

VADIM B. KASEVICH, APOLLINARIA S. AVRUTINA

**Indexical typology in linguistic diachrony
(with special reference to Turkish)**

Abstract

This paper focuses on the typological characteristics of Turkish seen in a historical perspective. The characteristics are presented as sets of quantitative indices calculated on the basis of two texts, one in Modern Turkish and the other one in Ottoman (i.e., Turkish before 1928). Such an approach makes it possible to replace the traditional qualitative descriptions with quantitative, and hence measurable, characteristics. Likewise, this method, applied to diachrony, provides a solid foundation for linguistic historical studies. The paper is a preliminary report that opens a series of publications within the framework of a project devoted to the development of Turkish in historical perspective.

Keywords: Turkish; Ottoman; typological indices; agglutinative languages; diachrony.

**Introduction: Historical Background.
Some Theoretical Aspects of Index-Based Typology.
The Statement of the Problem**

In his Inauguration Lecture *On the Mixed Nature of All the Languages*¹ presented in St. Petersburg University on September 21, 1900, Jan Baudouin de Courtenay insisted that, in the course of its centuries-long evolution, any language is affected by other languages, borrows some of their features and, as a result, acquires a “mixed” type.

Baudouin’s views were resonated by Edward Sapir² who expressed rather similar views with respect to linguistic typology. Edward Sapir emphasizes the well-known fact

¹ И.А. Бодуэн де Куртенэ, *О смешанном характере всех языков*, in: И.А. Бодуэн де Куртенэ *Избранные труды по общему языкознанию*, АН СССР, Москва 1963, т. 1., p. 362–372.

² Э. Сепир, *Избранные труды по языкознанию и культурологии*, Прогресс-Универс, Москва 1993.

that, strictly speaking, none of existing languages can be considered purely inflective, or purely agglutinative, or purely isolating: some elements of inflexion, agglutination, and isolation can be found in any language, and even within the framework of a specific paradigm typologically different word forms may co-exist. One could recall the case of Urdu, where the formation of noun plurals is generally of the agglutinative type, while in Arabic borrowings, the so-called broken plural of the donor language is maintained.³ In Russian, in the imperfective tense paradigm, present and past tense verbs are of synthetic type (*читаю, читал*), while those in Future – of analytical (*буду читать*). However, the *percentage* of typologically dissimilar elements within the same languages usually differs: in one language, the inflexion is dominant and agglutination is less represented, in another, the reverse is true.

As soon as one speaks of *percentage*, one naturally supposes that the differences between the languages should be expressed *numerically*. Precisely this idea was expressed by Josef Greenberg in his well-known work⁴: he introduced estimation of the language type using quantitative typological indices. Each such index is a fraction whose numerator is the number of, for example, prefixes in a representative text, and the denominator is the number of words in it. Thus, for a given language, the prefix(ality) index is PREF/W, where PREF is the number of prefixes, and W – the number of words. The corresponding parameters were calculated for a number of languages.⁵

It is extremely important that the comparability of the indices and, respectively, languages must be provided by sufficiently rigorous universal definitions of the units in terms of which the typological indices are expressed. Otherwise, any typology, especially quantitative, loses its explanatory adequacy. The above technique was developed in the book *Квантитативная типология языков Азии и Африки*.⁶ The method suggested by J. Greenberg is known as quantitative typology; it can also be labeled *the method of indexical typology*.

Typological indices provisionally described above may differ in their logical nature. For example, when evaluating the degree of agglutination in a given text (and, respectively, in the language), Greenberg relies upon the number of agglutinative morphemic (morphous) *junctures*, rather than upon the number of agglutinative affixes. In this case, the agglutination index takes up the form of A/J, i.e. the ratio of the number of agglutinative junctures (the latter is defined trivially as n-1, where n is the number of words) to the total number of

³ According to J.Greenberg, who defined agglutination with the help of the formula “one marker per one grammatical meaning”, the very fact of the presence of two markers for plurality provides evidence for non-agglutinativity of the category of number as such. However, such an approach eliminates the notion of exception, which is hardly defensible (see Joseph H. Greenberg, *A quantitative approach to the morphological typology of language*, “International Journal of American Linguistics” 1960, vol. 26, no. 3, pp. 178–194; В.Б. Касевич and С.Я. Яхонтов (ed.), *Квантитативная типология языков Азии и Африки*, ЛГУ, Ленинград 1982; В.Б. Касевич, *Семантика. Синтаксис. Морфология*, Наука, Москва 1988).

⁴ J. Greenberg. *A quantitative approach to the morphological typology of language*.

⁵ Clearly, the numerical values were affected not only by typological parameters of the languages, but also by parameters of the texts selected for the analysis. This is a separate problem, which is not discussed here.

⁶ Касевич and Яхонтов, *Квантитативная типология языков Азии и Африки*.

junctures. In contrast to that, the abovementioned index of prefixality is of another nature. The sum of the agglutinative and non-agglutinative junctures should naturally result in the total number of junctures in the text; as distinct from that, prefixes are not opposed to “non-prefixes”. In other words, while a morphemic juncture appears automatically in any combination of morphemes,⁷ this is definitely not the case with prefixes, as the prefix is not a “compulsory” component of a wordform.

If the set of indices is optimal and the text is representative, the quantitative typological study results in the establishment of a hierarchical set of parameters, or *indices*,⁸ which yields a complex multisided typological portrait of the given language; the hierarchy alluded to above reflects the “weight” of each of the parameters in the total structure of the language studied.

Naturally, the values of the indices may vary not only at the borders between more or less homogeneous classes, but also inside such classes. However, as it was shown already⁹, sharp changes in their values, which correspond to inter-class boundaries, are relatively easy to detect.

In other words, the classes distinguished on the quantitative basis are not only of quantitative but also of *categorical* nature. This situation closely resembles the categorical perception of the phonemes. Way back in the 1960-ies, it was experimentally found that if the subjects were stimulated acoustically with the stimuli monotonously varying according to some parameter (for example, the second formant), then within a certain interval the stimuli were perceived as identical. However, normally one could find a point where the stimulus was equiprobably referred to paradigmatically “adjacent” stimuli, and precisely this point corresponded to the phonological shift from one phoneme to another.¹⁰

As it follows from the title of this paper, we are primarily concerned with the languages traditionally referred to as *agglutinative*. In contrast to the dominating tradition, we divide the set of affixes into *four* subsets or sub-classes (agglutinative/non-agglutinative/inflective/non-inflective), rather than into two (agglutinative/inflective).

An agglutinative affix satisfies the condition “one grammatical meaning – one affix”. Otherwise, the affix is treated as non-agglutinative. In other words, in order to answer whether the given affix is agglutinative, one should find out whether in the same language there exist other affixes that convey the same meaning, or essentially are synonymous to the original one. If the answer to this question is affirmative the affix under consideration is agglutinative, otherwise it is non-agglutinative. For example, in English, the past tense form with *-ed* should be treated as agglutinative. Such forms as *took, ate, slept, etc.*,

⁷ Here, we simplify the problem to some extent: the situation is more complicated in the languages in which discontinuous affixes and/or discontinuous roots (as in Semitic languages) exist.

⁸ Strictly speaking, indices should be treated as independent; however, determination of their independence presents a separate formal and substantial problem, which we do not consider here (see Касевич and Яхонтов, *Квантитативная типология языков Азии и Африки*).

⁹ Касевич and Яхонтов, *Квантитативная типология языков Азии и Африки*.

¹⁰ Even animals were found to display some elements of categorical perception (see, for example, W.T. Fitch, *The evolution of language*, Cambridge University Press, Cambridge 2011).

are classified as exceptions or “irregular” forms; this is testified by the data where the subjects choose as default versions precisely those with *-ed* when asked to form the past tense from innovative verbs¹¹. In contrast, Russian plural endings *-Ø* (*нож*), *-a* (*рука*) etc. should be treated as non-agglutinative: as it is seen from the simple examples above, plural nouns in Russian are formed with synonymous endings.

An inflective affix introduces a corresponding word form into no fewer than *two* oppositions. For example, *-a* in the wordform *рука* introduces this word form into two oppositions: according to the case (*рука – руку* etc.) and according to the number (*рука – руки*). Respectively, a non-inflective affix is “responsible” for one opposition. For example, prefix *вс-* in the word form *вспахать* provides only one opposition, the aspectual one.

As a result, we obtain four attributes for the classification of affixes: agglutination, non-agglutination, inflexion, non-inflexion. Since classes of affixes are distinguished according to different criteria, the fact that some specific affix belongs to one class is not exclusive of its simultaneous belonging to another. Eventually, if we ignore for the moment the order of application of the affixes, we obtain the following system of affixes:

- 1) agglutinative, non-inflective;
- 2) agglutinative, inflective;
- 3) non-inflective, non-agglutinative;
- 4) inflective, non-agglutinative.

Each combination is related to a specific typological index: A/AF, non-A/AF etc., where A is the number of agglutinative affixes, AF – the total number of affixes etc. Clearly, according to this approach, an agglutinative language is that with a sufficiently high A/AF and sufficiently low F/AF indices.

The quantitative (indexical) approach makes it possible to expand this technique to *diachronic studies*. The connection between historical and typological linguistics was realized only recently¹²; meanwhile, this connection is actually obvious. In particular, it manifests (or *may* manifest) itself in the fact that, varying in the course of its evolution, the language in some cases maintains its typological portrait, while in other cases the latter undergoes more or less serious changes sometimes leading to switching from one typological class to another. Taking into account the abovementioned instability of the boundaries between different typological classes we can state that it is the introduction of some kind of a typological “metrics” that will provide the objective basis for historical and typological studies¹³.

¹¹ Experiments performed by Derwing and Skousen have shown that given the choice between regular (“correct”) and irregular (“incorrect”) forms, subjects always choose the former (B. Derwing and R. Skousen, *Morphology in the mental lexicon: A new look at analogy*, in: G. Booij, J. van Marle (eds), *Yearbook of Morphology*, vol. 2, Kluwer, Dordrecht 1989, pp. 56–71).

¹² А.К. Оглоблин, *Очерк диахронической типологии малайско-яванских языков*, УРСС, Москва 2009.

¹³ Here the psychological component is obviously present: when observing a language in its historical dynamics, the scholar cannot help noticing material variations, but sometimes remains “blind” to structural shifts which alter the type of the language.

The Set of Indices and the Set of Texts. The Formation of a Diachronically Oriented Set of Typological Indices

The selection of typological indices whose historical dynamics would most naturally reflect typological trends of the language system is far from trivial a problem. In the Introduction, we already noted that no tradition can help us in this project followed. That is why we essentially followed our own results¹⁴ and as well as some general ideas and natural consequences from theoretical notions widely spread in the classical linguistics.

Ultimately, we have formed a set of indices presumably reflecting the historical and typological dynamics. Below, we present their list with brief comments.

The first index that we introduce is

- (1) *The asyllabicity index*. It is calculated as Syl_0/M , i.e. it provides information on the proportion of asyllabic morphemes in the language. The typological meaning of this index lies primarily in the fact that the absence of the asyllabic morphemes as Russian *-л*, *-к-*, English *-z* etc. is an important attribute of syllabic languages¹⁵.
- (2) *The polysyllabicity index*. This index is calculated as Syl_n/W , i.e. the ratio of the number of *n*-syllable words to the total number of words in the text. In other words, the proportions of one-, two-, three-, and four-syllable words are calculated separately. This index is also known as the word length¹⁶.

The sense of introduction of this index is seen from below. It follows from some well-known reconstructions¹⁷ that most words in pre-Turkish language were likely to be one-syllable. This fact characterizes this language in a very definite way, and not only phonologically and morphonologically, but also morphologically (see below for more detail). Inversely directed processes were also attested in special literature. For example, some continental Austronesian languages lost their typical two-syllable morpheme structure under influence of their monosyllabic neighbors (e.g., *guru* → *kru*,¹⁸).

- (3) *The word depth index*. This index is determined as M/W , i.e. as the ratio of the number of morphemes to the number of words (cf. the synthesis index introduced by J.Greenberg, and also Москович¹⁹).
- (4) *The agglutination index*. It has been already discussed (see Introduction). The treatment of agglutination, given above, suggests that with the increase in its index, the use of synonymous affixes increases while, respectively, the representation of invariable

¹⁴ В.Б. Касевич (ed.), *Грамматика и семантика восточного текста: количественные характеристики*, РХГА, Санкт-Петербург 2011.

¹⁵ В.Б. Касевич, *Фонологические проблемы общего и восточного языкознания*, Наука, Москва, 1983.

¹⁶ В.А. Москович, *Глубина и длина слова в естественных языках*, in: *Вопросы языкознания*, Наука, Москва 1967, no. 6. pp. 17–33.

¹⁷ А.М. Щербак, *Очерки по сравнительной морфологии тюркских языков (глагол)*, Наука, Ленинград 1981.

¹⁸ В.Б. Касевич, *Фонологические проблемы общего и восточного языкознания*.

¹⁹ В.А. Москович, *Глубина и длина слова в естественных языках*.

or allomorphologically variable affixes decreases. The role of this index largely determining the morphological build-up of the language is self-explanatory.

- (5) *The inflexion index.* The meaning of this index was already explained in Introduction. It is essential that all introduced indices appear to be logically interconnected. Indeed, if each auxiliary morpheme conveys exactly one meaning (polysemy apart), then “supplementing” each additional meaning within the framework of a word suggests addition of a morpheme. If such auxiliary morpheme is one- or more than one-syllable, then both the depth and the length of the word increase. If the morpheme is asyllabic, then only the depth of the word increases, while its length is maintained.
- (6) *The standard word-order maintenance index.* Unlike the previous ones, this index is important for the syntax characteristics of the language. In fact, two “sub-indices” are implied here, which further may be treated separately. The first of these is the basic syntactic structure of a sentence (SOV or SVO); here, the index takes the form $\text{ORDN}_{\text{nat}}/\text{SENT}$, i.e. is the ratio of the number of sentences with the “natural” word order to the total number of sentences. The SOV, SVO, and VSO structures reflect the historical and genre stability of the parameter. It is known that the degree of its stability in different languages and genres may differ. For example, in the Burmese language, the maintenance of the basic SOV structure is an inviolable law; on the other hand, in modern Turkish poetry, the dislocation of the verbal predicate to the initial position is far from being an exception. The second “sub-index” is the type of marking of syntactic relations in noun phrases of the NN type. According to Joan Nichols [1986], the type of this marking (head-marking vs. dependent-marking) appears to be typologically relevant for very different languages and, given the type, one can predict a number of typological features of the language, which apparently do not seem to be related either to this type or between themselves. Formally, this index is expressed as NN/SENT (with necessary diacritics that specify the type of marking).
- (7) *The vernacular vocabulary index (here “borrowed/non-borrowed vocabulary index”).* It is calculated as Frn/W , i.e. as the ratio of the number of borrowed words (foreignisms, i.e. Persian & Arabic borrowings) to the total number of words in the text. Here we leave the sphere of grammar and enter the area of *vocabulary parameters*. Strictly speaking, this index is not formally typological, however, the characteristics that it conveys is far from being useless.

Language Material for the Testing of the Technique of Indexical Typology in Diachrony

Here, we use the term “language material” in the sense of Lev Shcherba’s: it denotes a collection of texts the analysis of which makes it possible to “discover” the language system in question or its fragment. For such analysis, we selected two parallel texts: one in the New Ottoman language and the other in Modern Turkish. The text itself is the novel entitled “The Awakening” (*İntibah*) written by Namik Kemal. Most scholars

consider this book, first published in 1876, to be one of the first novels in the history of Turkish literature.²⁰ The modern text is a translation from that in Ottoman, with the time distance between the texts of about 100 years. It seems justifiable to compare the two closely related idioms of such kind. Although the time span between the idioms is unlikely to be sufficient for demonstrating clear typologically significant differences between them to have appeared, indexical typology may prove to be a sensitive tool to detect trends in typological development of the languages. In addition, there is an idea that Turkic languages vary relatively rapidly as compared to those of another type and genetics, and that they may be to a certain extent opposite to such languages as Icelandic or Lithuanian (the latter, according to Saussure, have remained virtually unchanged over the latest 500 years).

In the current special literature, the term “Turkish language” denotes the modern official language of the Republic of Turkey declared in 1923 as a result of the national liberation movement and the decay of the Ottoman Empire after the First World War.

Before the 20th century, the literary language of the Ottoman Empire was Ottoman. It featured the abundance (up to 90%) of lexical and syntactic borrowings from Arabic and Persian. This literary language (Ottoman) began to form on the verge of the 14th and 15th centuries on the basis of the Old Anatolian Turkic language (the language of Oghuz-Seljuk Turkic tribes, which lived in the Central Asia, but were in the 10th century driven out by other Turkic tribes, the Uighur). The Old Anatolian Turkic had been formed by the 11th-12th centuries, when Oghuz tribes from the Central Asia finally migrated into Asia Minor. The new language combined the Oghuz language of the Central Asia and the language of mixed Turkic population of Anatolia.²¹

The literary Ottoman also substantially affected the everyday Turkic conversational language, which also contained a number of borrowings – Arabic and Persian syntax constructions, foreign to the structure of Turkic languages.²²

After the establishment of the Republic of Turkey in 1923, the process of purification of the Turkish language started: archaic Arabic and Persian words were changed for indigenous Turkish ones. The latter were frequently created by linguists using Old Turkic languages. To this end, in 1932 the state-run “Turkish Linguistic Society” was founded, aimed primarily at “turkization” and modernizing of the Turkish language. This aim is still being pursued, and the “purification” of Turkish continues.²³

As a result, different generations sometimes can hardly understand one another. Moreover, the youth that uses the modern Turkish language finds it difficult to read books written even in the 1930–1940-ies, leaving alone those written in the 19th century.²⁴

²⁰ A. Kabaklı *Türk Edebiyatı*, 2. cilt, Türk Edebiyatı Vakfı Yayınları, İstanbul 1975, p. 561.

²¹ Э.Р. Тенишев (ed.), *Языки мира. Тюркские языки*, Индрик, Москва 1997. pp. 116–118.

²² J. Nemeth, *Zur Kenntnis der Mischsprachen (das doppelte Sprachsystem des Osmanischen)*, in: *Acta Linguistica Hungarica*, III, 1–2. Budapest 1953, pp. 153–159.

²³ *Языки мира. Тюркские языки*, pp. 394–395.

²⁴ See В.Б. Касевич, А.С. Аврутина, Е.В. Глазанова, *О количественной оценке сравнительной трудности восприятия разновременных текстов (на материале турецкого и османского языка) часть I*, in: „Вестник

For example, current publications of the works of the Turkish novelist *Sabahattin Ali*, who wrote in the 1930–1940-ies, are accompanied with abundant footnotes, in which outmoded words are explained using their modern equivalents.

A Comparative Analysis of Text Indices in Ottoman and Turkish Languages

For our comparative analysis, we have selected parallel fragments of the novel, identical in meaning in both languages, with the approximate size of 500 words. We calculated all the indices listed above. All the formulae are also given in the first part of the article.

From now on, the text in Ottoman will be denoted as “Text 1”, while that in modern Turkish – “Text 2”.

The calculated indices are presented in Table 1.

Table 1. Typological indices for Text 1 and Text 2

Index	Ottoman (Text 1)	Modern Turkish (Text 2)
<i>Asyllabicity</i>	0	0
<i>Monosyllabicity</i>	0.18	0.18
<i>Disyllabicity</i>	0.4	0.3
<i>Word depth</i>	1.54	1.8
<i>Agglutination</i>	1	1
<i>Inflexion</i>	0	0
<i>Vernacular vocabulary</i>	0.48	0.22

Below we will explain and comment the results presented in the table above.

In the forefront, we can see the indices that indicate the highest degree of stability. Such are, first of all, the indices that reflect basic typological parameters related to agglutination/inflexion. Analyzing the texts statistically, we did not use splitting the “agglutination/inflexion” feature into two partitions as suggested above. The reason for this is clear: in both texts, neither non-agglutinative nor inflective affixes are present. The synonymy of affixes and their “bi-functionality” are nonexistent either (at least, within the texts 1 and 2).

Equally high stability is shown by the indices related to parameters, which are, in essence, morphonological: to the absence of asyllabic and to the stability of the number

череповецкого университета” 2015, 5 (66), pp. 54–58; В.Б. Касевич, А.С. Аврутина, Е.В. Глазанова, *О количественной оценке сравнительной трудности восприятия разновременных текстов (на материале турецкого и османского языка) часть II*, in: „Вестник череповецкого университета” 2015, 6 (67), 62–66.

of monosyllabic words. In Text 2, the disyllabicity index is somewhat lower, while the monosyllabicity index is invariable; this is apparently explained by the decrease in the number of borrowings, which are mostly polysyllabic.

We cannot currently propose any hypothesis that could convincingly explain some increase in the word depth index that marks the transition from Ottoman to Modern Turkish (see the table).

The standard word-order index appears to be irrelevant to our analysis: in both texts, violation of the basic word order was observed in very few cases. Therefore, this index was not calculated.

Another isolated index, which may be called “attributivity”, is not shown in the table, either. The numerator of this index is the sum of the number of adjectives and that of dependent-marking constructions, while the denominator is the total number of sentences. The index is equal to 1.3 for Text 1 and 0.8 for Text 2. This difference is apparently due to style factors: the author of an archaic text often tends to “embellish” the nouns with a number of epithets, while the “modern style” gravitates towards certain minimalism.

Finally, the greatest (and predictable) difference between Text 1 and Text 2 is displayed in the values of the vernacularity index: 0.48 for Ottoman and 0.22 for Modern Turkish.

We admit that essentially the same results could have been obtained without the application of the relatively cumbersome technique of indexical typology. It is true, as well, that our data have not been verified with the special statistical criteria for showing their significance. Note, however, that our method was applied in order to estimate relations between idioms that are very close in every respect, and we aimed at testing this technique in a kind of a frontier situation. When this method is supplemented with statistical data processing, it may become a reliable and sensitive way to estimate relations between idioms in comparative studies.