* during 1995 – 2015. The results are given in Table 1.

Table 1
The total number of new anglicisms in a Carinthian newspaper
"Kleine Zeitung" (1995 – 2015)

Year of issue	***Number***
1995	49
1996	38
1997	71
1998	46
1999	37
2000	55
2001	**109**
2002	68
2003	65
2004	79
2005	71
2006	69
2007	87
2008	54
2009	51
2010	54
2011	59
2012	50

2013	27
2014	50
2015	25

Judging from Table 1, it is possible to observe that the period from 2001 till 2008 has the highest number of anglicisms. The issue of 2001 turned out to have 109 English borrowings (the highest value in comparison with all issues). However, a statement of this kind must be tested statistically. If we take into account that newspapers depend on outer circumstances and each article is written in different domain by possibly different authors, we may expect that there will be great non-homogeneity in borrowing anglicisms. Let us first perform the test for homogeneity of all years using the chi-square test. The numbers of all anglicisms used in individual years are presented in Table 2.

Table 2
All anglicisms used in the newspaper "Kleine Zeitung"

Year of issue	***Number***
1995	49
1996	45
1997	78
1998	54
1999	43
2000	72
2001	**124**
2002	78
2003	89
2004	98
2005	93
2006	86
2007	76
2008	79
2009	62
2010	71
2011	75
2012	56
2013	27
2014	67
2015	43

The expected number for each cell is E = 1465/21 = 69.7619. The test for homogeneity is defined as

$$(1) \qquad X^2 = \sum_i \frac{(f_i - E)^2}{E}$$

yielding X^2 = 143,60 which is with 20 DF highly significant. Hence the use of anglicism in the individual years is not homogeneous.

One can set up separate hypotheses, e.g. conjecturing that the years 2002 to 2008 are homogeneous. In order to test it, one computes *E* as the sum of anglicisms in these years divided by 7 and obtains *E* = 599/7 = 85.5714. Inserting this value in (1) and fit for the given years, one obtains X^2 = 4.8347 which is, with 6 *DF* not significant. Hence these years are really homogeneous.

Adding the individual numbers in the second column of Table 1 to obtain cumulative frequencies and rescaling the independent variable "year" simply in 1,2,3,…, we obtain a different image of the acquisition of new Anglicism. The result is presented in Table 1a. As is usual, one can fit the cumulative data using the Piotrowski approach (c.f. Altmann et al. 1983; Altmann 1983, Best, Beöthy, Altmann 1990; Leopold 2005) in some of its forms. Here we shall use the simplest variant

$$(2) \qquad y = \frac{1}{c + a \bullet \exp(-bx)} \ ,$$

and obtain the results in the fourth column of Table 3.

Table 3
Fitting the Piotrowski law to the increase of anglicisms in German

Year	**New anglicisms**	**Cumulative**	**Computed**
1	49	49	100.78
2	38	87	127.85
3	71	158	161.19
4	46	204	201.67
5	37	241	250.02
6	55	296	306.60
7	**109**	405	371.28
8	68	473	443.29
9	65	538	521.08
10	79	617	602.49
11	71	688	684.85
12	69	757	765.41
13	87	844	841.63
14	54	898	911.51
15	51	949	973.75
16	54	1003	1027.78
17	59	1062	1073.64

18	50	1112	1111.84
19	27	1139	1143.15
20	50	1189	1168.48
21	25	1214	1188.76
a = 0.011860, b = 0.261538, c = 0.00079, R^2 = 0.9968			

In such a way, it is possible to summarize the above-mentioned results:

- The tendency to the usage of NEW anglicisms is observed in each year (in our case, from 1995 till 2015) of the Carinthian newspaper "Kleine Zeitung";
- The usage of anglicisms varies in individual years to a certain extent;
- We set out the hypothesis that the years from 2002 till 2008 are homogeneous which was statistically confirmed.

3. The investigation of English-German (German-English) hybrid compounds

It is worth mentioning here that German compounds are extremely productive in order to enrich the lexical word-stock. A careful attention should be drawn here to the fact that the compounds can be of two types: 1. The compounds consisting of the basic and the determinative words of English origin (*das Skateboard, die Lovestory, der Ladykiller, der Eurocity)*; 2. The compounds of mixed type consist of one lexeme which is of English origin, the other – German (*der Fitnessraum, die Chartermaschine, das Recyclingpapier, die Teenagersprache).*

The aim of the analysis is to find the structural patterns of English-German (or German-English) hybrid compounds in the Austrian newspaper "Kleine Zeitung".

The material of the research consists of 20 newspapers from 1995 till 2015. 483 English-German as well as German-English hybrid compounds have been found. In such a way, we have found the following 27 patterns of the compounds in question:

1) **Noun + Noun:** *Hobbysportler, Showprogramme, Musikvideo, Machtpoker, Kartoffelchips, Wusten-Rally, Faxgeräte, Pfefferspray, Laborversuch, Desserprodukten, Immobilienservice, Eishockey, Beachtemperaturen, Speed-disziplinen, Papiercontainer, Jobsucher, Eishockey, Sturmerstar, Schlagstar, Computerprogramme, Journaldienst, Fotoalbum, Tourneeleaders, Golfturnier, Machtpoker, Laborversuch, Computer, Auslandsengagement, Wusten-Rally, Faxnummer, Strukturfonds, Cheftrainer, Chartermaschine, Golfkrieg, Couchtisch, Gerätset, Chefrolle, Budgetdefizit, Währungsexperten, Horror-Jahr, Splengercup, Milleniumsparty*;
2) **N+N+N:** *Fussballstars, Weltcuperfolg, Tankstellenshops, Europacupfighter, Teamtorhüter, Arbeitsmarktservice, Eishockey-Star, Snowboardbekleidung, Snowboardschule, Volleyball-Herren, Landwirtschaftsmanagement, Ski Flug-Fans, Vorteilclub-Info, Weltcup-Rennen, Fußballfan, Champions-League Duelle, Gänsehaut-Feeling, Weltcup-Finale, Wellness-Botenstoff, Volksmusik-Events, Fußball-Team, Football-Begegnung, Champions-League-Final, Eishockey-Runde, Baby-Trager Tuch, Trickfilme-Spektale, Benefiz-Hockey-Night, Handicap-start, Eishockey-Welt, Weltcup-kalender, Weltcup-Damen, Comedy-Wochenwucht, Autofilter-Fabrik, „Lakeside"-Areal, Weltcupteam, Eishockey-Meister, Weltcupsieger, Eishockey-Experten, Autohaus-Chefin, Fenstertag-Chance, Fenstertage-Fans, Krisen-*

Interventions-Team, Autofahrer-clubs, Automobilclubs, Family Club Karte, Gästeservicecenter, Color Quality Club, Zeitungsdesign-Gurus, Mediendesign-Studenten, Spitting-Image-sendung, Notenbankinchefin, Cross-Weltmeister, Bank-austria region, Weltcuperfolg, Rally-Cross Auto, Fusball-Profi, High-Tech-Schlacht, Star-Trek Film, Fantasy-Action-Gemisch.

3) **Abbreviation + Noun:** *TV-Produktion, TV-format, UM-Team, TV-Rede, VC-Chef, TV-Phänomen, TV-Theater, TV-Sender, TV-Stationen, TV-serie, TV-Auftritt, SKT-team, TV-Gerät, VW-Golf, VIP-Tickets*
4) **Adjective + Noun + Noun***: High-Tech-Schlacht, Schwergewichts-Champ, Edelstahlcontainer, Broadway-Aufführung, Neujahrs-Interview, Fair-Play-Komitee, Softwarenentwicklung, Leichtathletik-Champion, Supercup-Sieg, Neujahrsbaby, Nationalbankchef, Doppeljackpot;*
5) **Adjective + Noun***: Nationalteam, Geheimtip, Hauptsponsor, Einzel-Interview, Celtic-Spielers, Minibaloons, Alkotest, Kaernten Card, Euro-Block, Euro-Baby, Top Qualität; 1 Wort?*
6) **Phrases:** *I love you virus, Support your local Tageszeitung, Best-of-seven-Serie. 24 Stunden News und Service, Villach city of music, Orchestra of the Swan, Fifty shades of Grey Trailer, Queen of the Dessert, You're beautiful, Order of the British Empire;*
7) **Noun +Noun + Noun + Noun***: Snowboardweltmeister, Football-Spitzenteam, Schauspieler-Traumteam, Frau-Jazz-Band, Eishockey-Teamstürmer, Autobahn-Funpark;*
8) **Verb (ing) + Noun***: Training-Sturz, Marketing-Leitner, Trainingspause, Marketingstrategie, Roaming-Gebühren, Trainigszwecken;*
9) **Numeral + Noun + Noun***: Ein-Mann-Team, Ein-Mann-Show, One-way-Kommunikation, One-World-Allianz, Six-Pack-Wirte;*
10) **Noun +Verb (ing):** *Drauconsulting, Standardmarketing, Fitnesstraining, Krafttrainig, Devisenfixing;*
11) **Preposition + Noun + Noun***: Übersee-Department, Online-anmeldung, Offroad-Spektakel, Offroad-Funktion;*
12) **Abbreviation + Noun + Noun:** *US-Kinocharts, EU-Chefsessel, US-Chefunterhalter*
13) **Verb + preposition + Noun***: Pay-off-Plätze, Play-off-Einzug;*
14) **Verb + Noun + Noun***: Open-Source-Variante, Open-source-program*
15) **Adjective + Numeral + Noun***: Top-Ten-Platzierungen , Top-Ten-Platz*
16) **Noun + Adjective +Noun***: Medienmegastar, Pepper-High-Effekt*
17) **Pronoun + preposition + noun + noun***: Allroundbürokraft*
18) **Adjective + Noun + Noun + Noun***: Hightech-Karbon-Prothesen*
19) **Noun + Noun +Adjective + Noun***: Eurostar-Fernzug*
20) **Preposition + Verb + ing***: Übertraining*
21) **Preposition + Verb + Noun***: Make-up-Spiegel*
22) **Adjective + Participle 2***: Top vorbereitet*
23) **Numeral + Noun + Adjective + Adjective:** Vier-Sterne-Superior-Prima
24) **Verb + Noun + Noun + Noun:** *Open-Source-Passwort-Management*
25) **Noun + Adjective***: Sony classical*
26) **Verb + Noun:** *Crash-Landung*
27) **Preposition + Verb***: durchboxen*

The results of the analyzed compounds are given in Table 4.

Table 4
The total values of hybrid compounds in the newspaper
"Kleine Zeitung" and their rank-frequency distribution

1	Noun + Noun	329
2	Noun + Noun + Noun	59
3	Abbreviation + Noun	15
4	Adjective + Noun + Noun	12
5	Adjective + Noun	11
6	Phrases	10
7	Noun + Noun + Noun + Noun	6
8	Verb + ing + Noun	6
9	Numeral + Noun + Noun	5
10	Noun + Verb + ing	5
11	Preposition + Noun + Noun	3
12	Abbreviation + Noun + Noun	3
13	Verb + Preposition + Noun	2
14	Verb + Noun + Noun	2
15	Adjective + Numeral + Noun	2
16	Noun + Adjective + Noun	2
17	Pronoun + Preposition +Noun + Noun	1
18	Adjective + Noun + Noun + Noun	1
19	Noun + Noun + Adjective + Noun	1
20	Preposition + Verb (ing)	1
21	Preposition + Verb + Noun	1
22	Adjective + Participle 2	1
23	Numeral + Noun + Adjective + Adjective	1
24	Verb + Noun + Noun +Noun	1
25	Noun + adjective	1
26	Verb + Noun	1
27	Preposition + Verb	1

Not caring for the qualitative side of these compounds one can fit to these rank-frequencies the power function $y = ax^b + 1$ as can be seen in Table 5

Table 5
Fitting the power function with additive constant
to the ranked frequencies in Table 4

Rank	**Frequency**	**Computed frequency**
1	329	328.89
2	59	58.46
3	15	21.74
4	12	11.07
5	11	6.75
6	10	4.63
7	6	3.47
8	6	2.76

9	5	2.31
10	5	2.01
11	3	1.79
12	3	1.64
13	2	1.52
14	2	1.43
15	2	1.36
16	2	1.31
17	1	1.27
18	1	1.23
19	1	1.20
20	1	1.18
21	1	1.16
22	1	1.14
23	1	1.12
24	1	1.11
25	1	1.10
26	1	1.09
27	1	1.08
$a = 327.8950, b = -2.5127, R^2 = 0.9987$		

Needless to say, a better result could be obtained by fitting some discrete distribution (e.g. the negative binomial) but our aim is merely to show the regularity of this process.

Conclusions:

- The above-mentioned analysis has revealed 27 structural patterns of English-German (German-English) compounds;
- The highest frequency of the hybrid compounds is observed in the patterns Noun + Noun (329) and Nouns + Nouns + Nouns (59). On the whole, the nouns are mostly borrowed parts of speech;
- The frequency of the structural patterns of compounds was determined with the help of the rank-frequency power function and negative binominal distribution. The results have turned out to be identical.

4. The investigation of cohesion for English-German and German-English compounds

According to Fan/Altmann (2007:8), "cohesion is a property which can be defined for any linguistic entity consisting of two or more components". *The aim of the given research* is to measure the cohesion of English-German and German-English compounds in the Austrian newspaper "Kleine Zeitung" and compare the results with Fan/Altmann's ones (2007). In such a way, we analyze 483 compounds. In order to do, it is necessary at first to illustrate which types of the cohesion for German compounds we have obtained in our analysis:

Fusion "takes place when at least one of the elements loses a part of its body or fused part belongs to both elements" (Fan/Altmann, 2007:2). We distinguish the following types of fusion in the analyzed newspapers:

a) abbreviation of the first elements: *NASA-Chef, TV-Zentrale, VC-Chef, OFB-Cup;*

b) the element of the compound loses partly his body: *Alkotest, Eurostar, Euroconsult, Infocard.*

Joining. The elements of the compounds are joined together. In this case we should take into account the degree of this joining:

a) compounds without changes: *Hauptsponsor, Europacupfighter, Teamtorhüter, Faxnummer, Gerätset, Couchtisch, Golfkrieg, Computersprachler.*

b) a phonetic or morphological fugue: *Auslandsengagement, Währungsexperten, Rettungsteam, Pistenquiz, Zeitungsdesign-Gurus, Tankstellenshop, Vorteilscard.*

Hyphenized compounds. The compounds are joined with the help of a hyphen: *Computer-Genie, Microsoft-Suchmaschine, Entertainment-Zentrale, Fun-Faktor, Comic-Figuren, Top-Preis, Mediendesign-Studenten.*

Blank compounds. The elements of the compounds are written separately: *Family Club Karte, Congress Center Villach, Premier League, Villach city of music.*

At this stage of research we have calculated the number of blank, joining and hyphenized compounds. The results are given in Table 6.

Table 6

The total values of English-German hybrid compound cohesion and their rank-frequency distribution in the Austrian newspaper "Kleine Zeitung"

Number	**Type**	**Number**
1.	Joining	243
2.	Hyphenized	187
3.	Fusion	48
4.	Blank	5
Total		483

The data can be fitting using the Zipf-Alekseev function $y = c \bullet x^{a+b*\ln(x)}$. The results are given in Table 7

Table 7

Fitting the Zipf-Alekseev function to the data in Table 6

Rank	**Frequency**	**Computed**
1	243	242.95
2	187	187.34
3	48	45.89
4	5	9.65
a = 1.5772, b = -2.8164, c = 242.9517, R^2 = 0.9993		

The above data can be considered in more detail. In that case, using the Zipf-Alekseev function with an additive 1, one obtains the results presented in Table 8.

Table 8
The extended values of English-German hybrid compound cohesion and their rank-frequency distribution in the Austrian newspaper "Kleine Zeitung"

Rank	**Type of cohesion**	**Number**	**Computed**
1.	Joining	207	207.23
2.	Hyphenized	180	178.60
3.	Joining with a joining element	33	41.68
4.	Fusion with 2 abbreviated elements	26	8.69
5.	Fusion	14	2.48
6.	Hyphenized with a joining element	7	1.31
7.	Blank	5	1.07
8.	Fusion with 3 abbreviated elements	4	1.02
9.	Fusion with one abbreviated element	3	1.00
10.	Blank with a joining element	3	1.00
11.	Fusion with 4 abbreviated elements	1	1.00
Total		483	
a = 1.9417, b = -3.1123, c = 206.2306, R^2 = 0.9898			

Conclusions: The above-done analysis on the cohesion of English-German (German-English) mixed (hybrid) compounds allow us making the following conclusions:

- The following types of cohesion is characteristic of German-English (English-German) compounds: joining, joining with a joining element; hyphenized, hyphenized with a joining element; fusion, fusion with 1, 2, 3, 4 elements; blank, blank with a joining element;
- The highest frequency of the hybrid compounds is observed in joining (42.9 %) and hyphenized (37.3 %) compounds; the lowest frequency is characteristic of fusion with 1, 3, 4 elements, blanks and blanks with a joining element.

Bibliography

Altmann, G., Buttlar, H.v., Rott, W., Strauss, U. (1983). A law of change in language. In: Brainerd, B. (ed.), *Historical Linguistics: 104-115.* Bochum: Brockmeyer.

Altmann, G. (1983). Das Piotrowski-Gesetz und seine Verallgemeinerungen. In: Best, K.-H., Kohlhase, J. (eds.), *Exakte Sprachwandelforschung: 54-90.* Göttringen: Herodot.

Best, K.-H., Beöthy, E., Altmann, G. (1990). Ein methodischer Beitrag zum Piotrowski-Gesetz. In: Hammerl, R. (ed.), *Glottometrika 12, 115-124.* Bochum: Brockmeyer.

Best, Karl-Heinz (2002). Der Zuwachs der Wörter auf *–ical* im Deutschen. *Glottometrics 2*, 11-16.

Best, Karl-Heinz (2004). Zur Ausbreitung von Wörtern arabischer Herkunft im Deutschen. *Glottometrics 8,* 75-78.

Best, Karl-Heinz (2005). Turzismen im Deutschen. *Glottometrics* 11, 56-63.

Best, Karl-Heinz (2013). Iranismen im Deutschen. *Glottometrics* 26, 1-8.

Best, Karl-Heinz (2014). Hebraismen im Deutschen. *Glottometrics* 27, 10-17.

Carstensen, B. (1965). Englische Einflüsse auf die deutsche Sprache nach 1945. *Beiheft zum Jahrbuch für Amerikastudien, 13.* Heidelberg.

Carstensen, B. (1980). Der Einfluß des Englischen auf das Deutsche – grammatische Probleme. *Arbeiten aus Anglistik und Amerikanistik 5, 37-63.*

Heller, K. (1965). Das Fremdwort als Modewort. *Neue Deutsche Presse 19,42.*

Heller, K. (1966). *Das Fremdwort in der deutschen Sprache der Gegenwart: Untersuchungen im Bereich der Gebrauchssprache* (VEB Bibliographisches Institut: Leipzig)

Leopold, E. (2005).Das Piotrowski-Gesetz. In: Köhler, R., Altmann, G., Piotrowski, R.G. (2005), *Quantitative Linguistics. An International Handbook: 627-633.* Berlin: de Gruyter.

Paul, H. (1920). Prinzipien der Sprachgeschichte. – Aufl. 5, Halle.

Zindler, H. (1959). *Anglizismen in der deutschen Presse nach 1945*. Diss. Universität Kiel.

Zindler, H. (1975). Anglizismen im heutigen Deutsch – beobachtet in der Sprache der Presse 1945-1966. In: *Beiträge zu den Fortbildungskursen des Goethe-Instituts für ausländische Deutschlehrern an Schulen und Hochschulen (München): 82-91*.

Glottometrics 31, 2015, 50-53

Malay borrowings in English

Karl-Heinz Best

Abstract. This study presents a further support of the logistic law, known in linguistics as Piotrowski law, using data which can be found in Cannon (1992).

Keywords: language change, Malay, English, Piotrowski law.

0. *Gecko*, *gong*, *ketchup*, *Malay*, *nasigoreng*, *orangutan*, *Penang*, *sago*, *sambal*: These are only some of the "368 primary Malay borrowings" from Cannon's list (1992: 154-160) which were taken over into English and survived up to now. There are further 538 "secondary items" comprising variants, obsolete forms, derivations etc., that is, in total 906 words. In the present article we are not interested in the individual borrowings but in the time interval and the dynamics of the borrowings by English: the investigation concerns only the "primary borrowings" because for the other ones Cannon does not give time specifications.

1. The most important task of quantitative linguistics is the derivation of hypotheses fro background knowledge, testing them in many languages and give them the status of laws. One of these hypotheses is known as Piotrowski law capturing the law-like process of changes in language. The law obtained its name in honour of R.G. Piotrovskiy who, as far as we know, together with his wife, was the first linguist trying to describe mathematically the change in language (Piotrovskaya, Piotrovskiy 1974; cf. also Piotrowski, Bektaev, Piotrowskaja 1985: 36ff, 81ff.).

The model has been revised by Altmann (1983, 1992) and Altmann et al. (1983) and obtained a form which has been positively tested in almost all respective investigations. It contains two basic types: the most usual is the change beginning with a linguistic change which is slow at the beginning but becomes more rapid and after a turning point slows down until an upper boundary accepted by the community is achieved (full or partial change). The second basic type differs from the first by the fact that after achieving a maximum a backward process begins leading to partial or full removing of the change (reversible process).

An idea of the course of change – without mathematical modelling – existed already before Piotrowskaya and Piotrowski (1974). Presentations in form of idealized curves can be found already with Lindgren (1953: 185; 1961: 56; Best 2008a); a verbal description of the "S-curve" can be found in Osgood, Sebeok (1965: 155), as well as in Wenreich, Labov, Herzog (1968: 113). In order to characterize the varying velocity of change, Aitchinson (1991: 83) speaks about a "slow – quick – quick – slow pattern". The so-called S-curve belongs not later than since 1987 to the common knowledge of linguistics, namely since its accepting by Crystal (1993: 332) in his *Cambridge Encyclopedia of Language.*.

2. In the meantime, it could be shown that the Piotrowski-law holds true with various language change phenomena: internal language change in writing, phonetics, grammar and lexicon, both in language system and its use, with borrowing phenomena and with children's language learning (for examples of all these phenomena cf. Best 2006: 196ff; Best 2008b: 108ff.). Some cases of rejection can be explained by scarcity of data: for example the borrowings of Chinese words in German could not be modelled because of a too small number of datable borrowings; however, it was no problem in English which is better exemplified (Best 2008c).

The Piotrowski law represents a special case of growth processes which were used first for the modelling of population increase (Verhulst 1838, 1845; Pearl 1926), later on used also in other domains (Banks 1994).

3. Even if the Piotrowski law belongs to the best corroborated language laws, it is worth of searching for additional testing possibilities because in every new case there is the danger to reject it, a situation leading to new considerations (cf. Schenke 2008). A further testing can be performed on the basis of data stated by Cannon (1992: 140) concerning the taking-over of Malay words into English "excluding purely Indonesian terms" (Cannon 1992: 136). They are presented in the table below under p_{obs} and as far as it was possible, presented in step of 50 years since 1400. Small deviations are observable according to Cannons data for the times 1700 – 1753, 1800 – 1899 and for 1950 – 1982. In the computations they were ignored and the trend was computed as if there were everywhere the 50 years steps; the time interval 1800-1899 has been treated as if it was the interval 1850-1899; for the interval 1800 – 1949, Cannon presents explicitly one case. In the column p_{obs} (cumulated) one finds the added observed values.

In order to test whether the Piotrowski law in form of incomplete change

$$(1) \quad p_t = \frac{c}{1+ae^{-bt}}$$

sufficiently captures also the borrowings from Malay into English formula (1) will be fitted to the cumulative frequency data in the column p_{obs} by means of the software NLREG. The results of fitting are displayed in column p_{comp}; the evaluation of the fitting is performed by the computation of the determination coefficient which, in this case, yields $R^2 = 0.99$, i.e. an excellent result. That means, the Piotrowski law holds in this case, too.

Table 1
Borrowings of Malay words by English (according to Cannon 1992)

t	Time	$p_{obs.}$	$p_{obs.}$ (cumulated)	p_{comp}
1	1449	1	1	1.4113
2	1450-99	0	1	2.7413
3	1500-49	0	1	5.3058
4	1550-99	19	20	10.1991
5	1600-49	10	30	19.3520
6	1650-99	18	48	35.8534
7	1700-53	8	56	63.7173
8	1754-99	35	91	105.9424
10	1800-99	142	233	218.0620
11	1900-49	32	265	267.2299
11.5	1950-82	17	282	286.5372
a = 482.1913 b = 0.6678 c = 350.4069 R^2 = 0.99				

p_{obs} – number of borrowings from Malay according to Cannon
p_{obs}(cumulated) – added values of p_{obs}
p_{comp} – fitting formula (1) to p_{obs}(cumulated)
a, b, c – parameters of (1); c indicated the approached final value

Figure 1 displays the positive result optically.

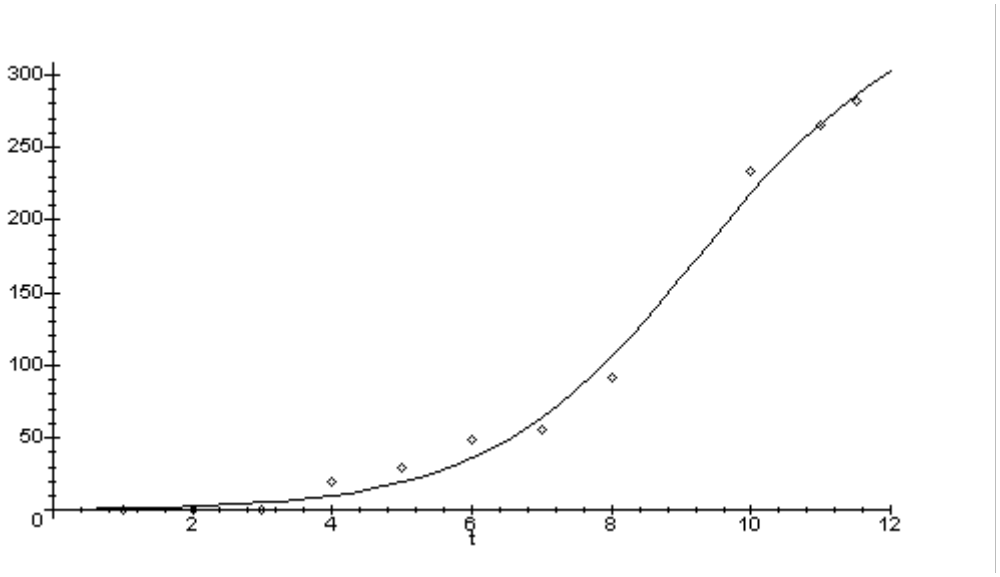


Figure 1. Fitting Piotrowski law to Malay borrowings in English

4. Hence one can state that the Piotrowski law holds true in this case, too. Parameter c is to be treated cautiously. It predicts the final state of the given process but it depends on the works from which the data were obtained and the number of datable borrowings. Cannon (1992: 136) leans against seven sources: six dictionaries and a periodical. If one had used other sources, one had obtained more or less datable borrowings, and if one had changed the way of sampling, one had automatically obtained a different parameter c.

References

Aitchison, J. (1991). *Language change: progress or decay?* Second edition. Cambridge: Cambridge University Press.

Altmann, Gabriel (1983). Das Piotrowski-Gesetz und seine Verallgemeinerungen. In: Best, Karl-Heinz, & Kohlhase, Jörg (Hrsg.) (1983). *Exakte Sprachwandelforschung* (S. 54-90). Göttingen: edition herodot.

Altmann, Gabriel (1992). Piotrowski's Law of Language Change. In: Saukkonen, Pauli (ed.), *What is Language Synergetics?* (S. 34-35). Oulu: Acta Universitatis Ouluensis, Series B: Humaniora, 16).

Altmann, Gabriel, von Buttlar, H., Rott, W., & Strauß, U. (1983). A law of change in language. In: Brainerd, B. (ed.), *Historical linguistics* (S. 104-115). Bochum: Brockmeyer.

Banks, Robert B. (1994). *Growth and Diffusion Phenomena. Mathematical Frameworks and Applications.* Berlin u.a.: Springer.

Best, Karl-Heinz (2006). *Quantitative Linguistik: Eine Annäherung.* 3., stark überarbeitete und ergänzte Auflage. Göttingen: Peust & Gutschmidt.

Best, Karl-Heinz (2008a). Kaj Brynolf Lindgren (1922-2007). Glottometrics, 127-131. (Auch in: Best, Karl-Heinz (2015). *Studien zur Geschichte der Quantitativen Linguistik. Band 1 (106-111).* Lüdenscheid: RAM-Verlag.)

Best, Karl-Heinz (2008b). *LinK. Linguistik in Kürze.*. 5., durchgesehene Auflage. Lüdenscheid: RAM-Verlag.

Best, Karl-Heinz (2008c). Sinismen im Deutschen und Englischen. *Glottometrics 17, 87-93.* (Auch in: Best, Karl-Heinz (2015). *Entlehnungen und Fremdwörter: Quantitative Aspekte (45-52).* Lüdenscheid: RAM-Verlag.

Cannon, Garland (1992). Malay(sian) borrowings in English. *American Speech 67, 134-162.*

Crystal, Davis (1993). *Die Cambridge Enzyklopädie der Sprache.* Frankfurt/New York: Campus.

Lindgren, Kaj B. (1953). *Die Apokope des mhd. –e in seinen verschiedenen Funktionen.* Helsinki (= Suomalainen tiedeakatemian toimituksia/ Annales academiae scientiarum fennicae; Sarja/ Ser. B, Nide/ Tom. 78,2)

Lindgren, Kaj B. (1961). *Die Ausbreitung der nhd. Diphthongierung bis 1500.* Helsinki (= Suomalainen tiedeakatemian toimituksia/ Annales academiae scientiarum fennicae; Sarja/ Ser. B, Nide/ Tom. 123,2)

Osgood, C.E., & Sebeok, T.A. (eds.) (1954/ 1965). *Psycholinguistics.* Bloomington/ London: Indiana University Press.

Pearl, Raymond (1926). *The Biology of Population Growth.* London: Williams and Norgate.

Piotrowski, R.G., Bektaev, K.B., & Piotrowskaja, A.A. (1985). *Mathematische Ling*uistik. Bochum: Brockmeyer.

Piotrovskaja, A.A., & Piotrovskij, R.G. (1974). Matematičeskie modeli diachronii i tekstoobrazovanija. In: *Statistika reči i avtomatičeskij analiz teksta* (S. 361-400). Leningrad: Nauka.

Schenke, M. (2008). *Sprachliche Innovation – lokale Ursache und globale Wirkungen*. Oldenburg: Diss.

Verhulst, P.-F. (1838). Notice sur la loi que la population suit dans son accroissement. *Correspondance Mathématique et Physique, Tome X, 3-21.*

Verhulst, P.-F. (1845). Recherches mathématiques sur la loi d'accroissement de la population. *Nouveaux Mémoires de l'Académie Royale des Sciences et Belles-Lettres de Bruxelles, Tome XVIII, 5-38.*

Weinreich, U., Labov, W., & Herzog, M.E. (1968). Empirical foundations for a theory of language change. In: Lehmann, W.P., Malkiel, Y. (eds.): *Directions for historical linguistics :95-188.* Austin: University of Austin Press.

Software

MAPLE V Release 4. 1996. Berlin u.a.: Springer.

NLREG. Nonlinear Regression Analysis Program. Ph.H. Sherrod. Copyright (c) 1991-2001.

Glottometrics 31, 2015, 54-75

Comparison of vocabulary richness in two translated *Hongloumeng*

Yu Fang & Haitao Liu[*]

Abstract: The hypothesis that vocabulary richness of translated language is lower than that of the native language has been found in previous studies and we also assume native speakers have a larger vocabulary than non-native speakers which would present in their writings. The present article, using two vocabulary richness indicators - STTR and lambda, reevaluates the two hypotheses based on the two versions of *Hongloumeng*, a Chinese literary classic, translated into English respectively by Hawkes and Yang Xianyi. The result indicates that the vocabulary richness in the native speaker's (Hawkes) version is no higher than that in the non-native speaker's version and the vocabulary richness of the two translated *Hongloumeng* is also lower than native English fictions, though the source text were written in 18th century. A deeper investigation into the wordlists and concordance lines reveals that Hawkes used words more frequently concerning the same word; what is more, Yang used more unfamiliar words in his version, especially in translating culture-loaded words.

Keywords: vocabulary richness; lambda; STTR; translation; corpus; Chinese, English

1. Introduction

The study of vocabulary richness, founded by Chotlos and Yule in 1940s (Chotlos, 1944; Yule, 1944), is of great importance in literary research. However, it is also "a complex problem both linguistically and mathematically" (Wimmer & Altmann, 1991, p.8). As far as we know, research in this field often follows three lines: (a) Testing whether vocabulary richness can be used for authorship attribution (Hoover, 2003; Jamak, Savatić & Can, 2012); (b) using vocabulary richness for authors' stylistic characteristics (Smith & Kelly, 2002); (c) applying vocabulary richness to genre analysis (Kubát & Milička, 2013).

The research of vocabulary richness mentioned above dealt with original works. However, we should not neglect the function of vocabulary richness in translation works, for vocabulary choice of translators has direct impact on the quality and readability of translations. As a result, researchers have attached great importance to the studies of vocabulary in literature translations (Laviosa, 1998; Jantunen, 2002; Kojima & Yamashita, 2014).

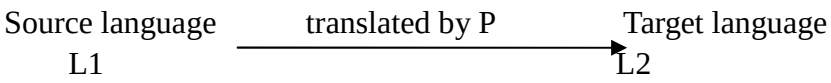
The development of corpus linguistics in recent years, together with corpus linguistics techniques in word analysis using WordSmith Tools, AntConc and Tree tagger etc., has made a

[*] Address correspondence to: Haitao Liu, Department of Linguistics, Zhejiang University, 310058, Hangzhou, Zhejiang, China. Email address: lhtzju@gmail.com

significant contribution to the study of vocabulary richness in translations. The first computer-assisted study in translation can be attributed to Gellerstam (1986), who explored translationese in novels translated from English into Swedish. From then on, corpus linguistics has brought about "a paradigmatic shift in Translation Studies" (Xiao & Dai 2014, p. 12). Now, "corpus-based translation study is an established subfield of the descriptive branch of the discipline" (Zanetti, 2013, p. 21).

One major field of corpus-based studies in translation works is translator comparison. Baker (2000) compared translations by Peter Clark and Peter Bush. Through statistical evidence like TTR and average sentence length, she found that Peter Clark's works are limited in a more restricted range of variation than Peter Clark's works. Qu (2012) compared *Six Chapters of a Floating Life* translated by two Chinese translators to examine their differences in computational stylistics and vocabulary richness and TTR was chosen as an indicator. The result showed that there is no significant difference between the two versions.

A translation process must have at least three elements: a source language (L1), a target language (L2) and one or more translators (P), whose relation can be displayed as:



The translator P may be the native speaker of L1 or L2 or another new language L3. In most cases, the first two are the majority, so the third one is not included in our discussion.

Xiao & Yue (2009), comparing five categories of fictions in the *LCMC* (written by native Chinese) and *The Contemporary Chinese Translated Fiction Corpus* (composed of novels translated from English by native Chinese speakers), found that the vocabulary richness in translated Chinese fictions is lower than that in native Chinese fictions. The same feature can be found in other languages like English. Laviosa (1998), studying the distinctive features of translational English (represented by TEC) in relation to native English (represented by the BNC), also found that the vocabulary richness in TEC is lower than that in BNC.

Source texts in these studies were written in the late 20th century, in other words, in modern language. However, will this rule apply to source texts completed earlier? What is more, translators in these studies are native speakers of L2, that is, the target language. Until now, little research has been done provided that translators are native speakers of L1. We usually intuitively assume that native speakers have a larger vocabulary than non-native speakers, thus L2 speakers are more likely to use more different words than L1 speakers in translating the same text written in L1 to text in L2. But whether this assumption is true needs further investigation.

In this paper, we will choose *Hongloumeng* and its two translation versions by David Hawkes and Yang Xianyi as the material. STTR (standardized type-token ratio) and lambda, as two vocabulary richness indicators, are used to measure their vocabulary richness. Considering the limitations and problems existing in previous research, we keep the following research questions in mind:

Question 1. Is the vocabulary richness of native speaker's (Hawkes) version higher than that of non-native speaker's (Yang) version?

Question 2. Is the vocabulary richness of the two translated *Hongloumeng* lower than native English texts, just like those texts translated from original texts in modern language?

Question 3. If the answer to the first or second question is negative, then how can we explain this phenomenon?

2. Materials and Method

To carry out the research smoothly, the literature work studied must have been translated by two translators - one is the native speaker of L1 and the other is the native speaker of L2, thus *Hongloumeng* is chosen in this study.

Hongloumeng, one of the masterpieces of Chinese literature and one of the Four Great Chinese Classical Novels written in the mid-eighteenth century, has been widely popular throughout the last two centuries. Until now, there have been already nine complete or selective English translations of the book (Chen & Jiang, 2003), among which, two of them stood out: one is *The Story of the Stone* translated by David Hawkes, a British Sinologist (the first 80 chapters) together with John Minford, his son-in-law (the remaining 40 chapters); the other is *A Dream of Red Mansions* translated by a Chinese translator Yang Xianyi and his wife Gladys Yang. Both of them were published in the late 1970s and the early 1980s and electronic versions can be obtained from the Internet. To reduce the workload while still reach the aim of the study, we randomly chose 30 chapters (Chapter 1, 8, 10, 15, 20, 24, 25, 28, 30, 32, 35, 40, 45, 48, 52, 56, 60, 64, 68, 70, 76, 80, 84, 88, 90, 100, 108, 110, 116, 120) out of the total 120 chapters from each translation and built two corpora: Yang's Version of *Hongloumeng* (YVH) and Hawkes' Version of *Hongloumeng* (HVH), Table 1 shows the information of the corpora.

Table 1
Total number of words in 30 chapters of the two translations

	HVH	**YVH**
Chapter 1	7,155	5,898
Chapter 8	5,870	4,774
Chapter 10	4755	3,494
Chapter 15	5,141	3,741
Chapter 20	5,187	3,956
Chapter 24	7,520	5,926
Chapter 25	7,181	5,589
Chapter 28	9,013	6,509
Chapter 30	5,501	3,716
Chapter 32	5,282	3,797
Chapter 35	7,539	5,055
Chapter 40	9,447	6,718
Chapter 45	8,532	6,409
Chapter 48	6,848	5,166
Chapter 52	8,408	5,863

Chapter 56	8,461	6,374
Chapter 60	7,862	5,942
Chapter 64	10,116	6,873
Chapter 68	9,110	6,135
Chapter 70	6,709	4,888
Chapter 76	6,814	5,541
Chapter 80	7,138	5,286
Chapter 84	6,655	5,223
Chapter 88	5,648	4,639
Chapter 90	5,207	4,300
Chapter 100	5,456	4,079
Chapter 108	6,463	5,098
Chapter 110	6,471	5,023
Chapter 116	6,437	4,880
Chapter 120	8,217	5,862

Various indicators can be used for vocabulary richness measurement, among which, one of the oldest and easiest ways is the type-token ratio (TTR). TTR refers to the relationship between the total number of running words in a corpus and the number of different words used (Olohan, 2004, p. 80). Thus TTR can reflect word choice of translators to some extent. A higher TTR means that the translator uses a wider range of words, while a lower TTR indicates that the translator chooses fewer word items in a certain text.

However, TTR has its own disadvantages. The stumbling block of TTR, as Kubát & Milička (2013) mentioned, is that "there is a dependence on text size" (p.339). That is to say, if texts chosen for comparison differ in length, TTR is not reliable. STTR (standard type token ratio), calculating TTR based on every 1000 words, may be a better choice.

Because of the limitations of TTR, researchers are seeking another indicator, which could minimize the influence of text length, to measure vocabulary richness of texts with various types and length. Popescu (2011) proposed a new indicator – lambda (Λ):

$$\Lambda = \frac{L(Log_{10} N)}{N}$$

N refers to the text size, *L* stands for the arc length that can be computed as

$$L = \sum_{i=1}^{V-1} [(f_i - f_{i+1})^2 + 1]^{1/2}$$

where f_i are the ordered absolute frequencies (i = 1,2,…,V) and V is the highest rank (= vocabulary).

In this research, to get a more reliable result, we apply both of the two indicators to calculate the vocabulary richness of two translations. If STTR and lambda point to the same result, we can then confirm the finding.

Applying the two indicators mentioned above, we firstly figure out $STTR_1$ (standing for STTR value of Hawkes' version), $STTR_2$ (standing for STTR value of Yang's version), $lambda_1$

(standing for lambda value of Hawkes version) and $lambda_2$ (standing for lambda value of Yang's version). Quantitative Index Text Analyzer (QUITA), a new software for quantitative studies, enables "researchers from various disciplines (linguistics, criticism, history, sociology, psychology, politics, biology, etc.) to analyze texts using quantitative methods" (Kubát, Matlach & Čech, 2014, p.1). Many indicators in this program are connected with rank frequency distribution of texts like type, token, h-point, L and lambda. One more feature of QUITA is that it can be used to create charts, that is to say, "there is no need to use any additional software such as spreadsheet applications or special statistical programs" (Kubát, Matlach & Čech, 2014, p.1), which is convenient for data export.

Another software used in this study is WordSmith Tools 6.0 (Scott, 2015), designed by Lexical Analysis Software Ltd. WordSmith Tools contain three main tools for corpus exploration, that is, Concord, KeyWords and WordList. We will use Concord and WordList in this study and their functions are explained in details as follows.

Wordlists provide frequency information of every word in the corpus, either in an ascending order or in an alphabetical order. It is applied here to capture the most frequently-used lexical words in each corpus. Concordance gives the context of every specific word, thus the interpretation of concordance lines can help us understand texts better. In this thesis, the concordance analysis will be used together with the wordlist to reveal the word usage of the two translators.

In this study, QUITA is used to compute $STTR_1$, $STTR_2$, $lambda_1$ and $lambda_2$. After obtaining all the data, SPSS 20 is applied to carry out significance analysis: $STTR_1$ and $STTR_2$ are compared first, then the p-values of $lambda_1$ and $lambda_2$ are tested. After that, we look at whether the two results are consistent.

3. Results and Discussion

3.1 Comparison of STTR and Lambda

Using QUITA and WordSmith Tools 6.0, we can get the values of type, token, STTR and lambda. The result is shown in Table 2.

Table 2
TTR and lambda in Yang and Hawkes' translations

	Hawkes				**Yang**			
Chapter	**type**	**token**	**$STTR_1$**	**$lambda_1$**	**type**	**token**	**$STTR_2$**	**$lambda_2$**
1	1,872	7,155	44.96	1.205983	1,714	5,898	47.12	1.283607
8	1,395	5,870	41.74	1.014032	1,252	4,774	43.10	1.06575
10	1,082	4,755	38.08	0.965711	950	3,494	40.00	1.074634
15	1,381	5,141	43.08	1.164036	1,087	3,741	43.87	1.195028
20	1,167	5,187	39.80	0.940358	1,035	3,956	41.80	1.02965
24	1,483	7,520	39.77	0.899358	1,317	5,926	43.32	0.958859
25	1,731	7,181	42.34	1.074288	1,452	5,589	45.08	1.118629

28	1,721	9,013	39.12	0.893348	1,472	6,509	43.25	0.990018
30	1,304	5,501	40.74	1.012277	1,033	3,716	44.03	1.096506
32	1,090	5,282	37.92	0.857417	961	3,797	43.50	0.981294
35	1,438	7,539	39.44	0.865397	1,202	5,055	43.38	1.00003
40	1,871	9,447	41.94	0.966619	1,527	6,718	45.45	1.056436
45	1,638	8,532	39.06	0.886149	1,355	6,409	39.88	0.925218
48	1,471	6,848	40.37	0.958452	1,298	5,166	42.92	1.040658
52	1,672	8,408	40.11	0.898427	1,390	5,863	41.92	1.017147
56	1,624	8,461	39.66	0.894149	1,383	6,374	40.33	0.961463
60	1,476	7,862	39.43	0.839734	1,237	5,942	40.28	0.888784
64	1,948	10,116	40.92	0.936403	1,487	6,873	41.63	0.968155
68	1,720	9,110	40.90	0.875887	1,288	6,135	40.70	0.934914
70	1,532	6,709	41.60	1.058447	1,290	4,888	44.75	1.128655
76	1,581	6,814	42.87	1.061891	1,398	5,541	43.00	1.127053
80	1,617	7,138	42.24	0.991579	1,358	5,286	42.80	1.063992
84	1,582	6,655	42.77	1.043004	1,182	5,223	39.74	0.960538
88	1,375	5,648	41.88	1.027401	1,128	4,639	41.50	1.000094
90	1,207	5,207	40.32	0.975589	1,043	4,300	39.78	0.973754
100	1,328	5,456	42.06	1.026472	1,019	4,079	43.60	1.002781
108	1,328	6,463	39.98	0.903359	1,159	5,098	43.05	0.988693
110	1,392	6,471	41.15	0.963448	1,187	5,023	43.33	1.023431
116	1,628	6,437	43.10	1.098634	1,271	4,880	43.15	1.073908
120	1,825	8,217	42.56	1.022989	1,467	5,862	44.84	1.089014

According to Table 2, we produce the curves of STTR and lambda distributions by EXCEL, as shown in Figure 1 and Figure 2 respectively.

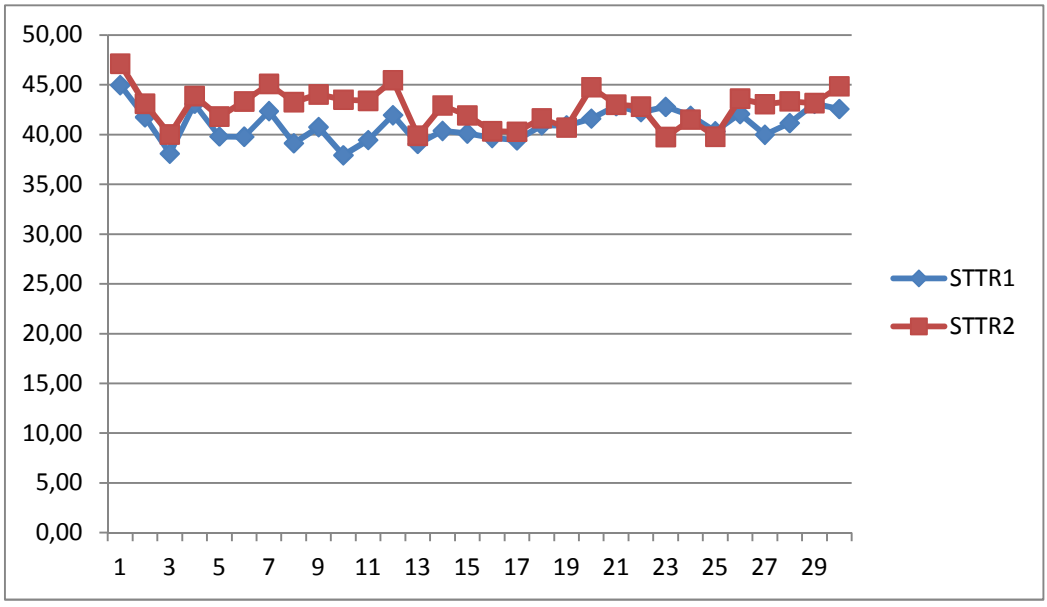


Figure 1. Comparison between $STTR_1$ and $STTR_2$

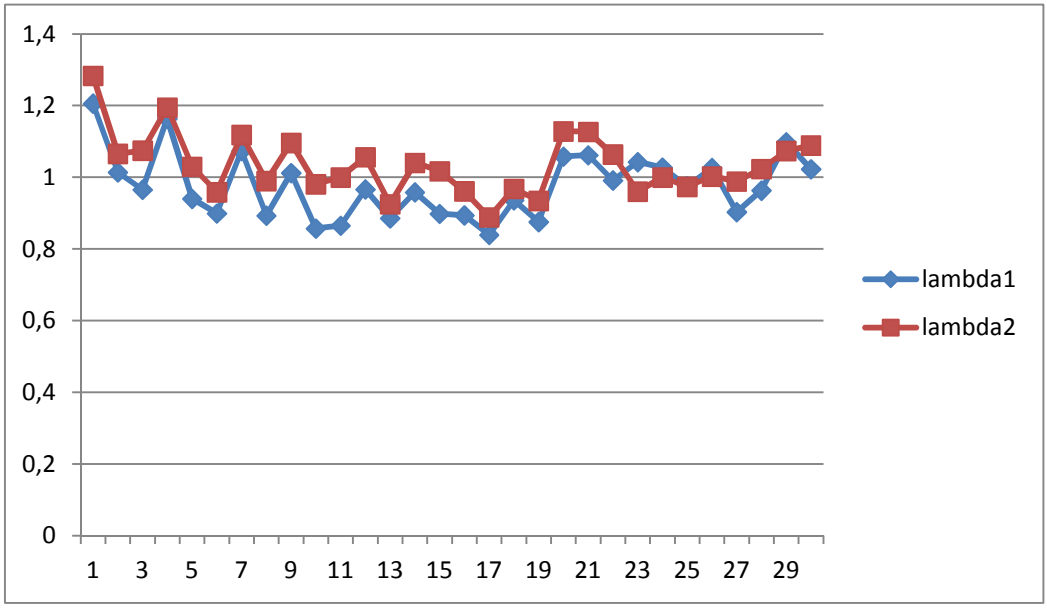


Figure 2. Comparison between $lambda_1$ and $lambda_2$

Figure 1 shows that STTR values of Hawkes' version fluctuate between 37 and 48 and the STTR values of each chapter do not differ much, so does Yang's version. If the STTR values of each chapter in the two translations are compared respectively, it can be concluded that except Chapter 68, 84, 88 and 90, STTR values of Yang's version are higher than that of Hawkes' version in the rest nine chapters listed here. At the same time, since the divergence is not wide, significance test is needed for further discussion.

Figure 2 tells that though the lambda values are in the interval of 0.8 to 1.3. If each chapter is observed separately, it can be found that, nearly all the lambda values of Yang's version are larger than those of Hawkes' version except for Chapters 84, 88, 90 and 116. Similarly, it should be tested later whether this difference is significant. The respective hypotheses are:

H_0: The vocabulary richness of the two translations is the same.
H_1: The vocabulary richness of the two translations is not the same.

Shapiro-Wilk tests show $STTR_1$, $STTR_2$, $lambda_1$ and $lambda_2$ are normally distributed: $W_{STTR1} = 0.9559$, $p_{STTR1} = 0.2426 > 0.05$; $W_{STTR2} = 0.949$, $p_{STTR2} = 0.1591 > 0.05$; $W_{lambda1} = 0.9793$, $p_{lambda1} = 0.8062 > 0.05$; $W_{lambda2} = 0.951$, $p_{lambda2} = 0.1798 > 0.05$. Then independent sample t-test is conducted and a significant difference between $STTR_1$ and $STTR_2$, $t_{(58)} = -3.766$, $p < 0.001$ has been stated. There is also significant difference between $lambda_1$ and $lambda_2$, $t_{(58)} = -2.510$, $p = 0.015 < 0.05$. The two results are consistent, which means that we have to reject H_0 and the vocabulary richness of Yang's version is considered to be higher than that of Hawkes' version.

Though Hawkes is a native English speaker, he does not use higher vocabulary richness than Yang, which contradicts to our previous conjecture.

3.2 Comparing STTR of the two translations with native English texts

British National Corpus (BNC), with 100 million word collection of samples from a wide range of sources like newspapers, journals for all ages and interests, academic books and fictions, is designed to "represent a wide cross-section of British English from the later part of the 20th century" (from BNC official website). So it can best represent the features of native English. Since the two corpora in this paper are translated fictions, we only concern the fiction samples of BNC (BNCFIC hereafter). BNCFIC has 4,947,688 tokens and 63,652 types, and its STTR is 44.48. To find out whether $STTR_1$ and $STTR_2$ is significantly different from $STTR_{BNCFIC}$, we conduct one-sample t-test. The hypotheses are:

H_0: $STTR_1$ and $STTR_2$ are equal to $STTR_{BNCFIC}$.
H_1: $STTR_1$ and $STTR_2$ are lower than $STTR_{BNCFIC}$.

One sample t-test shows that $STTR_1$ is signigificantly lower than $STTR_{BNCFIC}$: $t_{(29)} = -11.164$, $p < 0.001$; the same result can be found between $STTR_2$ and $STTR_{BNCFIC}$: $t_{(29)} = -5.233$, $p < 0.001$.

Although *Hongloumeng* was written in the 18th century and full of cultural and historical information, STTRs of Yang and Hawkes' version are both lower than that in BNCFIC (44.48). This result conforms to previous findings: the vocabulary richness in translated English fictions is lower than native English fictions, which is true both for native and non-native English translators.

3.3 Reasons for Differences between the Two Translations

From the above sections, we know the vocabulary richness in Hawkes' version is lower than that in Yang's version, which contradicts to our assumption. In this part, reasons for the differences are sought by putting words in their specific context. Table 3 shows Top 50 words in the wordlists of two translations.

Table 3
Top 50 words in the wordlists*

rank	Yang			Yang		
	word	**frequency**	**standard frequency**	**word**	**frequency**	**standard frequency**
1	the	8,643	411	the	6,273	400
2	to	6,983	332	to	5,247	335
3	and	6,043	288	and	4,291	274
4	of	4,714	224	a	2,878	184
5	a	4,212	200	you	2,830	181
6	you	3,791	180	of	2,581	165
7	I	3,203	152	I	2,200	140
8	it	2,867	136	**she**	2,048	131
9	in	2,859	136	in	2,037	130

10	**her**	2,810	134	**her**	2,016	129
11	that	2,703	129	he	1,828	117
12	**she**	2,419	1159	it	1,770	113
13	was	2,229	1069	that	1,673	107
14	he	2,135	102	was	1,410	90
15	for	1,889	90	for	1,331	85
16	with	1,701	81	with	1,300	83
17	**said**	1,596	76	this	1,247	80
18	on	1,480	70	on	1,077	69
19	had	1,474	70	had	1,033	66
20	his	1,315	63	as	991	63
21	be	1,310	62	his	989	63
22	have	1,248	59	but	960	61
23	as	1,222	58	they	854	54
24	this	1,158	55	him	793	51
25	but	1,122	53	so	793	51
26	is	1,107	53	have	790	50
27	at	1,065	51	be	789	50
28	all	1,044	50	**Pao-yu**	753	48
29	him	1,040	49	if	750	48
30	they	999	48	me	741	47
31	Jia	947	45	at	708	45
32	if	941	45	is	703	45
33	them	864	41	lady	701	45
34	me	843	40	all	700	45
35	when	837	40	not	692	44
36	what	826	39	we	646	41
37	so	815	39	your	641	41
38	not	796	38	when	601	38
39	there	773	37	then	586	37
40	one	758	36	them	583	37
41	from	743	35	out	572	36
42	out	743	35	what	564	36
43	about	732	35	up	554	35
44	your	728	35	by	551	35
45	we	720	34	can	540	34
46	by	704	33	**said**	530	34
47	up	687	33	now	503	32
48	**Bao-yu**	682	32	one	495	32
49	now	673	32	no	492	31
50	were	673	32	my	484	31

* Standard frequency = (frequency/token)*10000, all figures in the table are rounded to the nearest integer.

Since the two translations differ in word number, standard frequency (per 10000 words) is calculated to get a more reliable analysis. Table 3 shows that the two translations share almost the same words in the Top 50 wordlists and most of them are functional words (e.g. prepositions, conjunctions, articles), which conforms to the regularity: functional words usually have high frequencies. Besides functional words, personal pronouns also have high frequencies, especially "her" and "she", which implies that there are more female characters than male characters in this novel. In addition, "Pao-yu", occurring 48 times per 10000 words in Yang's version and "Bao-yu", occurring 32 times in Hawkes' version, is the main character in this story.

There is also another word that demands our attention: "said". This word occurs 33.81 times per 10000 words in Yang's version and over twice higher (75.93 times per 10000 words) in Hawkes' version. In the original work, the author used "说, 道" (said) to refer to direct speech and if they were all translated as "said", there must be too much monotony in the whole work. Since the frequency of "said" in Yang's version is significantly lower than that in Hawkes' version, it could be supposed that Yang used more synonyms to express the same meaning. Liu & Yan (2010) classified the translation of "said" into several types according to their own research: "explained, suggested, protested" refer to speech content; "exclaimed, shouted, cried" indicate speech act, "asked, replied, told" show word essence and "chuckled, smiled, laughed" convey "said" meaning indirectly. For the purpose of understanding this classification better, some specific examples are provided to see how the two translators deal with this word.

E.g. 1 (from Chapter 1)
Hawkes' version: "You will laugh when I tell you", **said** the monk.
Yang's version: "It's an amusing story." The monk **smiled**.

The original text: 那僧笑道："此事说来好笑，竟是千古未闻的罕事。"

Both Yang and Hawkes shifted the direct speech behind "said" to the front of it. Hawkes directly used "said" to lead this speech, while Yang used "smiled" to imply this speech came from the monk and show his facial expression at the same time.

E. g. 2 (from Chapter 40)
Hawkes' version: You are very busy, Mrs Zhu, **said** Grannie Liu. Li Wan smiled, "I told you you'd never get away. You kept saying yesterday that you had to go, but I knew they wouldn't let you."
"It was Her Old Ladyship that kept me", **said** Grannie Liu, "She said she wanted me to enjoy myself for a day or two before I went back."
Yang's version: "How busy you are, madam!" **remarked** Granny Liu.
"I knew you wouldn't be able to leave yesterday," replied Li Wan with a smile. "Yet you were in such a hurry to get away."
"The old lady made me stay to enjoy myself for a day," **chuckled** Granny Liu.

The original text: 只见丰儿带了刘姥姥板儿进来，说"大奶奶倒忙的紧。"李纨笑道："我说你昨

儿去不成，只忙着要去。”刘姥姥笑道：“老太太留下我，叫我也热闹一天去。

In this dialogue, Hawkes translated the word into “said Grannie Liu” twice. Yang, to put it into a specific context, translated it as “remarked Granny Liu” and “chuckled Granny Liu”. Though these verbs have nearly the same function, “remark” and “chuckle” convey more context and feelings than “said”. And if readers see too many “said” in the text, they would be bored.

E.g. 3 (from Chapter 70)

Hawkes' version: “Yes”, **said** Xiang-yun, “We founded it in the autumn, which is a time of decay. Perhaps that's why it didn't thrive. If we re-establish it now, when everything is burgeoning, it is bound to flourish! And this 'Flower of the Peach' is such a splendid poem: I think we ought to rename our club 'The Peach-flower Club'. What do the rest of you think?”

Yang's version: “We started the club in autumn”, **added** Hsiang-yun, “That's why it didn't prosper. If we start it again in spring when everything burgeons, it's bound to come to life. And this poem on peach-blossom is so good, why not change our Begonia Club into Peach-Blossom Club?”

The original text: 湘云笑道：“一起诗社时是秋天，就不应发达。如今却好万物逢春，皆主生盛。况这首桃花诗又好，就把海棠社改作桃花社。”

The original text wrote “笑道” (said with a smile), but translators often added their own understandings based on the context. Hawkes, again, chose “said” to lead the direct speech. Yang, instead, used “added” to imply Xiang-yun said these words after someone's speech.

From Table 3, it could be concluded that the standard frequency in Hawkes' work is mostly larger than that in Yang's work concerning each corresponding rank. We just mentioned that the two wordlists share many words, so if word frequency distributions are compared, another reason which lead to the overall difference in vocabulary richness might be found. Since the wordlists are too long, here we only choose words whose standard frequencies are over seven times. Appendix 1 displays the word list.

H_0: The word frequency distributions of the two translations are the same, that is, the two translators use the same word equally often.

H_1: The word frequency distributions of the two translations are not the same, that is, Hawkes uses a word more often than Yang in each corresponding word.

The Pearson Chi-Square test shows that there is a significant difference between word frequencies of Hawkes and word frequencies of Yang, $\chi^2(83, N = 12042) = 7080$, $p < 0.001$, Cramer's $\phi = 0.767$. This result implies that Hawkes did use words more frequently than Yang.

Besides those Top 50 words, the different background of the two translators is another reason that affects the vocabulary richness of texts. David Hawkes, a famous sinologist who had studied in Oxford and Beijing University, was familiar with Chinese traditional culture, which helped him in translating *Hongloumeng*. While Yang Xianyi, who was exposed to Chinese culture since he was born and went to Oxford for British literature study, must have a

deeper understanding toward the classic than Hawkes. This difference is shown through their translation strategies.

Hongloumeng, as a classical Chinese novel, abounds in Chinese cultural concepts. Li, Zhang & Liu (2011) found that Yang adopted a faithful and literal translation due to his primary goal of introducing the Chinese literature and culture to the English-speaking world. Usually, it is not easy to find corresponding words in English to express the same meaning of the cultural terms and allusions in Chinese, so a wide range of words were required to express these concepts and many of them were unusual and unfamiliar words in English. Hawkes, trying to make his translation easier to understand for English readers, went for free and fluent translation. When he met those cultural terms, he just left them out or converted them based on English culture. Just as he himself emphasized repeatedly in the book:

> . . . [T]he text abounds in passages containing references to books, plays, and poems which to the Western reader, lacking the literary background that Cao Xueqing was able to take for granted in his Chinese contemporaries, might often seem puzzling or incomprehensible (Hawkes 1979, p. 17).

Obviously, those culture-loaded words would not appear frequently in the corpora, so if hapax legomena in the wordlist were analyzed, some phenomena may be uncovered. We now turn to some words that occur once only in the wordlist of Yang's version but not in Hawkes' version. Concordance lines are analyzed together to help us understand those words in the specific context.

E.g. 1. Amida (from Chapter 20)

Hawkes' version: All I can say it's that I hope you marry a lisping husband, so that you have "I see, I see" in your ears every minute of the day. Ah, **Holy Name**! I think I can see that blessed day already before my eyes!

Yang's version: I just pray that you'll marry a husband who talks like me, so that you hear nothing but 'love' the whole day long. **Amida Buddha**! May I live to see that day!

The original text: 这一辈子我自然比不上你。我只保佑着明儿得一个咬舌的林姐夫，时时刻刻你可听阿弥陀佛，那才现在我眼里！

Xiang-yun (a cousin of bao-yu) said those words to rebut bao-yu. "阿弥陀佛" (Amida) in this sentence is a culture-based term in Buddhism and Yang translated it as "Amida Buddha", while Hawkes translated it as "Holy Name". Chinese readers could easily understand the meaning behind this term because most of them are familiar with Buddhism, but for English readers, they are more familiar with Christianity, so Hawkes replaced it by "Holy Name". In addition, Hawkes also tried to avoid other Buddhism terms throughout the translation.

E.g. 2. ai and erl (from Chapter 20)

Hawkes' version: Dai-yu burst out laughing: "**Lisping** doesn't seem to make you any less talkative! Listen to you: Cousin! Cousin Presently, when you're playing Racing Go, you'll be all sixth and seventh!"

Yang's version: "The lisper loves to rattle away," said Tai-yu with a laugh. "Fancy saying ***ai*** instead of ***er1*** like that. I suppose, when we start dicing, you'll be shouting one, love, three, four, five...."

The original text: 黛玉笑道："偏是咬舌子爱说话，连个'二'哥哥也叫不出来，只是'爱'哥哥'爱'哥哥的．回来赶围棋儿，又该你闹'幺爱三四五'了。"

Dai-yu makes fun of Xiang-yu here for Xiang pronounce "爱" (ai) instead of "二"(erl). Hawkes just used a "lisping" here to indicate Xiang's wrong articulation; while Yang, besides using "lisper" first, created *"ai"* and *"erl"*, which imitate the Chinese pronunciation, to present vivid and lively scene of the two cousins in the original text. This is easy to understand for those who know Chinese, but most English readers cannot get the point.

E.g. 3 Cowherd (Weaving Maid) (from Chapter 40)
Hawkes' version: On Seventh Night **the lovers** meet in heaven.
Yang's version: The **Weaving Maid** and **Cowherd** meet in Heaven.

The original text: 织女牛郎会七夕。

This is a game during drinking that all people should say something related to figures. Obviously, Yang created two figures "Weaving Maid" and "Cowherd" and did not give any figure in the translation, so only readers with some Chinese culture background should know the figure behind the sentence. But Hawkes just used "lovers" instead of the two names and explained the date exactly.

E.g. 4. incantation (from Chapter 120)
Hawkes' version: The second was when Bao-yu was seriously ill and the monk came and **said a prayer** over the jade, which seemed to cure Bao-yu at once.
Yang's version: The second time, when Pao-yu was so ill and the monk took the jade in his hand and **intoned some incantation** to cure him.

The original text: 第二次便是宝玉病重，他来了将那玉持诵了一番，宝玉便好了。

"诵" is also a culture-loaded word, which is, to some extent, similar to the witchcraft. This action, carried out by a Taoist priest, is of course related to Taoism, a religion originated from China. Yang translated it into "intoned some incantation", which can somewhat reveal the essence of this Chinese word. While for Hawkes, "said a prayer" seemed to be a better choice because it was easier for English readers to understand. But at the same time, it lost the original meaning conveyed by the author.

4. Conclusion

Based on the above analyses, we came to conclusions corresponding to the three research questions posed in the introduction section.

(1) Independent sample t-tests of the two indicators - STTR and lambda are consistent, which shows that there is significant difference in vocabulary richness of the two translations

and the vocabulary richness of Hawkes' version is lower than that of Yang's version. Though Hawkes is a native English speaker, he does not use higher vocabulary richness than Yang, which contradicts to our previous assumption.

(2) The source text was written in the mid of 18th century, but just like those texts translated from modern language, the vocabulary richness of the two translated *Hongloumeng* is also lower than native English.

(3) Further analyses through wordlists and concordance lines indicate that two main reasons explain their difference in vocabulary richness: (a) Hawkes used words more frequently concerning the same rank in the wordlists, and the Chi-square test shows that there is a significant difference between word frequency distributions of Hawkes and that of Yang. (b) Yang used more unfamiliar words in his version, especially in translating culture-loaded words.

In the paper, we have tried to explore the differences of vocabulary richness in the two translated *Hongloumeng* and do come to some conclusions. This study, however, still has its limitations. Firstly, due to the restrictions of time and space, we only choose 30 chapters out of the whole book, which may affect the accuracy of results. For further research, larger corpora could be built. Secondly, since there are many hapax legomena in the wordlist, finding culture-loaded words out of them is very time-consuming and we only manage to deal with four distinct words in this paper. Maybe we can find a better way to carry out this task in the future.

Acknowledgments

This work is partly supported by the National Social Science Foundation of China (Grant No. 11&ZD188).

References

Baker, M. (2000). Towards a Methodology for Investigating the Style of a Literary Translator. *Target*, 12(2), 241-266.

British National Corpus. Retrieved March 14, from http://www.natcorp.ox.ac.uk/

Chen, H. & Jiang, F. (2003). The translation of *Hong Lou Meng* into English: A descriptive study. *Chinese Translators Journal*, 5, 46–52.

Chen, Q. (2012). *A Corpus-based Critical Discourse Analysis of English Disaster News in America and China*. MA dissertation, Henan University of Science and Technology.

Chotlos J.W. (1944). A statistical and comparative analysis of individual written language samples. *Psychological Monographs*. 56: 75-111.

Corpas-Pastor, G. (2008). *Investigar con corpus en traducción: los retos de un nuevo paradigma*. Bern: Peter Lang.

Gellerstam, M. (1986). Translationese in Swedish Novels Translated from English. In: L. Wollin & H. Lindquist (Eds.), *Translation Studies in Scandinavia* (pp. 88-95). Lund: CWK Gleerup.

Hawkes, D. (Trans.) (1979). The Story of the Stone, *The Crab-Flower Club*, vol. 2. Bloomington: Indiana University Press.

Hoover, D.L. (2003). Another Perspective on Vocabulary Richness. *Computers and the Humanities*, 37, 151–178.

Jamak, A., Savatić, A., & Can, M. (2012). Principal component analysis for authorship attribution. *Business Systems Research*, 3(2), 49–56.

Jantunen, J.H. (2002). Synonymity and lexical simplification in translations: a corpus-based approach. *Across Languages and Cultures*, 2(1), 97–112.

Kojima, M., Yamashita, J. (2014). Reliability of lexical richness measures based on word lists in short second language productions. *System*, 42, 23–33.

Koizumi, R. & In'nami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System*, 40(4), 554–564.

Kubát, M. & Milička, J. (2013). Vocabulary Richness Measure in Genres. *Journal of Quantitative Linguistics*, 20(4), 339–349.

Kubát, M., Matlach, V & Čech, R. (2014). *Quantitative Index Text Analyzer*. Lüdenscheid: RAM-Verlag.

Laviosa, S. (1998). Core patterns of lexical use in a comparable Corpus of English narrative prose. *Meta*, 43(4), 557-570.

Li, D., Zhang, C., & Liu, K. (2011)Translation Style and Ideology: a Corpus-assisted Analysis of two English Translations of *Hongloumeng*. *Literary and Linguistic Computing*, 26(2), 153-166.

Lian, Z. (2014). A Parallel Corpus-Based Study of Emotion Words in *Hong Lou Meng* and Their Translations. *Canadian Social Science*, 10(1), 64-71.

Olahan, M. (2004). *Introducing Corpora in Translation Studies*. London: Routledge.

Popescu, I.-I., Čech, R. & Altmann, G. (2011). *The Lambda-structure of Texts*. RAM-Verlag.

Puurtinen, T. (2003). Genre-specific Features of Translationese? Linguistic Differences between Translated and Non-translated Finnish Childrens Literature. *Literary and Linguistic Computing*, 18(4), 389-406.

Qu, X. (2012). A Computational Stylistics Study on Two Translation Versions of *Six Chapters of a Floating Life*. MA thesis of Daliann Maritime University.

Scott, M. (2015). WordSmith Tools 6.0. Stroud, Gloucestershire, UK: Lexical Analysis Software Ltd. Available from http://www.lexically.net/wordsmith/

Smith, J. A. & Kelly, C. (2002). Stylistic Constancy and Change across Literary Corpora: Using Measures of Lexical Richness to Date Works. *Computers and the Humanities*, 36(4), 411-43**0**

Sotov, A. (2009). Lexical Diversity in a Literary Genre: A Corpus Study of the Rgveda. *Literary and Linguistic Computing* 24(4), 435-447.

Wimmer, G. & Altmann, G. (1991). Review Article: On Vocabulary Richness. *Journal of Quantitative Linguistics*, 6(1), 1-9.

Xiao, R. & Dai, G. (2014). Lexical and grammatical properties of Translational Chinese: Translation universal hypotheses reevaluated from the Chinese perspective. *Corpus Linguistics & Linguistic Theory*, 10 (1), 11-55.

Xiao, R. & Yue, M. (2009). "Using corpora in translation studies: The state of the art". In P. Baker (Ed.), *Contemporary Corpus Linguistics*. London: Continuum, 237–262.

Yule, G. U. (1944). *A Statistical Study of Literary Vocabulary*. Cambridge: University Press.

Zanettin, F. (2013). Corpus Methods for Descriptive Translation Studies. *Procedia - Social and Behavioral Sciences* 95, 20-32.

Appendix 1

Standard frequencies of shared words in the two translations

version	rank	Word	Frequency	Standard frequency
1	5	a	4212	200
2	4	a	2878	184
1	43	about	732	35
2	86	about	285	18
1	60	after	470	22
2	59	after	416	27
1	123	again	246	12
2	111	again	219	14
1	28	all	1044	50
2	34	all	700	45
1	106	am	293	14
1	192	am	148	7
1	75	an	407	19
2	106	an	235	15
1	3	and	6043	288
2	3	and	4291	274
1	178	another	162	8
2	189	another	117	7
1	109	any	278	13
2	140	any	169	11
1	190	anything	150	7
1	52	are	625	30
2	64	are	395	25
1	23	as	1222	58
2	20	as	991	63
1	174	ask	164	8
2	156	ask	151	10
1	124	asked	242	12
2	71	asked	344	22
1	27	at	1065	51
2	31	at	708	45
1	105	aunt	294	14
2	123	aunt	199	13
1	144	away	202	10
2	124	away	196	13
1	67	back	433	21
2	67	back	389	25
1	177	baochai	163	8
2	74	baochai	323	21
2	28	Pao-yu	753	48
1	48	Bao-yu	682	32
1	21	be	1310	62
2	27	be	789	50
1	184	because	155	7
2	171	because	131	8
1	51	been	634	30
2	77	been	312	20
1	139	before	213	10

2	134	before	180	11
1	141	better	210	10
2	125	better	196	13
1	25	but	1122	53
2	22	but	960	61
1	46	by	704	33
2	44	by	551	35
1	100	came	301	14
2	104	came	239	15
1	54	can	590	28
2	45	can	540	34
1	82	come	370	18
2	84	come	293	19
1	69	could	428	20
2	95	could	267	17
1	138	cousin	217	10
2	181	cousin	126	8
1	150	daiyu	196	9
2	76	daiyu	314	20
1	97	day	317	15
2	81	day	300	19
1	122	did	248	12
2	113	did	212	14
1	53	do	619	29
2	66	do	392	25
1	119	down	253	12
2	122	down	204	13
1	115	even	264	13
2	102	even	241	15
1	156	family	187	9
2	127	family	193	12
1	265	father	107	5
2	224	father	93	6
1	162	few	180	9
2	154	few	154	10
1	128	first	235	11
2	151	first	158	10
1	15	for	1889	90
2	15	for	1331	85
1	41	from	743	35
2	58	from	427	27
1	188	garden	153	7
2	180	garden	127	8
1	83	get	363	17
2	109	get	224	14
1	189	girl	151	7
2	163	girl	141	9
1	167	give	178	8
2	183	give	125	8
1	56	go	558	27
2	63	go	404	26
1	140	going	213	10
2	202	going	110	7
1	95	good	322	15

2	70	good	348	22
1	19	had	1474	70
2	19	had	1033	66
1	92	has	326	16
2	115	has	211	13
1	22	have	1248	59
2	26	have	790	50
1	14	he	2135	102
2	11	he	1828	117
1	169	heard	174	8
2	161	heard	144	9
1	10	her	2810	134
2	10	her	2016	129
1	87	here	341	16
2	73	here	324	21
1	29	him	1040	49
2	24	him	793	51
1	20	his	1315	63
2	21	his	989	63
1	170	home	170	8
2	147	home	162	10
1	86	how	343	16
2	62	how	405	26
1	7	I	3203	152
2	7	I	2200	140
1	32	if	941	45
2	29	if	750	48
1	9	in	2859	136
2	9	in	2037	130
1	77	into	400	19
2	143	into	165	11
1	26	is	1107	53
2	32	is	703	45
1	8	it	2867	136
2	12	it	1770	113
1	31	Jia	947	45
2	53	Chia	459	29
1	72	just	422	20
2	54	just	455	29
1	88	know	339	16
2	97	know	264	17
1	70	lady	425	20
2	33	lady	701	45
1	194	left	148	7
2	144	left	165	11
1	112	let	273	13
2	96	let	266	17
1	129	lian	234	11
2	172	lian	131	8
1	62	like	460	22
2	80	like	301	19
1	80	little	380	18
2	203	little	110	7
1	191	long	149	7

2	205	long	108	7
1	120	look	251	12
2	167	look	134	9
1	107	made	284	14
2	105	made	239	15
1	143	maids	207	10
2	135	maids	176	11
1	110	make	277	13
2	98	make	253	16
2	153	may	155	10
1	34	me	843	40
2	30	me	741	47
1	195	mind	148	7
2	176	mind	128	8
1	186	miss	154	7
2	198	miss	112	7
1	79	more	390	19
2	93	more	272	17
1	134	mother	224	11
2	132	mother	182	12
1	111	much	277	13
2	159	much	148	9
1	102	must	299	14
2	107	must	230	15
1	55	my	588	28
2	50	my	484	31
1	145	never	202	10
2	141	never	169	11
1	57	no	539	26
2	49	no	492	31
1	38	not	796	38
2	35	not	692	44
1	175	nothing	164	8
2	158	nothing	151	10
1	49	now	673	32
2	47	now	503	32
1	4	of	4714	224
2	6	of	2581	165
1	91	off	327	16
2	103	off	241	15
1	84	old	362	17
2	52	old	480	31
1	18	on	1480	70
2	18	on	1077	69
1	147	once	199	9
2	170	once	132	8
1	40	one	758	36
2	48	one	495	32
1	71	only	425	20
2	90	only	280	18
1	64	or	450	21
2	72	or	335	21
1	85	other	358	17
2	83	other	299	19

1	164	others	179	9
2	142	others	168	11
1	137	our	220	10
2	91	our	273	17
1	42	out	743	35
2	41	out	572	36
1	176	outside	164	8
2	184	outside	123	8
1	90	over	331	16
2	85	over	290	19
1	103	own	298	14
2	173	own	131	8
1	183	put	159	8
2	128	put	190	12
1	181	right	160	8
2	152	right	156	10
1	131	room	229	11
2	191	room	116	7
1	172	round	167	8
1	17	said	1596	76
2	46	said	530	34
1	113	say	268	13
2	150	say	159	10
1	76	see	407	19
2	92	see	273	17
1	180	sent	161	8
2	138	sent	172	11
1	12	she	2419	115
2	8	she	2048	131
1	89	should	334	16
2	129	should	190	12
1	37	so	815	39
2	25	so	793	51
1	61	some	461	22
2	56	some	448	29
1	153	something	191	9
2	194	something	114	7
1	118	still	255	12
2	130	still	188	12
1	165	such	179	9
2	131	such	188	12
1	116	take	262	12
2	94	take	270	17
1	121	tell	250	12
2	139	tell	172	11
1	158	than	184	9
2	155	than	152	10
1	11	that	2703	129
2	13	that	1673	107
1	1	the	8643	411
2	1	the	6273	400
1	65	their	444	21
2	75	their	321	20
1	33	them	864	41

2	40	them	583	37
1	63	then	460	22
2	39	then	586	37
1	39	there	773	37
2	51	there	483	31
1	154	these	190	9
2	114	these	212	14
1	30	they	999	48
2	23	they	854	54
1	130	things	231	11
2	148	things	161	10
1	94	think	323	15
2	169	think	133	8
1	24	this	1158	55
2	17	this	1247	80
1	133	thought	225	11
2	192	thought	116	7
1	78	time	393	19
2	88	time	281	18
1	2	to	6983	332
2	2	to	5247	335
1	151	told	195	9
2	100	told	247	16
1	101	too	300	14
2	60	too	414	26
1	161	took	181	9
2	137	took	175	11
1	74	two	412	20
2	89	two	281	18
1	47	up	687	33
2	43	up	554	35
1	117	us	259	12
2	117	us	207	13
1	99	Wang	308	15
2	101	Wang	242	15
1	159	want	182	9
2	179	want	128	8
1	13	was	2229	106
2	14	was	1410	90
1	104	way	295	14
2	121	way	205	13
1	45	we	720	34
2	36	we	646	41
1	114	well	266	13
2	116	well	211	13
1	96	went	319	15
2	79	went	303	19
1	50	were	673	32
2	55	were	454	29
1	36	what	826	39
2	42	what	564	36
1	35	when	837	40
2	38	when	601	38
1	166	where	179	9

2	136	where	176	11
1	93	which	325	15
2	126	which	194	12
1	125	while	242	12
2	120	while	206	13
1	68	who	433	21
2	61	who	411	26
1	126	why	237	11
2	78	why	309	20
1	73	will	422	20
2	118	will	207	13
1	16	with	1701	81
2	16	with	1300	83
1	58	would	525	25
2	87	would	282	18
1	132	xifeng	228	11
2	65	xifeng	394	25
1	6	you	3791	180
2	5	you	2830	181
1	148	young	198	9
2	112	young	217	14
1	44	your	728	35
2	37	your	641	41

Glottometrics 31, 2015, 76-84

Quantitative Studies in Chinese Language

Wei HUANG[1]

Abstract. The article contains a commented bibliography of Chinese quantitative studies in quantitative linguistic and textology. The English title, the Chinese transcription of the title and the Chinese title are presented

Keywords: Chinese, bibliography, quantitative linguistics

In the early years of 1980s, Zipf's and Herdan's work were introduced into China. Then a few Chinese researchers in linguistics and information science drew attentions to quantitative studies of languages, including theoretical studies of Zipf's law and its application to frequency distribution of Chinese characters and words.

In recent years, Chinese linguists in increasing numbers are focusing on quantitative linguistics. Both of the theories and the methods of modern quantitative linguistics are comprehensively introduced into China. Meanwhile, Chinese researchers are carrying out quantitative studies in lexicology, syntax, discourse analysis and other branches of linguistics. They have published some quantitative findings of Chinese (Mandarin), English, Russian and other languages.

The following are 32 articles published in Chinese language. Each item consists of 4 parts, the translated bibliographical information in English, the Romanized transliteration according to ISO-7098, the original Chinese information and a brief summary of the article. The bibliography is sorted by years of the publications are in ascending order.

Xiao Shensheng (1982). G. Herdan's stylo-statistics. *Language Study 2, 104-117.*
Xiao Shensheng (1982). G. Herdan de yanyu fengge tongjixue. Yuyan yanjiu *2, 104-117.*

萧申生. (1982). G.Herdan 的言语风格统计学. 语言研究 *2, 104-117.*

> The theory and method of stylo-staticstics are introduced from the first part of Herdan's book *Type-token Mathematics: A Textbook of Mathematical Linguistics.* It may be the first time when the Chinese researchers made acquaintance with quantitative linguistics.

Feng Zhiwei (1983). The origin and development of Zipf's law. *Information Science 4(2), 37-42.*
Feng Zhiwei (1983). Qipufu dinglü de lailongqumai. *Qingbao kexue 4(2), 37-42.*

[1] Beijing Language and Culture University: hwstudio@263.net

冯志伟. (1983). 齐普夫定律的来龙去脉. 情报科学 *4(2), 37-42.*

The derivation and development of Zipf-Mandelbrot law including the work by Estoup, Condon, Zipf, Joos and Mandelbrot are reviewed.

Shi Guiqing, Xu Bingzheng (1984). On the frequency distribution, optimum coding and input scheme for Chinese characters. *ACTA Electronica Sinica 12(4), 94-96.*

Shi Guiqing, Xu Bingzheng (1984). Hanzi zipin fenbu、zuijia bianma yu shuru wenti. *Dianzi xuebao 12(4), 94-96.*

石贵青、徐秉铮. (1984). 汉字字频分布、最佳编码与输入问题. 电子学报 *12(4), 94-96.*

The head part of rank-frequency distribution of Chinese characters in a corpus with size of one million characters can be fitted by the Zipf's law while the tail part approaches exponential distribution. These findings can be applied in evaluation of optimum code of Chinese characters and evaluation of input scheme.

Xu Wenxia (1986). Zipf's law and word frequency distribution of Chinese. *Information Science 7(1), 29-36.*

Xu Wenxia (1986). Qipufu dinglü yu zhongwen cipin fenbu jili. *Qingbao kexue 7(1), 29-36.*

许文霞. (1986). 齐普夫定律与中文词频分布机理. 情报科学 *7(1), 29-36.*

The word frequency distribution of a Chinese academic article follows Zipf’s law.

Cao Congsun (1987). Zipf's law and entropy of languages. *Journal of Tianjin Normal University 4, 80-85+73.*

Cao Congsun (1987). Qifu dinglü he yuyan de shang. *Tianjin shifan daxue xuebao 4, 80-85+73.*

曹聪孙. (1987). 齐夫定律和语言的熵. 天津师范大学学报 *4, 80-85+73.*

Zipf's law and the concept of entropy have been involved in discussion of language evolution. The author claims that Zipf's law shortens the length of language constituent while the entropy enlarges it macroscopically.

Wang Dejin (1988). The probability distribution and entropy in printed Chinese. *Journal of Beijing Institute of Aeronauties and Astronautics 4, 89-94.*

Wang Dejin (1988). Hanyu zi、ci de gailü fenbu he yi jie shang de yanjiu. *Beijing hangkong xueyuan xuebao 4, 89-94.*

王德进. (1988). 汉语字、词的概率分布和一阶熵的研究. 北京航空学院学报 *4, 89-94.*

The frequency distribution of characters and words in contemporary written Chinese does not obey Zipf's law. The entropy of Chinese words is larger than that of English words.

Wang Chongde, Lai Ling (1989). The Chinese collected work of Zipf's distribution. *Information Science 10(2), 1-8+42+79.*

Wang Chongde, Lai Ling (1989). Hanyu wenji de Qifu fenbu. *Qingbao kexue 10(2), 1-8+42+79.*

王崇德、来玲. (1989). 汉语文集的齐夫分布. 情报科学 *10(2), 1-8+42+79.*

The word frequency distribution of a Chinese academic article follows Zipf's law.

Zhao Laiyuan (1996). Fractal representation of Zipf's law. *Journal of the China Society for Scientific and Technical Information 15(4), 74-79.*
Zhao Laiyuan (1996). Qifu dinglü de fenxing tixian. *Qingbao xuebao 15(4), 74-79.*

赵来远. (1996). 齐夫定律的分形体现. 情报学报 *15(4), 74-79.*

The self-similarity of Zipfian distribution has been studied experimentally with the fractal method.

Chen Hailun (1996). A brief introduction to quantitative linguistics. *Journal of Yulin Teachers College (Philosophy and Socical Science) 17(1), 37-41+56.*
Chen Hailun (1996). Jiliang yuyanxue shuo lüe. *Yulin shizhuan xuebao 17(1), 37-41+56.*

陈海伦. (1996). 计量语言学说略. 玉林师专学报 *17(1), 37-41+56.*

The elementary theories and methods of quantitative linguistics, stylo-statistics and mathematical linguistics have been presented. And the early quantitative studies on Chinese, including word frequency and character frequency distribution, stylo-statistics, dialects and speech evolution, have been reviewed.

Guan Yi, Wang Xiaolong, Zhang Kai (1999).The frequency-rank relation of language units in modern Chinese computational language model. *Journal of Chinese Information Processing 13(2), 9-16.*
Guan Yi, Wang Xiaolong, Zhang Kai (1999). Xiandai hanyu jisuan yuyan moxing zhong yuyan danwei de pindu - pinji guanxi. *Zhongwen xinxi xuebao 13(2), 9-16.*

关毅、王晓龙、张凯. (1999). 现代汉语计算语言模型中语言单位的频度-频级关系. 中文信息学报 *13(2), 9-16.*

The exploration in Chinese corpus shows that the frequency distribution of Chinese constituents including characters, words and word bigrams follows Zipf's law. The authors claim that Zipf's law has great effect on many technologies of Chinese automatic processing, especially the construction of Chinese computational language model.

You Rongyan (2000). Zipf's law and the distribution of Chinese character frequency. *Journal of Chinese Information Processing 14(3), 60-65.*
You Rongyan (2000). Zipf dinglü yu hanzi zipin fenbu. *Zhongwen xinxi xuebao 14(3), 60-65.*

游荣彦. (2000). Zipf 定律与汉字字频分布. 中文信息学报 *14(3), 60-65.*

Zipf's law does not fit the whole frequency distribution of Chinese characters. And a method has been presented to describe only the tail of the distribution by using Zipf's law.

Jiang Wangqi (2005). Zipf and the principle of the least effort. *Journal of Tongji University (Social Science Section) 16 (1), 87-95.*

Jiang Wangqi (2005). Zipf yu shengli yuanze. *Tongji daxue xuebao (shehui kexue ban) 16(1), 87-95*.

姜望琪. (2005). Zipf 与省力原则. 同济大学学报(社会科学版) *16(1), 87-95.*

Zipf's law and the principle of the least effort, the modified Occam's Razor Principle, the Q-principle and R-principle, the Principle of Relevance have been theoretically discussed.

Fan Fengxiang (2006). Quantitative lexical description of marine engineering English. *Journal of Dalian Maritime University (Social Sciences Edition) 5(3), 161-164.*

Fan Fengxiang (2006). Lunji yingyu cihui de lianghua tezheng. *Dalian haishi daxue xuebao (shehui kexue ban) 5(3), 161-164.*

范凤祥. (2006). 轮机英语词汇的量化特征. 大连海事大学学报(社会科学版) *5(3), 161-164.*

This article presents investigations of the quantitative lexical characteristics in marine engineering English, including lexical density, zero order, word entropy and perplexity, coverage by the Chinese English Test (CET) Band 4 and CET Band 6 wordlists and the goodness of fit by the Herdan-Heaps model and other models.

Zheng Yabin, Liu Zhiyuan, Sun Maosong (2007). Statistical features of Chinese song lyrics and its application to retrieval. *Journal of Chinese Information Processing 21(5), 61-67.*

Zheng Yabin, Liu Zhiyuan, Sun Maosong (2007). Zhongwen geci de tongji tezheng ji qijiansuo yingyong. *Zhongwen xinxi xuebao 21(5), 61-67.*

郑亚斌、刘知远、孙茂松. (2007). 中文歌词的统计特征及其检索应用. 中文信息学报 *21(5), 61-67.*

The frequency distribution of words and characters in a Chinese lyrics corpus follow Zipf's law. And other experiments on Chinese lyrics in natural language processing including analysis based on time annotation, detecting the repetition of songs identifying rhythms, retrieving songs have been presented.

Gong Xiaoqing, Wang Zhan (2008). A note on Zipf's law. *Complex Systems and Complexity Science 5(3), 73-78.*

Gong Xiaoqing, Wang Zhan (2008). Guanyu Zipf lü de yidian zhuji. *Fuza xitong yu fuzaxing kexue 5(3), 73-78.*

龚小庆、王展. (2008). 关于 Zipf 律的一点注记. 复杂系统与复杂性科学 *5(3), 73-78.*

Based on numerical simulation and regression analysis, this article confirms that Zipf's law is statistically equivalent with the power-law distribution.

Liu Haitao (2008). Quantitative study of Chinese grammar based on dependency treebank. *Yangtze River Academic 3, 120-128.*

Liu Haitao (2008). Jiyu yicun shuku de hanyu jufa jiliang yanjiu. *Changjiang xueshu 3, 120-128.*

刘海涛. (2008). 基于依存树库的汉语句法计量研究. 长江学术 *3, 120-128.*

The syntactic characteristics of Chinese, including dependent distance and dependent direction, are quantitatively investigated based on 5 dependency treebanks of Chinese. The mean of dependent distance in Chinese is 2.84. And the percentage of the dependency relation of words which are not neighbors is between 40% and 50%. According to the dependency and the direction, Chinese is topologically a SV, VO and AdjN language.

Fan Fengxiang (2008). Inter-textual vocabulary repetition of marine engineering English. *Journal of Dalian Maritime University (Social Sciences Edition) 7(2). 128-132.*

Fan Fengxiang (2008). Lunji yingyu de pianji cihui chongfulü. *Dalian haishi daxue xuebao (shehui kexue ban) 7(2), 128-132*.

范凤祥. (2008). 轮机英语的篇际词汇重复率. 大连海事大学学报(社会科学版) *7(2), 128-132.*

The inter-textual vocabulary repetition of marine English has been examined based on a corpus with size of one million words. According to the Brunet's model of text length and vocabulary size, the inter-textual vocabulary repetition and its 95% confidence interval are calculated. And the 96% of the observed vocabulary repetitions is within the computed 95% confidence interval.

Huang Wei, Liu Haitao (2009). Application of quantitative characteristics of Chinese genres in text clustering. *Computer Engineering and Applications 45(29), 25-27*.

Huang Wei, Liu Haitao (2009). Hanyu yuti de jiliang tezheng zai wenben julie zhong de yingyong. *Jisuanji gongcheng yu yingyong 45(29), 25-27*.

黄伟、刘海涛. (2009). 汉语语体的计量特征在文本聚类中的应用. 计算机工程与应用 *45(29), 25-27.*

The method of applying the findings in quantitative study of linguistics for scrutinizing text clustering is presented. 16 linguistic structures , which are distributed distinctively between oral and written Chinese ,are investigated based on two sample corpora with size of half million words in each.

Wang Hui (2009). Polysemy: meaning, length and frequency. *Chinese Language 2, 120-130+191*.

Wang Hui (2009). Ciyi·cichang·cipin --《Xiandai hanyu cidian》(di 5 ban) duoyi ci jiliang fenxi. *Zhongguo yuwen 2, 120-130+191*.

王惠. (2009). 词义·词长·词频——《现代汉语词典》(第 5 版)多义词计量分析. 中国语文 *2, 120-130+191.*

The analysis of more than 10 thousand Chinese polysemy words extracted from *The Contemporary Chinese Dictionary* shows a strong correlation between polysemy and word frequency.

Luo Weihua, Deng Yaochen (2009). Pattern of inter-textual vocabulary repetition in English based on BNC. *Foreign Language Teaching and Research 41(3), 224-229.*

Luo Weihua, Deng Yaochen (2009). Jiyu BNC yuliaoku de yingyu pianji cihui chongfu moshi yanjiu. *Waiyu jiaoxue yu yanjiu 41(3), 224-229.*

罗卫华、邓耀臣. (2009). 基于 BNC 语料库的英语篇际词汇重复模式研究. 外语教学与研究 *41(3), 224-229.*

The inter-textual vocabulary repetition of written English is investigated based on the British National Corpus. And some inspirations in English vocabulary acquisition are discussed.

Wang Yang, Liu Yufan, Chen Qinghua (2009). Zipf distribution of words used in Chinese literature. *Journal of Beijing Normal University (Natural Science) 45(4), 424-427.*

Wang Yang, Liu Yufan, Chen Qinghua (2009). Hanyuyan wenxue zuopin zhong cipin de Zipf fenbu. *Beijing shifan daxue xuebao (ziran kexue ban) 45(4), 424-427.*

王洋、刘宇凡、陈清华. (2009). 汉语言文学作品中词频的 Zipf 分布. 北京师范大学学报(自然科学版) *45(4), 424-427.*

The word frequency distributions in Chinese literature, including *Dream of Red Chamber*, *Selected Works of Mao Tse-tung* and *Selected Works of Deng Xiaoping*, follow Zipf's law.

Deng Yaochen, Li Bingbing (2010). Study of English compounding propensity in synergistic linguistics. *Foreign Languages and Their Teaching 4, 19-24.*

Deng Yaochen, Li Bingbing (2010). Yingyu fuheci shengcheng quxiang de xietong yuyanxue yanjiu. *Waiyu yu waiyu jiaoxue 4, 19-24.*

邓耀臣、李冰冰. (2010). 英语复合词生成趋向的协同语言学研究. 外语与外语教学 *4, 19-24.*

The investigation in technical English text based on the Jiao Da English for Science and Technology corpus confirms that both the relationship between compounding propensity and stem frequency, and the relationship between compounding propensity and length of stem follow $\square = \square\square^{\square}$, and that the relationship between compounding propensity and polylexy follows $\square = \square + \square\square^{\square}$.

Wang Hua, Gulila Altenbek (2010).A corpus–based study on frequency statistics of Kazak words. *Computer Engineering 36(24), 59-61.*

Wang Hua, Gulila Altenbek (2010). Jiyu yuliao de hasakeyu cipin tongji yanjiu. *Jisuanji gongcheng, 36(24), 59-61.*

王花、古丽拉·阿东别克. (2010). 基于语料的哈萨克语词频统计研究. 计算机工程 *36(24), 59-61.*

The word frequency distribution in a Kazak corpus with size of 300 thousand words follows Zipf's law.

Tang Lian, Wang Dahui (2011). An explanation of shift parameter ρ in Zipf-Mandelbrot's law. *Journal of Beijing Normal University (Natural Science) 47(1), 97-100.*

Tang Lian, Wang Dahui (2011). Guanyu Zipf-Mandelbrot lü zhong canshu ρ de yizhong jieshi. *Beijing shifan daxue xuebao (ziran kexue ban) 47(1), 97-100.*

唐莲、王大辉. (2011). 关于 Zipf-Mandelbrot 律中参数 ρ 的一种解释. 北京师范大学学报(自然科学版) *47(1), 97-100.*

The experiment of simulating random data series following Zipf-Mandelbrot's law reveals that the parameter ρ in the formula $S \propto (r + \rho)^{\square\square}$ reflects saturation effect of sampled data.

Xia Huiyan, Sun Fenglan (2011). Quantitative relationship between word length and polysemy in English. *Foreign Languages and Their Teaching 3, 44-49.*

Xia Huiyan, Sun Fenglan (2011). Yingyu cihui changdu yu cihui yiyi guanxi de jiliang yuyanxue yanjiu. *Waiyu yu waiyu jiaoxue 3, 44-49.*

夏慧言、孙凤兰. (2011). 英语词汇长度与词汇意义关系的计量语言学研究. 外语与外语教学 *3, 44-49.*

The investigation in the British National Corpus shows that the relationship between polysemy and word length follows the power model $\square = \square\square^{\square\square}$, while the relationship between synonymy and word length does not follow the model $\square = \square\square^{\square}\square^{\square\square}$, but the model $\square = \square + \square\square + \square\square^{\square} + \square\square^{\square}$ does well.

Lu Gaofei, Han Pu, Shen Si (2012). Comparative empirical study on Zipf's law with two fitting methods. *Library and Information Service 56(24), 71-76+126.*

Lu Gaofei, Han Pu, Shen Si (2012). Liang zhong Zipf dinglü nihe fangfa de duibi shizheng yanjiu. *Tushu qingbao gongzuo 56(24), 71-76+126.*

路高飞、韩普、沈思. (2012). 两种 Zipf 定律拟合方法的对比实证研究. 图书情报工作 *56(24), 71-76+126.*

The comparison of Ordinary Least Square and Maximum Likelihood Estimation in fitting of the distribution of Zipf's law with 6 corpora, including 3 Chinese ones and 3 English ones, shows that the Maximum Likelihood Estimation is much better.

Feng Zhiwei (2012). Studying language by quantitative method. *Foreign Language Teaching and Research 44(2), 256-269+321.*

Feng Zhiwei (2012). Yong jiliang fangfa yanjiu yuyan. *Waiyu jiaoxue yu yanjiu 44(2), 256-269+321.*

冯志伟. (2012). 用计量方法研究语言. 外语教学与研究 *44(2), 256-269+321.*

The relationship of quantitative linguistics and mathematical linguistics is discussed, and Zipf's law, the Mentherath-Altmann law, the Fucks-Čebanov law and the Piotrowski-Altmann law are reviewed.

Liu Haitao, Huang Wei (2012). Quantitative linguistics: state of the art, theories and methods. *Journal of Zhejiang University (Humanities and Social Sciences) 42(2), 178-192.*

Liu Haitao, Huang Wei (2012). Jiliang yuyanxue de xianzhuang、lilun yu fangfa. *Zhejiang*

daxue xuebao (renwen shehui kexue ban) 42(2), 178-192.

刘海涛、黄伟. (2012). 计量语言学的现状、理论与方法. 浙江大学学报(人文社会科学版) *42(2), 178-192.*

This article reviews the history and development of quantitative linguistics, introduces the classical theories and achievements, and discusses the methodology of quantitative linguistics. It is the first time that the quantitative linguistics and synergetic linguistics are comprehensively introduced into Chinese.

Dong Hong (2012). Inter-textual vocabulary growth patterns for Shakespearean drama. *Journal of Qiqihar University (Philosophy and Social Science) 2, 125-127.*

Dong Hong (2012). Shashibiya xiju de pianji cihui zengzhang moshi. *Qiqihar daxue xuebao (zhexue shehui kexue ban) 2, 125-127.*

董红. (2012). 莎士比亚戏剧的篇际词汇增长模式. 齐齐哈尔大学学报(哲学社会科学版) *2, 125-127.*

The investigation in the Shakespearean Play Corpus shows that the inter-textual vocabulary growth of the hapax legomena follows the Brunet's model $V = V_1(N_1 N)^{a}$.

Deng Yaochen, Feng Zhiwei (2013). A quantitative linguistic study on the relationship between word length and word frequency. *Journal of Foreign Lanugages 36(3), 29-39.*

Deng Yaochen, Feng Zhiwei (2013). Cihui changdu yu cihui pinshuo guanxi de jiliang yuyanxue yanjiu. *Waiguoyu (Shanghai waiguoyu daxue xuebao) 36(3), 29-39.*

邓耀臣、冯志伟. (2013). 词汇长度与词汇频数关系的计量语言学研究. 外国语(上海外国语大学学报) *36(3), 29-39.*

The investigation in the Lancaster Corpus of Mandarin Chinese and the Spoken Corpus of Mandarin Chinese shows that the relationship between word length and word frequency in Chinese texts follows the power law $y = ax^{b}$, and the parameter *a* is distinguishable between oral speech and written texts.

Wang Yong, Liu Haitao (2013). Quantitative properties of Russian nouns. *Journal of Zhejiang University (Humanities and Social Sciences) 43(6), 174-186.*

Wang Yong, Liu Haitao (2013). Eyu mingci de jiliang tezheng yanjiu. *Zhejiang daxue xuebao (renwen shehui kexue ban) 43(6), 174-186.*

王永、刘海涛. (2013). 俄语名词的计量特征研究. 浙江大学学报(人文社会科学版) *43(6), 174-186.*

The quantitative investigation in a Russian corpus shows that 1) the Russian is a dependent-final language, 2) the major types of nominal structures in Russian are concordant attribute and non-concordant attribute relations, and 3) the word orders in the Russian nominal structures tend to follow certain patterns.

Zheng Chen, Hu Manfeng (2013). Zipf's law in words and characters in Moyan's literature. *Journal of Jiangnan University (Natural Science Edition) 12(3), 347-350.*

Zheng Chen, Hu Manfeng (2013). Moyan zuopin zhong zipin, cipin de Zipf fenbu. *Jiangnan daxue xuebao (ziran kexue ban) 12(3), 347-350.*

郑晨、胡满峰. (2013). 莫言作品中字频、词频的 Zipf 分布. 江南大学学报(自然科学版) *12(3), 347-350*.

The word and the character frequency distribution in 3 novels by Mo Yan, who is a Nobel Prize winner for literature, follow Zipf's law.

Acknowledgement

This work was supported by the Municipal Social Science Foundation of Beijing (12WYC032) and the National Social Science Foundation of China (13CYY022).

Glottometrics 31, 2015, 85-88

Bibliography of Quantitative Linguistics of Chinese Reseachers in International Academic Journals

Ruina Chen[1]

Chinese researchers began to take an increasingly important role in the international academic journals. We will briefly introduce their contributions, which are published in English and are related to quantitative linguistics. Literatures are arranged according to the alphabetical name of the first author, and are mainly from the *Journal of Quantitative Linguistics*, *Glottometrics*, *Glottotheory* and some others.

Chen Heng and Liu Haitao (2014). A diachronic study of Chinese word length distribution. *Glottometrics* 29: 81-94.

Chen Ruina and Liu Haitao (2014). Quantitative Aspects of Journal of Quantitative Linguistics. *Journal of Quantitative Linguistics* 21(4): 299-340.

Chen Qinghua, Guo Jinzhong and Liu Yufan (2012). A Statistical Study on Chinese Word and Character Usage in Literatures from the Tang Dynasty to the Present. *Journal of Quantitative Linguistics* 19(3): 232-248.

Cong Jin and Liu Haitao (2014). Approaching human language with complex networks. *Physics of Life Reviews* 4: 598-618.

Cong Jin and Liu Haitao (2014). Linguistic complex networks: rationale, application, interpretation, and directions. Reply to comments on "Approaching human language with complex networks". *Physics of Life Reviews*. 4: 644-649.

Cong Jin (2011). Quantitative Linguistics in China. *Glottometrics* 21: 85.

Deng Yaochen (2009). A Synergetic Linguistic study on compounding propensity in technical English. *Glottotheory* 2 (2): 54-61.

Fan Fengxiang (2013). Text Length, Vocabulary Size and Text Coverage Constancy. *Journal of Quantitative Linguistics* 20(4): 288-300.

Fan Fengxiang (2012). A Quantitative Study on the Lexical Change of American English. *Journal of Quantitative Linguistics* 19(3): 171-180.

Fan Fengxiang, Grzybek Peter and Altmann Gabriel (2010). Dynamics of word length in sentence. *Glottometrics* (20): 70-109.

Fan Fengxiang and Altmann Gabriel (2008). On meaning diversification in English. *Glottometrics* (17):66-78.

[1] Zhejiang University, Hangzhou, P. R. China, chenruina@sina.com

Fan Fengxiang, Popescu Ioan-Iovitz and Altmann Gabriel (2008). Arc length and meaning diversification in English. *Glottometrics* (17):79-86.

Fan Fengxiang (2006). A Corpus-based empirical study on inter-textual vocabulary growth. *Journal of Quantitative Linguistics* 13(1): 111-127.

Fan Fengxiang (2006). Models for dynamic inter-textual type-token relationship. *Glottometrics* (12):1-10.

Ferrer-i-Cancho Ramon and Liu Haitao (2014). The risks of mixing dependency lengths from sequences of different length. *Glottotheory* (2): 143-155.

Gao Song, Zhang Hongxin and Liu Haitao (2013). Synergetic Properties of Chinese Verb Valency. *Journal of Quantitative Linguistic*s 21(1): 1-21.

Gao Yuyang, Liang Wei, Shi Yuming and Huang Qiuling (2014). Comparison of directed and weighted co-occurrence networks of six languages. *Physica A: Statistical Mechanics and its Applications* 393(1): 579–589.

Hou Renkui, Jiang Yang and Jiang Minghu (2014). A study on Chinese quantitative stylistic features and relation among different styles based on text clustering. *Journal of Quantitative Linguistics* 21(3), 246-280.

Ji Meng (2009). Corpus Statistics in Contrastive Literary Studies. *Journal of Quantitative Linguistics* 16(3): 243-255.

Ke Jinyun and Yao Yao (2008). Analysing Language Development from a Network Approach. *Journal of Quantitative Linguistics* 15(1): 70-99.

Ke Jinyun (2006). A cross-linguistic quantitative study of homophony. *Journal of Quantitative Linguistics* 13(1): 129-159.

Lei Lei and Liu Zhehua (2013). A Word Type-based Quantitative Study on the Lexical Change of American and British English. *Journal of Quantitative Linguistics* 21(1): 36-49.

Li Bin, Zhang Caicai and Wayland Ratree (2012). Acoustic Characteristics and Distribution of Variants of /l/ in the Nanjing Dialect. *Journal of Quantitative Linguistics* 19(4): 281-300.

Li Jianyu, Zhou Jie, Luo Xiaoyue and Yang Zhanxin (2012). Chinese lexical networks: the structure, function and formation. *Physica A: Statistical Mechanics and its Applications* 391(21):5254-5263.

Li J**ianyu and Zhou Jie** (2007). Chinese character structure analysis based on complex networks. *Physica A: Statistical Mechanics and its Applications* 380: 629-38.

Li Wentian (2013). Characterizing Ranked Chinese Syllable-to-Character Mapping Spectrum: A Bridge between the Spoken and Written Chinese Language. *Journal of Quantitative Linguistics* 20(2): 153-167.

Li Wentian and Miramontes Pedro (2011). Fitting Ranked English and Spanish Letter Frequency Distribution in US and Mexican Presidential Speeches. *Journal of Quantitative Linguistics* 18(4): 359-380.

Li Yong, Wei Luoxia, Niu Yi and Yin Junxun (2005). Structural organization and scale-free properties in Chinese Phrase Networks. *Chinese Science Bulletin*. 50(13):1305-1308.

Li Yong, Wei Luoxia, Li Wei, Niu Yi and Luo Shiyu (2005). Small-world patterns in Chinese phrase networks. *Chinese Science Bulletin*. 50(3): 286—288.

Liang Junying and Liu Haitao (2013). Noun Distribution in Natural Languages. *Poznań*

Studies in Contemporary Linguistics 49(4): 509-529.

Liang Wei, Shi Yuming, Tse Chi K. and Wang Yanli (2012). Study on co-occurrence character networks from Chinese essays in different periods. *Science China Information Sciences* 55(11): 2417-2427.

Liang Wei, Shi Yuming, Tse Chi K, Liu Jing, Wang Yanli and Cui Xunqiang (2009). Comparison of co-occurrence networks of the Chinese and English languages. *Physica A: Statistical Mechanics and its Applications* 388(23): 4901–4909.

Liang Wei, Shi Yuming, Tse Chi K, and Wang Yanli (2008). Structural Equivalence between co-occurrences of characters and words in the Chinese language. In *International Symposium on Nonlinear Theory and its Applications:* 94-97.

Liu Haitao and Cong Jin (2014). Empirical characterization of modern Chinese as a multi-level system from the complex network approach. *Journal of Chinese Linguistics*. (42):1-38.

Liu Haitao and Cong Jin (2013). Language clustering with word co-occurrence networks based on parallel texts. *Chinese Science Bulletin* 58(10): 1139-1144.

Liu Haitao and Xu Chunshan (2012). Quantitative typological analysis of Romance languages. *Poznań Studies in Contemporary Linguistics*. 48(4): 597-625.

Liu Haitao (2011). Quantitative analysis of Zamenhof's Esenco kaj estonteco. *Language Problems and Language Planning* 35(1): 57-81.

Liu Haitao and Xu Chunshan (2011). Can syntactic network indicate morphological complexity of a language? *EPL*. 93(2), 28005.

Liu Haitao (2011). Quantitative Properties of English Verb Valency. *Journal of Quantitative Linguistics* 18(3): 207-233.

Liu Haitao (2010). Dependency direction as a means of word-order typology: A method based on dependency treebanks. *Lingua* 120: 1567-1578.

Liu Haitao, Zhao Yiyi and Huang Wei (2010). How do Local Syntactic Structures Influence Global Properties in Language Networks? *Glottometrics* (20): 38-58.

Liu Haitao and Li Wenwen (2010). Language Clusters based on Linguistic Complex Networks. *Chinese Science Bulletin* 55(30): 3458-3465.

Liu Haitao (2009). Probability Distribution of Dependencies Based on a Chinese Dependency Treebank. *Journal of Quantitative Linguistics* 16(3): 256-273.

Liu Haitao (2009). Statistical properties of Chinese semantic networks. *Chinese Science Bulletin*. 54(16):2781-2785.

Liu Haitao (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science* 9(2):159-191.

Liu Haitao and Hu Fengguo (2008). What role does syntax play in a language network? *EPL*, 83.18002.

Liu Haitao (2008). The complexity of Chinese syntactic dependency networks. *Physica A: Statistical Mechanics and its Applications* 387(12): 3048–3058.

Liu Haitao (2007). Probability distribution of dependency distance. *Glottometrics* (15): 1-12.

Ming Delie, Liu Jian and Tian Jinwen (2002). The Design and Implementation of a Chinese Financial Invoice Recognition System. *Journal of Quantitative Linguistics* 9(1): 19-33.

Narisong, Jiang Jingyang and Liu Haitao (2014). Word Length Distribution in Mongolian. *Journal of Quantitative Linguistics* 21(2), 123-152.

Pan Xiaxing and Liu Haitao (2014). Adnominals in Modern Chinese and their Distribution Properties. *Glottometrics* 29: 1-30.

Peng Gang, Minett James W. and Wang S Y William (2008).The networks of syllables and characters in Chinese. *Journal of Quantitative Linguistics* 15(3): 243-255.

Qiao Wei, Sun Maosong and Menzel Wolfgang (2010). Chinese Word Frequency Approximation Based on Multitype Corpora. *Journal of Quantitative Linguistics* 17(2): 142- 166.

Sheng Long and Li Chunguang (2009). English and Chinese languages as weighted complex networks. *Physica A: Statistical Mechanics and its Applications*. 388(12): 2561–2570.

Shi Daming, Damper Robert I and Gunn Steve R (2003). An Approach to Off-Line Handwritten Chinese Character Recognition Based on Hierarchical Radical Decomposition. *Journal of Quantitative Linguistics* 10(1): 41-69.

Wang Hua and Liu Haitao (2014). The Effects of Length and Complexity on Constituent Ordering in Written English. *Poznań Studies in Contemporary Linguistics* 50(4): 477–494.

Wang Hua (2012). Length and complexity of NPs in Written English. *Glottometrics* (24): 79-87.

Wang Lu (2014). Synergetic Studies on Some Properties of Lexical Structures in Chinese. *Journal of Quantitative Linguistics* 21(2), 177-197.

Wang Lu (2011). Polysemy and word length in Chinese. *Glottometrics* (21): 73-84.

Yang Chih-Chien and Yang Chih-Chiang (2013). Categorical Latent Interactions of Multinomial Linguistic Components. *Journal of Quantitative Linguistics* 20(2): 105-124.

Yang Yuting, Qu Yunhua, Bao Chenyao and Zhang Xiaowen (2015). A Model-based Feature Optimization Approach to Chinese Language Processing. *Journal of Quantitative Linguistics* 22 (1), 55-81.

Yu Shuiyuan, Liu Haitao and Xu Chunshan (2011). Statistical properties of Chinese phonemic networks. *Physica A: Statistical Mechanics and its Applications* 390(7): 1370-1380.

Yu Shuiyuan, Xu Chunshan, Liu Haitao and Chen Yudong (2011). Statistical analysis of Chinese phonemic contrast. *Phonetica*. 68(4): 201-214.

Yue Ming and Liu Haitao (2011). Probability Distribution of Discourse Relations Based on a Chinese RST-annotated Corpus. *Journal of Quantitative Linguistics* 18(2): 107-121.

Zhao Xiaodong and Feng Zhiwei (2013). A Dynamic Study of English Intertextual Lexical Repetition Rates. *Journal of Quantitative Linguistics* 21(1): 65-84.

Zhao Xiaodong (2012). Vocabulary growth of content words of ESP and General English. A contrastive study based on CMTE and BNC. *Glottometrics* (24): 12-24.

Zhang Yanhui (2011). Similarity and Degree of Perplexity Analysis of Chinese Characters. *Journal of Quantitative Linguistics* 18(3): 189-206.

Zhang Zhiyi and Huang Hongwei (2007). Turing's formula revisited. *Journal of Quantitative Linguistics* 14(2-3): 222-241.

Zhou Shuigeng, Hua Guobiao, Zhang Zhongzhi and Guan Jihong (2008). An empirical study of Chinese language networks. *Physica A: Statistical Mechanics and its Applications* 387(12):3039–3047.

Glottometrics 1 – 30

Using the jubilee of *Glottometrics*, we are glad to present a complete bibliography of all publications of the first 30 issues. The contributions are ordered in 5 sections: (1) General articles, (2) History, (3) Reviews, (4) Bibliographies, and (5) Miscellanea. Within each of these sections, the contributions are ordered according to authors' names and year of publication. The Bibliography can be downloaded as PDF-file from: http://www.ram-verlag.eu/wp-content/uploads/2015/03/biblioglo-1-30.pdf

Peter Grzybek, Emmerich Kelih

1. General Articles

Adamic, Lada A.; Huberman, Bernardo A. (2002). Zipf's law and the Internet. *Glottometrics,* 3; 143–150.

Altmann, Gabriel (2002). Zipfian Linguistics. *Glottometrics,* 3; 19–26.

Altmann, Gabriel (2004). Script complexity. *Glottometrics,* 8; 68–74.

Altmann, Gabriel (2014). On morphological complexity in Indonesian. *Glottometrics,* 29; 59–69.

Altmann, Gabriel; Popescu, Ioan-Iovitz; Zotta, Dan (2013). Stratification in texts. *Glottometrics,* 25; 85–93.

Andersen, Simone (2002). Freedom of choice and psychological interpretation of word frequencies in texts. *Glottometrics,* 2; 45–52.

Andersen, Simone (2002). Speaker's information content: frequency-length correlation as partial correlation. *Glottometrics,* 3; 90–109.

Andersen, Simone (2005). Word length balance in texts: Proportion constancy and word-chain-lengths in Proust's longest sentence. *Glottometrics,* 11; 32–50.

Andreev, Sergey (2007). Some properties of a metalinguistic verbal system (in the metalanguage of the Macmillan English Dictionary's defining vocabulary). *Glotto-metrics,* 14; 21–31.

Andreev, Sergey (2012). Literal vs. Liberal Translation – Formal Estimation. *Glottometrics,* 23; 62–69.

Andres, Jan; Benešová, Martina (2011). Fractal analysis of Poe's *Raven*. *Glottometrics,* 21; 73–98.

Antić, Gordana; Altmann, Gabriel (2005). On letter distinctivity. *Glottometrics,* 9; 46-53.

Araujo, Leonardo C.; Cristófaro-Silva, Thaïs; Yehia, Hani Camille (2013). Entropy of a Zipfian Distributed Lexicon. *Glottometrics,* 26; 38–49.

Baevskij, Vadim S. (2011). Academician Andrey Nikolaevich Kolmogorov as a Scholar of Verse Theory. *Glottometrics,* 22; 17–43.

Baevskij, Vadim S. (2012). Non-traditional approach to the study of the rhythmics of Russian verse. *Glottometrics,* 24; 5–11.

Balasubrahmanyan, Viddhachalam K.; Naranan, Sundaresan (2002). Algorithmic information, complexity and Zipf's law. *Glottometrics,* 4; 1–26.

Beliankou, Andrei; Köhler, Reinhard (2010). The distribution of parts-of-speech in Russian texts. *Glottometrics,* 20; 59–69.

Best, Karl-Heinz (2001). Zur Gesetzmäßigkeit der Wortartenverteilungen in deutschen Pressetexten. *Glottometrics,* 1; 1–26.

Best, Karl-Heinz (2002). Der Zuwachs der Wörter auf **-***ical* im Deutschen. *Glottometrics,* 2; 11–16.

Best, Karl-Heinz (2002). The distribution of rhythmic units in German short prose. *Glottometrics,* 3; 136–142.

Best, Karl-Heinz (2003). Spracherwerb, Sprachwandel und Wortschatzwachstum in Texten. Zur Reichweite des Piotrowski-Gesetzes. *Glottometrics,* 6; 9–34.

Best, Karl-Heinz (2004). Zur Ausbreitung von Wörtern arabischer Herkunft im Deutschen. *Glottometrics,* 8; 75–78.

Best, Karl-Heinz (2005). Sprachliche Einheiten in Textblöcken. *Glottometrics,* 9; 1–12.

Best, Karl-Heinz (2005). Turzismen im Deutschen. *Glottometrics,* 11; 56–63.

Best, Karl-Heinz (2005). Zur Häufigkeit von Buchstaben, Leerzeichen und anderen Schriftzeichen in deutschen Texten. *Glottometrics,* 11; 9–31.

Best, Karl-Heinz (2006). Deutsche Entlehnungen im Englischen. *Glottometrics,* 13; 66–72.

Best, Karl-Heinz (2006). Gesetzmäßigkeiten im Erstspracherwerb. *Glottometrics,* 12; 39–54.

Best, Karl-Heinz (2006). Wie viele Morphe enthalten Wörter in deutschen Pressetexten? *Glottometrics,* 13; 47–58.

Best, Karl-Heinz (2007). Quantitative Untersuchungen zum deutschen Wörterbuch. *Glottometrics,* 14; 32–45.

Best, Karl-Heinz (2007). Zur Entwicklung des Wortschatzes der Elektrotechnik, Informationstechnik und Elektrophysik im Deutschen. *Glottometrics,* 15; 24–27.

Best, Karl-Heinz (2008). Das Fremdwortspektrum im Türkischen. *Glottometrics,* 17; 8–11.

Best, Karl-Heinz (2008). Sinismen im Deutschen und Englischen. *Glottometrics,* 17; 87–93.

Best, Karl-Heinz (2008). Word length in Persian. *Glottometrics,* 16; 27–30.

Best, Karl-Heinz (2008). Zur Diversifikation lateinischer und griechischer Hexameter. *Glottometrics,* 17; 43–50.

Best, Karl-Heinz (2009). Diversifikation des Phonems /r/ im Deutschen. *Glottometrics,* 18; 26–31.

Best, Karl-Heinz (2009). Wortlängen im Englischen. *Glottometrics,* 19; 1–10.

Best, Karl-Heinz (2009). Zur Entwicklung der Entlehnungen aus dem Japanischen ins Deutsche. *Glottometrics,* 19; 80–84.

Best, Karl-Heinz (2010). Zur Entwicklung des Wortschatzes der deutschen Umgangssprache. *Glottometrics,* 20; 34–37.

Best, Karl-Heinz (2011). Diversification of a single sign of the Danube script. *Glottometrics,* 22; 1–4.

Best, Karl-Heinz (2011). Silben-, Wort- und Morphlängen bei Lichtenberg. *Glottometrics,* 21; 1–13.

Best, Karl-Heinz (2011). Word length distribution in French. *Glottometrics,* 22; 57–61.

Best, Karl-Heinz (2012). Diversifikation der starken Verben im Deutschen. *Glottometrics,* 24; 1–4.

Best, Karl-Heinz (2012). Längen von Komposita im Deutschen. *Glottometrics,* 23; 1–6.

Best, Karl-Heinz (2012). Zur Verslänge bei G.A. Bürger. *Glottometrics,* 23; 56–61.

Best, Karl-Heinz (2013). Iranismen im Deutschen. *Glottometrics,* 26; 1–8.

Best, Karl-Heinz (2013). Zur Verslänge im Altisländischen. *Glottometrics,* 25; 22–29.

Best, Karl-Heinz (2014). Hebraismen im Deutschen. *Glottometrics,* 27; 10–17.

Best, Karl-Heinz; Altmann, Gabriel (2005). Some properties of graphemic systems. *Glottometrics,* 9; 29–39.

Best, Karl-Heinz; Zhu, Jinyang (2010). Ein Modell für die Zunahme chinesischer Schriftzeichen. *Glottometrics,* 20; 29–33.

Brüers, Nina; Heeren, Anne (2004). Pluralallomorphe in Briefen Heinrich von Kleists. *Glottometrics,* 7; 85–90.

Buk, Solomija; Mačutek, Ján; Rovenchak, Andrij (2008). Some properties of the Ukrainian writting system. *Glottometrics,* 16; 63–79.

Čech, Radek (2014). Four reasons for a revision of the transitivity hypothesis. *Glottometrics,* 27; 1–9.

Čech, Radek; Mačutek, Ján (2009). Word form and lemma syntactic dependency networks in Czech: a comparative study. *Glottometrics,* 19; 85–98.

Čech, Radek; Popescu, Iovitz; Altmann, Gabriel (2011). Euphony in Slovak lyric poetry. *Glottometrics,* 22; 5–16.

Čech, Radek; Popescu, Iovitz; Altmann, Gabriel (2011). Word length in Slovak poetry. *Glottometrics,* 22; 44–56.

Chen, Ruina; Altmann, Gabriel (2015). Conceptual inertia in texts. *Glottometrics,* 30; 73–85.

Chen, Heng; Liu, Haitao (2014). A diachronic study of Chinese word length distribution. *Glottometrics,* 29; 81–94.

Dębowski, Łukasz (2002). Zipf's law against the text size: a half-rational model. *Glottometrics,* 4; 49–60.

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

Dzhuriuk, Tetiana (2006). Sentence length as a feature of style (applied to works of German writers). *Glottometrics,* 12; 55–62.

Eeg-Olofsson, Mats (2009). A word length regularity and its genesis. *Glottometrics,* 19; 49–69.

Fengxiang, Fan (2006). Models for dynamic inter-textual type-token relationship. *Glottometrics,* 12; 1–10.

Fengxiang, Fan; Altmann, Gabriel (2008). On meaning diversification in English. *Glottometrics,* 17; 66–78.

Fengxiang, Fan; Grzybek, Peter; Altmann, Gabriel (2010). Dynamics of word length in sentence. *Glottometrics,* 20; 70–109.

Fengxiang, Fan; Popescu, Ioan-Iovitz; Altmann, Gabriel (2008). Arc length and meaning diversification in English. *Glottometrics,* 17; 79–86.

Fengxiang, Fan; Yaquin, Wang; Zhao, Gao (2014). Some macro quantitative features of low-freqeuency word classes. *Glottometrics,* 28; 1–12.

Fenk-Oczlon, Gertraud; Fenk, August (2002). Zipf's Tool Analogy and Word Order. *Glottometrics,* 5; 22–28.

Ferrer-i-Cancho, Ramon (2005). Hidden communication aspects in the exponent of Zipf's law. *Glottometrics,* 11; 98–119.

Ferrer-i-Cancho, Ramon (2013). Hubiness, length, crossing and their relationships in dependency trees. *Glottometrics,* 25; 1–21.

Ferrer-i-Cancho, Ramon; Servedio, Vito (2005). Can simple models explain Zipf's law for all exponents? *Glottometrics,* 11; 1–8.

Gnatchuk, Hanna (2015). Phonosemantic features of English and German consonants. *Glottometrics,* 30; 1–18.

Grzybek, Peter (2007). On the systematic and system-based study of grapheme frequencies. A re-analysis of German letter frequencies. *Glottometrics,* 15; 82–91.

Grzybek, Peter; Altmann, Gabriel (2002). Oscillation in the frequency-length relationship. *Glottometrics,* 5; 97–107.

Grzybek, Peter; Kelih, Emmerich (2005). Häufigkeiten von Buchstaben / Graphemen / Phonemen: Konvergenzen des Rangierungsverhaltens. *Glottometrics,* 9; 62–73.

Grzybek, Peter; Kelih, Emmerich; Stadlober, Ernst (2008). The relation between word length and sentence length: An intra-systemtic perspective in the core data structure. *Glottometrics,* 16; 111–121.

Gumenjuk, A.S.; Kostyshin, A.; Simonova S. (2002). An approach to the research of the structure of linguistic and musical texts. *Glottometrics,* 3; 61–89.

Gumenjuk, A.S. et al. (2004). On the acoustic elements of a poem and the formal procedures of their segmentation. *Glottometrics,* 8; 42–67.

Heinicke, Nora (2008). Wortlängenverteilungen in französischen Briefen eines Autors. *Glottometrics,* 16; 38–45.

Hilberg, Wolfgang (2002). The Unexpected Fundamental Influence of Mathematics upon Language. *Glottometrics,* 5; 29–50.

Hilberg, Wolfgang (2004). Some results of quantitative linguistics derived from a structural language model. *Glottometrics,* 7; 1–24.

Hřebíček, Luděk (2002). The elements of symmetry in text structures. *Glottometrics,* 2; 17–32.

Hřebíček, Luděk (2002). Zipf's law and text. *Glottometrics,* 3; 27–38.

Hřebíček, Luděk (2003). Some Aspects of the Power Law. *Glottometrics,* 6; 1–8.

Hřebíček, Luděk (2005). Contextual relationships. *Glottometrics,* 9; 54–61.

Hua, Wang (2012). Length and complexity of NPs in Written English. *Glottometrics,* 24; 79–87.

Hussien, Osama Abdelaziz (2004). The Lerchianness plot. *Glottometrics,* 7; 50–64.

Ishida, Motohiro; Ishida, Kazue (2007). On distributions of sentence lengths in Japanese writing. *Glottometrics,* 15; 28–44.

Jahn, Thomas; Uckel, Annika (2008). Verteilung von Wortlängen in englischen Spam-E-Mails. *Glottometrics,* 17; 1–7.

Jayaram, Bijapur D.; Vidya, Matummal N. (2006). Word Length Distribution in Indian Languages. *Glottometrics,* 12; 16–38.

Joyce, Terry (2005). Constructing a Large-Scale Database of Japanese Word Associations. *Glottometrics,* 10; 82–98.

Kantemir, Sergej; Levickij, Viktor V. (2005). Die statistische Analyse des semantischen Feldes der Farbbezeichnungen im Deutschen. *Glottometrics,* 11; 64–97.

Kaßel, Anja; Livesey, Eleanor (2001). Untersuchungen zur Satzlängenhäufigkeit im Englischen: Am Beispiel von Texten aus Presse und Literatur (Belletristik). *Glottometrics,* 1; 27–50.

Kazartsev, Evgeny (2006). Zum Problem der Entstehung des syllabotonischen Versmaßsystems im europäischen Vers. *Glottometrics,* 13; 1–22.

Kelih, Emmerich (2008). Modelling polysemy in different languages: A continuous approach. *Glottometrics,* 16; 46–56.

Kelih, Emmerich (2009). Graphemhäufigkeiten in slawischen Sprachen: Stetige Modelle. *Glottometrics,* 18; 52–68.

Kelih, Emmerich (2010). The type-token relationship in Slavic parallel texts. *Glottometrics,* 20; 1–11.

Kelih, Emmerich; Grzybek, Peter (2004). Häufigkeiten von Satzlängen. Zum Faktor der Intervallgröße als Einflussvariable (am Beispiel slowenischer Texte). *Glottometrics,* 8; 23–41.

Kelih, Emmerich; Popescu, Ioan-Iovitz; Altmann, Gabriel (2014). Some aspects of Slavic phonemics and graphemics. *Glottometrics,* 27; 18–53.

Kelih, Emmerich; Zörnig, Peter (2012). Models of morph lengths: Discrete and continuous approaches. *Glottometrics,* 24; 70–78.

Kess, Joseph F. (2005). On the History, Use, and Structure of Japanese Kanji. *Glottometrics,* 10; 1–15.

Kiyko, Svitlana (2006). Wortlängen im Gotischen. *Glottometrics,* 13; 59–65.

Kiyko, Svitlana (2007). Wortlängen im Weißrussischen. *Glottometrics,* 14; 46–57.

Knaus, Marina (2008). Zur Verteilung rhythmischer Einheiten in russischer Prosa. *Glottometrics,* 16; 57–62.

Knight, Róisín (2013). Laws governing rank frequency and stratification in English texts. *Glottometrics,* 25; 30–42.

Köhler, Reinhard (2002). Power Law Models in Linguistics: Hungarian. *Glottometrics,* 5; 51–61.

Köhler, Reinhard (2005). Quantitative Untersuchungen zur Valenz deutscher Verben. *Glottometrics,* 9; 13–20.

Köhler, Reinhard (2014). Towards a theory of compounding. *Glottometrics,* 28; 75–86.

Köhler, Reinhard; Naumann, Sven (2013). Syntactic Complexity and Position in Hungarian. *Glottometrics,* 26; 27–37.

Köhler, Reinhard; Rapp, Reinhard (2007). A psycholinguistic application of synergetic linguistics. *Glottometrics,* 15; 62–70.

Kornai, András (2002). How many words are there? *Glottometrics,* 4; 61–86.

Körner, Helle (2002). Der Zuwachs der Wörter auf –ion im Deutschen. *Glottometrics,* 2; 82–86.

Körner, Helle (2004). Zur Entwicklung des deutschen (Lehn-)Wortschatzes. *Glottometrics,* 7; 25–49.

Kotsyuba, Oxana (2007). Russizismen im deutschen Wortschatz. *Glottometrics,* 15; 13–23.

Krause, Marion (2002). Subjektive Bewertung von Vorkommenshäufigkeiten: Methode und Ergebnisse. *Glottometrics,* 2; 53–81.

Kromer, Viktor V. (2001). Word length model, based on the one-displaced Poisson-uniform distribution. *Glottometrics,* 1; 87–96.

Kromer, Viktor V. (2002). Zipf's law and its modification possibilities. *Glottometrics,* 5; 1–13.

Laufer, Janja; Nemcová, Emília (2009). Diversifikation deutscher morphologischer Klassen in SMS. *Glottometrics,* 18; 13–25.

Lehfeldt, Werner; Altmann, Gabriel (2002). Der altrussische Jerwandel. *Glottometrics,* 2; 33–44.

Levickij, Viktor V.; Hikow, Leonid (2004). Zum Gebrauch der Wortarten im Autorenstil. *Glottometrics,* 8; 12–22.

Levitskij, Viktor V.; Melnyk, Yulia P. (2011). Sentence length and sentence structure in English prose. *Glottometrics,* 21; 14–24.

Lewizkij, Victor; Matskulyak, Yulia (2009). Semantische Kombinierbarkeit von Komponenten in der Struktur der deutschen Komposita. *Glottometrics,* 19; 11–41.

Li, Wentian (2002). Zipf's Law everywhere. *Glottometrics,* 5; 14–21.

Liu, Haitao (2007). Probability distribution of dependency distance. *Glottometrics,* 15; 1–12.

Liu, Haitao; Zhao, Yiyi; Huang, Wei (2010). How do Local Syntactic Structures Influence Global Properties in Language Networks? *Glottometrics,* 20; 38–58.

Long, Eric; Yokoyama, Shoichi (2005). Text Genre and Kanji Frequency. *Glottometrics,* 10; 55–72.

Lvova, Nadija L. (2005). Semantic functions of English initial consonant clusters. *Glottometrics,* 9; 21–28.

Mačutek, Ján (2005). Discrete distributions connected by partial summations. *Glottometrics,* 11; 51–55.

Mačutek, Ján (2008). Runes: complexity and distinctivity. *Glottometrics,* 16; 1–16.

Mačutek, Ján; Popescu, Ioan-Iovitz; Altmann, Gabriel (2007). Confidence intervals and tests for the *h*-point and related text characteristics. *Glottometrics,* 15; 45–52.

Mačutek, Ján; Švehlíková, Zuzana; Cenkerová, Zuzana (2011). Towards a model for rank-frequency distributions of melodic intervals. *Glottometrics,* 21; 60–64.

Majerník, Vladimír (2002). A conceptualization of the configurational and functional organization. *Glottometrics,* 3; 110–135.

Martináková, Zuzana et al. (2008). Some problems of musical texts. *Glottometrics,* 16; 80–110.

Marx, Michael (2001). Zu den Wortlängen in polnischen Briefen. *Glottometrics,* 1; 51–62.

Masuda, Hisashi; Joyce, Terry (2005). A Database of Two-Kanji Compound Words Featuring Morphological Family, Morphological Structure, and Semantic Category Data. *Glottometrics,* 10; 30–44.

Melka, Tomi S.; Altmann, Gabriel (2014). Script complexity: A Case Study. *Glottometrics,* 28; 56–74.

Meuser, Katharina; Schütte, Jana Madlen; Stremme, Sina (2008). Pluralallomorphe in den Kurzgeschichten von Wolfdietrich Schnurre. *Glottometrics,* 17; 12–17.

Meyer, Peter (2002). Laws and Theories in Quantitative Linguistics. *Glottometrics,* 5; 62–80.

Miyaoka, Yayoi; Tamaoka, Katsuo (2005). A Corpus Investigation of the *Right-hand Head Rule* Applied to Japanese Affixes. *Glottometrics,* 10; 45–54.

Mohanty, Panchanan; Popescu, Ioan-Iovitz (2014). Word length in Indian languages. *Glottometrics,* 29; 95–109.

Montemurro, Marcelo A.; Zanette, Damián H. (2002). New perspectives on Zipf's law in linguistics: from single texts to large corpora. *Glottometrics,* 4; 87–99.

Naumann, Sven; Popescu, Ioan-Iovitz; Altmann, Gabriel (2012). Aspects of nominal style. *Glottometrics,* 23; 23–55.

Nemcová, Emília (2009). Nominal suffixes in German press texts. *Glottometrics,* 18; 69–76.

Overbeck, Anja et al. (2010). Analysis of Italian word classes. *Glottometrics,* 20; 12–28.

Pan, Xiaxing; Liu, Haitao (2014). Adnominal Constructions in Modern Chinese and their Distribution Properties. *Glottometrics,* 29; 1–30.

Pauli, Francesco; Tuzzi, Arjuna (2009). The End of Year Addresses of the Presidents of the Italian Republic (1948-2006): discoursal similarities and differences. *Glottometrics,* 18; 40–51.

Pavlyshenko, Olha (2013). The lexical-semantic fields of verbs in English texts. *Glottometrics,* 25; 69–84.

Peust, Carsten (2006). Script complexity revisited. *Glottometrics,* 12; 11–15.

Popescu, Ioan-Iovitz (2003). On a Zipf's Law Extension to Impact Factors. *Glottometrics,* 6; 83–93.

Popescu, Ioan-Iovitz; Altmann, Gabriel (2006). Some aspects of word frequencies. *Glottometrics,* 13; 23–46.

Popescu, Ioan-Iovitz; Altmann, Gabriel (2007). Writer's view of text generation. *Glottometrics,* 15; 71–81.

Popescu, Ioan-Iovitz; Altmann, Gabriel (2008). On the regularity of diversification in language. *Glottometrics,* 17; 94–108.

Popescu, Ioan-Iovitz; Altmann, Gabriel (2008). Zipf's mean and language typology. *Glottometrics,* 16; 31–37.

Popescu, Ioan-Iovitz; Altmann, Gabriel (2014). Clause centrality. *Glottometrics,* 28; 13–36.

Popescu, Ioan-Iovitz; Altmann, Gabriel (2014). The lambda structure of language levels. *Glottometrics,* 27; 54–88.

Popescu, Ioan-Iovitz; Altmann, Gabriel (2015). A simplified lambda indicator in text analysis. *Glottometrics,* 30; 19–44.

Popescu, Ioan-Iovitz; Best, Karl-Heinz; Altmann, Gabriel (2007). On the Dynamics of Word Classes in Text. *Glottometrics,* 14; 58–71.

Popescu, Ioan-Iovitz; Čech, Radek; Altmann, Gabriel (2011). On stratification in poetry. *Glottometrics,* 21; 54–59.

Popescu, Ioan-Iovitz; Čech, Radek; Altmann, Gabriel (2011). Vocabulary richness in Slovak poetry. *Glottometrics,* 22; 62–72.

Popescu, Ioan-Iovitz; Čech, Radek; Altmann, Gabriel (2014). Descriptivity in special texts. *Glottometrics,* 29; 70–80.

Popescu, Ioan-Iovitz et al. (2009). Diversification of the case. *Glottometrics,* 18; 32–39.

Popescu, Ioan-Iovitz; Mačutek, Ján; Altmann, Gabriel (2008). Word frequency and arc length. *Glottometrics,* 17; 18–42.

Popescu, Ioan-Iovitz; Martinákova-Rendeková, Zuzana; Altmann, Gabriel (2012). Stratification in musical texts based on rank-frequency distribution of tone pitches. *Glottometrics,* 24; 25–40.

Popescu, Ioan-Iovitz; Zörnig, Peter; Altmann, Gabriel (2013). Arc length, vocabulary richness and text size. *Glottometrics,* 25; 43–53.

Popescu, Ioan-Iovitz et al. (2013). Some statistics for sequential text properties. *Glottometrics,* 26; 50–95.

Prün, Claudia; Zipf, Robert (2002). Biographical notes on G.K. Zipf. *Glottometrics,* 3; 1–10.

Robbins, Jeff (2002). Technology, Ease, and Entropy: A Testimonial to Zipf's Principle of Least Effort. *Glottometrics,* 5; 81–96.

Roelcke, Thorsten (2002). Efficiency of communication. A new concept of language economy. *Glottometrics,* 4; 27–38.

Roelcke, Thorsten; Altmann, Gabriel (2013). Kant‘s Terminology of Cognitive Capacities. A Quantitative Study on Lexicographic Polysemy in the “Critique of Pure Reason”. *Glottometrics,* 26; 17–26.

Rottmann, Otto A. (2002). Sylable Lengths in Russian, Bulgarian, Old Church Slavonic and Slovene. *Glottometrics,* 2; 87–94.

Rottmann, Otto A. (2003). Word lengths in the Baltic languages – are they of the same type as the word lengths in the Slavic languages? *Glottometrics,* 6; 52–60.

Rousseau, Ronald (2002). George Kingsley Zipf: life, ideas, his law and informetrics. *Glottometrics,* 3; 11–18.

Rovenchak, Andrij (2011). A naïve conception of the uncertainty principle in the multiparametric attibution of texts. *Glottometrics,* 21; 65–72.

Rovenchak, Andrij; Mačutek, Ján; Riley, Charles (2009). Distribution of complexities in the Vai script. *Glottometrics,* 18; 1–12.

Sanada, Haruko (2001). New *Kango* of the early Meiji era: Their survival and disappearance from Meiji to the present. *Glottometrics,* 1; 63–86.

Sanada, Haruko; Altmann, Gabriel (2009). Diversification of postpositions in Japanese. *Glottometrics,* 19; 70–79.

Sandalescu, C. George et al. (2015). Quantifying Joyce’s *Finnegans Wake*. *Glottometrics,* 30; 45–72.

Schroeder, Manfred (2002). Power laws: from *Alvarez* to *Zipf*. *Glottometrics,* 4; 39–44.

Serdelová, Kvetoslava (2005). Some properties of slang words. *Glottometrics,* 9; 40–45.

Smith, Reginald (2007). Investigation of the Zipf-plot of the extinct Meroitic language. *Glottometrics,* 15; 53–61.

Smith, Reginald (2012). Distinct word length frequency and symbol entropies. *Glottometrics,* 23; 7–22.

Strauß, Udo; Altmann, Gabriel (2003). Age and polysemy of words. *Glottometrics,* 6; 61–64.

Tamaoka, Katsuo; Altmann, Gabriel (2004). Symmetry of Japanese Kanji Lexical Productivity on the Left- and Right-hand Sides. *Glottometrics,* 7; 65–84.

Tamaoka, Katsuo; Altmann, Gabriel (2005). Mathematical Modelling for Japanese Kanji Strokes in Relation to Frequency, Asymmetry and Readings. *Glottometrics,* 10; 16–29.

Tamaoka, Katsuo; Makioka, Shogo; Murata, Tadao (2004). Are the effects of vowel repetition influenced by frequencies? A corpus study on CVCVCV-structured nouns with and without vowel repetition. *Glottometrics,* 8; 1–11.

Tamaoka, Katsuo et al. (2005). Predicting Attachment of the Light Verb *-suru* to Japanese Two-kanji Compound Words Using Four Aspects. *Glottometrics,* 10; 73–81.

Tatar, Doina; Lupea, Mihaiela; Altmann, Gabriel (2014). Hreb-like analysis of Eminescu’s poems. *Glottometrics,* 28; 37–55.

Ternes, Katharina (2011). Entwicklungen im deutschen Wortschatz. *Glottometrics,* 21; 25–53.

Tuzzi, Arjuna; Popescu, Ioan-Iovitz; Altmann, Gabriel (2009). Parts-of-speech diversification in Italian texts. *Glottometrics,* 19; 42–48.

Tuzzi, Arjuna et al. (2012). Aspects of the behaviour of parts-of-speech in Italian texts. *Glottometrics,* 24; 41–69.

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

Uhlířová, Ludmila (2002). Zipf's notion of 'economy' on the text level. *Glottometrics,* 3; 39–60.

Uhlířová, Ludmila (2007). Word frequency and position in sentence. *Glottometrics,* 14; 1–20.

Vulanović, Relja (2008). A mathematical analysis of parts-of-speech systems. *Glottometrics,* 17; 51–65.

Wang, Lu (2011). Polysemy and word length in Chinese. *Glottometrics,* 22; 73–85.

Wheeler, Eric S. (2002). Zipf's Law and why it works everywhere. *Glottometrics,* 4; 45–48.

Wheeler, Eric S. (2003). Multidimensional Scaling to Visualize Text Separation. *Glottometrics,* 6; 65–69.

Wilson, Andrew (2003). Word-Length Distribution in Modern Welsh Prose Texts. *Glottometrics,* 6; 35–39.

Wimmer, Gejza; Altmann, Gabriel (2001). Some statistical investigations concerning word classes. *Glottometrics,* 1; 109–123.

Xiaodong, Zhao (2012). Vocabulary growth of content words of ESP and General English. A contrastive study based on CMTE and BNC. *Glottometrics,* 24; 12–24.

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

Yokoyama, Shoichi; Wada, Yukiko (2006). A logistic regression model of variant preference in Japanese kanji: an integration of mere exposure effect and the generalized matching law. *Glottometrics,* 12; 63–74.

Yujia, Zhu (2013). Sentence length and syntactic complexity in spoken and written English. *Glottometrics,* 26; 9–16.

Ziegler, Arne; Best, Karl-Heinz; Altmann, Gabriel (2001). A contribution to text spectra. *Glottometrics,* 1; 97–108.

Zörnig, Peter (2013). A continuous model for the distances between coextensive words in a text. *Glottometrics,* 25; 54–68.

Zörnig, Peter; Melka, Tomi S. (2014). Ethnographic Study of the *pintaderas* of Gran Canaria: A Measure Theoretic Approach for Quantifying Symmetry. *Glottometrics,* 29; 31–58.

2. History

Best, Karl-Heinz (2005). Georg Philipp Harsdörffer (1607-1658). *Glottometrics,* 9; 86–88.

Best, Karl-Heinz (2005). Georg von der Gabelentz (1840-1893). *Glottometrics,* 9; 77–79.

Best, Karl-Heinz (2005). Gottfried Wilhelm Leibniz (1646-1716). *Glottometrics*, 9; 79–82.

Best, Karl-Heinz (2005). Karl Marbe (1869-1953). *Glottometrics*, 9; 74–76.

Best, Karl-Heinz (2006). Adolf Lucas Bacmeister (1827-1873). *Glottometrics*, 13; 79–84.

Best, Karl-Heinz (2006). August Friedrich Pott (1802-1887). *Glottometrics*, 12; 94–96.

Best, Karl-Heinz (2006). August Schleicher (1821-1868). *Glottometrics*, 13; 73–75.

Best, Karl-Heinz (2006). Ernst Wilhelm Förstemann (1822-1906). *Glottometrics*, 12; 77–86.

Best, Karl-Heinz (2006). Hans Arens (1911-2003). *Glottometrics*, 13; 75–79.

Best, Karl-Heinz (2006). Jean Paul (1763-1825). *Glottometrics*, 12; 75–77.

Best, Karl-Heinz (2006). Karl Knauer (1906-1966). *Glottometrics*, 12; 86–94.

Best, Karl-Heinz (2006). Siegfried Behn (1884-1970). *Glottometrics*, 13; 85–88.

Best, Karl-Heinz (2007). Erwin Kunath (1899-1983). *Glottometrics*, 14; 78–80.

Best, Karl-Heinz (2007). Lorenzo Bianchi (1889-1960). *Glottometrics*, 14; 72–74.

Best, Karl-Heinz (2007). Manfred Faust (1936-1997). *Glottometrics*, 14; 74–78.

Best, Karl-Heinz (2007). Otto Behaghel (1854-1936). *Glottometrics*, 14; 80–86.

Best, Karl-Heinz (2007). Paul Menzerath (1883-1954). *Glottometrics*, 14; 86–98.

Best, Karl-Heinz (2008). Adolf Busemann (1887-1967). *Glottometrics*, 16; 124–127.

Best, Karl-Heinz (2008). Helmut Meier (1897-1973). *Glottometrics*, 16; 122–124.

Best, Karl-Heinz (2008). Kaj Brynolf Lindgren (1922-2007). *Glottometrics*, 16; 127–131.

Best, Karl-Heinz (2008). Moritz Wilhelm Drobisch (1802-1896). *Glottometrics*, 17; 109–114.

Best, Karl-Heinz (2009). Eduard Sievers (1850-1932). *Glottometrics*, 18; 87–91.

Best, Karl-Heinz (2009). Ferdinand Schrey (1850-1938). *Glottometrics*, 18; 91–94.

Best, Karl-Heinz (2009). Friedrich Wilhelm Kaeding (1843-1928). *Glottometrics*, 18; 81–87.

Best, Karl-Heinz (2009). Heinrich August Kerndörffer (1769-1846). *Glottometrics*, 18; 94–96.

Best, Karl-Heinz (2009). William Palin Elderton (1877-1962). *Glottometrics*, 19; 99–101.

Best, Karl-Heinz (2010). Laut- und Buchstabenzählungen im frühen 19. Jahrhundert. François Dujardin aîné (1834). Josef Nowak (1848). *Glottometrics*, 20; 110–114.

Best, Karl-Heinz; Altmann, Gabriel (2007). Gustav Herdan (1897-1968). *Glottometrics*, 15; 92–96.

Best, Karl-Heinz; Kotrasch, Brita (2005). Albert Thumb (1865-1915). *Glottometrics*, 9; 82–84.

Grzybek, Peter (2003). Viktor Jakovlevič Bunjakovskij (1804-1889). A forerunner of quantitative linguistics? *Glottometrics*, 6; 103–104.

Grzybek, Peter (2004). Nikolaj Gavrilovič Černyševskij. A Forerunner of Quantitative Stylistics in Russia. *Glottometrics*, 7; 91–93.

Grzybek, Peter (2006). Jan Svatopluk Presl (1791-1849). *Glottometrics*, 13; 88–91.

Grzybek, Peter (2006). Tomo Maretić's First Croatian and/or Serbian Sound Statistics (1899). *Glottometrics,* 13; 92–96.

Grzybek, Peter (2012). Harry Dexter Kitson (1886-1959). *Glottometrics,* 24; 88–94.

Grzybek, Peter (2012). Michail Lopatto: *Attempt at an Introduction into the Theory of Prose* (1918). *Glottometrics,* 23; 70–80.

Grzybek, Peter (2013). Historical Remarks on the Consonant-Vowel Proportion – From Cryptoanalysis to Linguistic Typology. The Concept of Phonological Stoichiometry (Francis Lieber, 1800-1872). *Glottometrics,* 26; 96–103.

Grzybek, Peter; Kelih, Emmerich (2004). Anton Semënovič Budilovič (1846-1908). A Forerunner of Quantitative Linguistics in Russia? *Glottometrics,* 7; 94–96.

Hřebíček, Luděk (2005). Jiří Krámský (1913-1991). *Glottometrics,* 9; 76–77.

Kelih, Emmerich (2004). Dmitrij N. Kudrjavskij (1867-1920) – ein Wegbereiter quantitativer Methoden in der russischen Sprachwissenschaft. *Glottometrics,* 8; 79–83.

Kelih, Emmerich (2007). B.I. Jarcho as a pioneer of the exact study of literature. *Glottometrics,* 15; 96–100.

Kelih, Emmerich (2009). Quantitative Hypothesen bei Mikołaj Kruszewski. *Glottometrics,* 18; 77–81.

Maruyama, Naoko (2005). Sizuo Mizutani (1926). The Founder of Japanese Quantitative Linguistics. *Glottometrics,* 10; 99–107.

Pawłowski, Adam (2004). Jerzy Woronczak (1923-2003). The Founder of Polish Quantitative Linguistics. *Glottometrics,* 8; 90–98.

Pawłowski, Adam (2004). Wincenty Lutosławski (1863-1954): A Forgotten Father of Stylometry. *Glottometrics,* 8; 83–90.

Pawłowski, Adam (2005). Jan Czekanowski (1882-1965) – a pioneer of multidimensional taxonomy. *Glottometrics,* 9; 84–86.

Uhlířová, Ludmila (2003). Bohumil Trnka: The first bibliography. *Glottometrics,* 6; 105–106.

3. Bibliographies

Best, Karl-Heinz (2014). Diversification. *Glottometrics,* 28; 87–91.

Gnatchuk, Hanna (2015). A bibliography of quantitative studies on sound symbolism. *Glottometrics,* 30; 86–90.

Köhler, Reinhard (2014). Motifs. *Glottometrics,* 27; 89–90.

4. Reviews

Andersen, Simone (2002). [Rev.] Best, K.-H. (ed.): *Häufigkeitsverteilungen in Texten*. Göttingen: Peust & Gutschmidt Verlag, 2001. [= Göttinger Linguistische Abhandlungen; 4]. *Glottometrics,* 2; 95–100.

Chen, Ruina (2014). [Rev.] Ji, Meng: *Exploratory Statistical Techniques for the Study of Literary Translation*. Lüdenscheid: RAM-Verlag, 2013. *Glottometrics,* 28; 92–95.

Hřebíček, Luděk; Altmann, Gabriel (2006). [Rev.] Skalička, Vladimír*: Souborné dílo. III. díl* [Collected works. Part III]. Praha: Nakladatelství Karolinum, 2006. *Glottometrics,* 13; 100–102.

Kelih, Emmerich (2006). [Rev.] Polikarpov, A.A.; Sil'nickij, G.G.; Poddubnyj, V.V. (eds.): *Kvantitativnaja Lingvistika: Issledovanija i modeli (Klim-2005). Materialy vserossijskoj naučnoj konferencii (6-10 ijunja 2005 g.).* Novosibirsk: Novosibirskij Gosudarstvennyj Pedagogičeskij Universitet, 2005. *Glottometrics,* 12; 97–106.

Kijko, Juri (2006). [Rev.] Gabriel Altmann; Viktor Levickij; Valentina Perebyinis (eds.): *Problemy kvantytatyvnoji lingvistyky: zbirnyk naukovych pracj // Probleme der Quantitativen Linguistik: Sammelband*. Černivci: Ruta, 2005. *Glottometrics,* 12; 106–108.

Kusendová, Jana (2004). [Rev.] Wimmer, Gejza; Altmann, Gabriel, Hřebíček, Luděk, Ondrejovič, Slavomír, Wimmerová, Soňa: *Úvod do analýzy textov* [Introduction to Text Analysis]. Bratislava: Veda, 2003. *Glottometrics,* 7; 97–99.

Kusendová, Jana (2005). [Rev.] Don McNico: *A Primer of Signal Detection Theory*. London: Lawrence Erlbaum Associates Publishers, 2005. *Glottometrics,* 9; 89–90.

Liu, Haitao (2012). [Rev.] Grzybek, Peter; Kelih, Emmerich; Mačutek, Ján (eds.): *Text and Language.* Wien: Praesens Verlag, 2010. *Glottometrics,* 23; 81–83.

Mačutek, Ján (2006). [Rev.] Grzybek, Peter (ed.), *Contributions to the Science of Text and Language. Word Length Studies and Related Issues.* (Text, Speech and Language Technology, vol. 31). Dordrecht, Springer 2006, 352 pp. *Glottometrics*, 13; 97–100.

Mačutek, Ján (2009). [Rev.] Jeehyeon Eom: *Rhythmus im Akzent. Zur Modellierung der Akzentverteilung als einer Grundlage des Sprachrhythmus im Russischen*. München: Verlag Otto Sagner, 2006. *Glottometrics,* 18; 97–99.

Nemcová, Emília (2012). [Rev.] Emmerich Kelih: *Die Silbe in slawischen Sprachen*. Von der Optimalitästheorie zu einer funktionalen Interpretation. München - Berlin - Washington, D.C.: Sagner, 2012. *Glottometrics,* 24; 97–99.

Nemcová, Emília (2012). [Rev.] Gordana Đuraš: *Generalized Poisson models for word length frequencies in texts of Slavic languages*. Diss. University of Graz. *Glottometrics,* 24; 95–96.

Uhlířová, Ludmila (2009). [Rev.] Haruko Sanada: *Investigations in Japanese Historical Lexicology (Revised Edition).* Göttingen: Peust & Gutschmidt Verlag, 2008. [= Göttinger Linguistische Abhandlungen; 6]. *Glottometrics,* 18; 99–100.

Yan, Jingqi (2013). [Rev.] James W. Pennebaker: *The Secret Life of Pronouns - What our Words Say about Us*. New York: Bloomsbury Press, 2011. *Glottometrics,* 25; 97–101.

Wachtel, Michael (2013). [Rev.] Barry P. Scherr; James Bailey; Evgeny V. Kazartsev (eds.): *Formal Methods in Poetics: A Collection of Scholarly Works Dedicated to the Memory of Professor M.A. Krasnoperova*. Lüdenscheid: RAM-Verlag, 2011. *Glottometrics,* 26; 104–109.

5. Miscellanea (Announcements, Project Reports, Software Presentations, etc.)

Benešová, Martina; Faltýnek, Dan (2013). Report: Research activities at the Department of General Linguistics of the Philosophical Faculty of Palacký University in Olomouc, Czech Republic. *Glottometrics,* 25; 94–96.

Cong, Jin (2011). Quantitative Linguistics funded by NSSFC. *Glottometrics,* 22; 85.

Jüngling, Ralf; Altmann, Gabriel (2003). Python for linguistics? *Glottometrics,* 6; 70–82.

Kelih, Emmerich; Grzybek, Peter; Stadlober, Ernst (2003). Das Grazer Projekt zu Wortlängen(häufigkeiten). *Glottometrics,* 6; 94–102.

Kubát, Miroslav; Maltach, Vladimír; Čech, Radek (2014). Quantitative Index Text Analyser (QUITA). *Glottometrics,* 27; 91–92.

Other linguistic publications of RAM-Verlag:

Studies in Quantitative Linguistics

Up to now, the following volumes appeared:

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, VI + 148 pp.
15. K.-H. Best, E. Kelih (Hrsg.), *Entlehnungen und Fremdwörter: Quantitative Aspekte.* 2014, IV + 163 pp.
16. I.-I. Popescu, K.-H. Best, G. Altmann, *Unified modeling of length in language.* 2014. III + 123 pp.
17. G. Altmann, R. Čech, J. Mačutek, L. Uhlířová (eds.), *Empirical approaches to text and language analysis.* 2014, IV + 230 pp.
18. M. Kubát, V. Matlach, R. Čech, *QUITA. Quantitative Index Text Analyzer*. 2014, IV + 106 pp.
19. K.-H. Best (Hrsg.), *Studies zur Geschichte der Quantitativen Linguistik. Band 1*. 2015, III + 159 pp.