Quantitative Methodology in Linguistics

Elizaveta Dmitrieva
Moscow State University of Railway Engineering, Moscow, Russia

The paper is devoted to the study of quantitative methods in linguistics and describes the studies conducted. The purpose is to give the general idea of these studies. The first one considers one of the principal logical categories—the quality. The basis of the research was comprised of lexicographical recourses. The text study finishes up the research. The second one dwells on the usage of the typological indices method in the research of comparatives and superlatives in English, German, and Russian texts. The principal method used is that of typological indices. As the result, people can observe the prospects of this method in linguistics.

Keywords: quantitative linguistics, method, quality, comparison

Introduction

The paper dwells upon the quantitative methodology in linguistics and describes two researches completed with the use of the method.

According to Stefan T. Gries (2003), over the past years, linguistics has taken a decidedly quantitative turn. While sub-disciplines such as phonetics, sociolinguistics, and experimental psycholinguistics have employed statistical methods for a long time, work in most other central sub-disciplines—morphology, syntax, semantics, to name but a few—has only done so since about the 1990s. This shift has, according to the author, no doubt been facilitated by a variety of developments.

The predominant quantitative method used is the statistical one. Stefan T. Gries describes the following methods: (1) distributions, (2) frequencies, (3) averages (such as means or medians), (4) dispersions (such as standard deviations or interquartile ranges), and (5) correlations (such as Pearson’s $r$ or Kendall’s $\tau$).

In general, the role of qualitative linguistics is now to unveil corresponding phenomena, to systematically describe them, and to find and formulate laws which explain the observed and described facts. Quantitative interrelations have an enormous value for fundamental research but they can also be used and applied in many fields such as computational linguistics and natural language processing, language teaching, optimisation of texts, etc..

Early modern linguistics, in the time after the seminal contribution of Saussure (1880), was mainly interested in the structure of language. Consequently, linguists adopted the qualitative means of mathematics: logics, algebra, and set theory. The historical development of linguistics and a subsequent one-sided emphasis on certain elements in the structuralist achievements resulted in the emergence of an absolutely static concept of system, which has prevailed until present days. The aspects of systems which exceed structure, viz. functions,

Elizaveta Dmitrieva, associate professor, Ph.D., Moscow State University of Railway Engineering.
dynamics, processes, were disregarded almost completely. To overcome this flaw, the quantitative parts of mathematics (e.g., analysis, probability theory and statistics, function theory, differential and difference equations) must be added to the qualitative ones, and this is the actual aim of qualitative linguistics.

Generally speaking, the development of quantitative methods aims at improving the exactness and precision of the possible statements on the properties of linguistic and textual objects. Exactness depends, in fact, on two factors: (1) the acuity of the definition of a concept and (2) the quality of the measurement methods with which the given property can be determined. Success in defining a linguistic property with sufficiently crisp concepts enables the author to operate it with mathematical means, provided the operations correspond to the scale level (cf. above) of the concepts.

Such operations help to derive new insights which would not be possible without them: Appraisal criteria which exist at the time being only in a subjective, tentative form can be made objective and operationalised (e.g., in stylistics), interrelations between units and properties can be detected, which remain invisible to qualitative methods, and workable methods for technical and other fields of application can be found where traditional linguistic methods fail or produce inappropriate results due to the stochastic properties of the data or to the sheer mass of them.

Now the study turns to the description of the two researches undertaken.

**Research 1**

The first one deals with the usage of suffixes in the Russian and German languages as a means of quality expression. The topic is urgent as the correct expression of quality is an important part of cross-culture communication. That helps to understand linguistic mentality of a nation. At the same time, the correct understanding of the derivational process can be helpful in language learning. The aim of the study is to consider the use of suffixes in the Russian and German languages as a means of quality expression. The novelty of the research conducted consists of the research method applied. That is the method of quantitative analysis. The results achieved can be used in the courses of theoretical and comparative grammar.

The paper is aimed at students as well as teaching staff of colleges and universities.

Derivation, known to be one of the principal means of quality expression, classifies the quality-expressing words into the noun group and the adjectival group, the latter being the centre of the semantic field of quality.

**Figure 1. Qualitative suffixes.**

The research undertaken consisted of two stages. The first one was the lexicographical data research, the aim of which was to study the number of qualitative suffixes in Russian and German. The conducted research of the Russian lexicographical data has discovered 47 quality suffixes (17 being adjectival and 30 nominal). The
bulk of the nominal suffixes are, however, stylistically marked and not frequently used. In the German language, there have been discovered 88 suffixes of quality (39 adjectival and 49 nominal). The results are presented in Figure 1.

The abovementioned facts can make the impression of the predominance of the nominal world-building method. Yet that is not totally true as a large number (53% in Russian as well as 25, 1% of German noun-building models are non-productive). Meanwhile, the adjective-building patterns are highly productive in both the languages.

The second stage was the text data research. The aim of this stage was to find out the frequency of the use of qualitative suffixes (compared to other suffixes).

The research was based on the analyses of fiction texts in Russian and German (over 35,000 words in each language).

To describe qualitative suffixes, the author introduces an index of qualitative suffixation. An index can be described as an indirect shortcut derived from and pointing into, a greater volume of values, data, information, or knowledge.

The formula of the index of qualitative suffixation is $s_q = \frac{n}{w}$, with $n$ being the number of words with qualitative suffixes and $w$ being the number of words in the text.

Now the study compares the qualitative suffixation index with suffixation index (J. Greenberg formula). The results are presented in Table 1.

<table>
<thead>
<tr>
<th></th>
<th>Suffixation index</th>
<th>Qualitative suffixation index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Russian</td>
<td>1.13</td>
<td>0.11</td>
</tr>
<tr>
<td>German</td>
<td>0.5</td>
<td>0.06</td>
</tr>
</tbody>
</table>

From the data above one can conclude that qualitative suffixes are frequently used due to their semantic importance.

**Research 2**

The second research deals with the usage of the typological indices method in the research of comparatives and superlatives in English, German, and Russian texts. The research is based on the sample of up to 5,355 examples. The urgency and practical importance are connected with the sphere of translation studies and teaching grammar.

At the morphological level, the main constituent of the field of comparative adjectives is the system of degrees of comparison. The author will use the definition of I. A. Melchuk (1998), according to which category of the degrees of comparison is a category, the grammatical units of which characterize the degree of intensity of a quality—either in relation to the same quality to another object or all objects in a given situation imaginable, or in relation to the same quality of the same object, but at different time periods. System degrees of comparison is a typologically important characteristics of the language system, and therefore can be analyzed in terms of typological classification of languages. Further work will consider the elements of synthetism, analytism, and
suppletism in the structure of the degrees of comparison of the language. The study is based on the novels by Dan Brown “Angels and Demons” and “The Lost Symbol” and their translations into Russian and German, made by Desyatova M., E. Romanova, G. Kosov and Axel Merz. The total sample comprises 5,355 examples.

For an objective and systematic description of the phenomena of analytism, synthetism, and suppletism in the structure and the comparative study of languages, the following indices were introduced into the research apparatus:

1. Index of synthetism—\(s/n\), where \(s\) stands for the number of inflectional comparative and superlative forms, \(n\) is the total number of cases analyzed;

2. Index of analytism—\(a/n\), where \(a\) denotes the number of periphrastic comparatives and superlatives, \(n\) is the total number of cases analyzed;

3. Index of suppletivism—\(sup/n\), where \(sup\) indicates the number of suppletive forms, \(n\) is the total number of cases analyzed.

All calculations were performed on the sets of examples, each containing 100 cases of the use of comparatives and superlatives.

Analyses of the English material showed that the average synthetism index value is 0.41, with the minimum of 0.31 and the maximum value of 0.56. The average analytism index is 0.31, with the minimum of 0.20 and the maximum value of 0.38. The average suppletive index value is 0.28, with the minimum of 0.17 and the maximum value of 0.39.

In Russian, the average value of synthetism index is 0.46, with the minimum of 0.33 and the maximum value of 0.56. The average value of the analytism index is 0.33, with the minimum of 0.24 and the maximum value of 0.45. The average value of the suppletive index is 0.21, with the minimum of 0.15 and the maximum value of 0.28.

Due to the means of the comparative expression (absence of an analytical form), in the German language the indices of synthetism and suppletism can be defined: The average value of the index of synthetism is 0.71, with the minimum of 0.60 and the maximum value of 0.79. The average index of suppletism is 0.29, with the minimum of 0.21 and the maximum value of 0.40.

The above data suggest that the index of synthetism in Russian exceeds index of synthetism in the English language that conforms to the Russian as the language of inflectionally-synthetic grammatical structure. In this case a relatively high index of analytism can be explained by a semantics of periphrastic forms (as it has both the meaning of greater and smaller extend). In English, the relatively high index of synthetism, surpassing that of analytism can be explained by typological characteristic of the English language that is observed in one-morpheme words, and therefore does not contradict the observation by J. Leach and J. Kulpepper (1997) about a tendency to prefer analytical forms of the degrees of comparison. Index of suppletivism figures in Russian and in English are relatively low. Index of synthetism in German is much higher than the indices of synthetism in the Russian and English languages, and the index of suppletism in German is higher than the index of suppletism in English at 0.01, in Russian, at 0.07. In the author’s opinion, this phenomenon can be explained by the absence of German analytical comparative forms.

As it is evident from this study, there is number of similarities of means of expressing comparison that can be explained by the universal logical meaning of comparison.
Conclusion

As one can see from the example researches described above, the quantitative method in linguistic can be widely applied especially in the sphere of grammar studies. The paper illustrates its usage for the study of qualitative suffixes in comparison with other suffixes and the distribution of synthetic, analytic, and suppletive comparatives.

References


