IEEE TÜRKİYE SECTION

# UBMK'24

## Bildiriler Kitabı
## Proceedings

Editor Eşref ADALI

# 9. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı

# 9th International Conference on Computer Science and Engineering

26-27-28 Ekim (October) 2024 Antalya - Türkiye

# 9. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı (UBMK'2024)

# 9th International Conference on Computer Science and Engineering

26-28 Ekim 2024 Akdeniz Üniversitesi Antalya Türkiye
26-28 October 2024 Akdeniz University Antalya Türkiye



IEEE Catalog Number : CFP24L97-CDR

ISBN : 979-8-3503-6587-0

Additional copies may be ordered from:
Curran Associates, Inc
57 Morehouse Lane Red Hook, NY 12571 USA
Phone: (845) 758 0400
Fax: (845) 758 2633
E-mail: curran@proceeding.com

# The 80/20 Principle in Morphemics-Morphology in the Educational Corpus of the Uzbek Language

Shahlo Khamroeva
*Computational linguistic and digital technology*
*Tashkent State University of Uzbek language and Literature*
Tashkent, Uzbekistan
0000-0002-5429-4708

Bakhtiyor Mengliyev
*Department of Uzbek Language and Literature*
*Tashkent University of Applied Sciences*
Tashkent, Uzbekistan
0000-0002-6757-0476

Muyassar Kholova
*Department of Uzbek Linguistics*
*Termez State University*
Termez, Uzbekistan
0000-0002-7604-2958

Shakhnoza Gulyamova
*Departament of Pedagogy Turkiston*
*New Innovation University*
Bukhara, Uzbekistan
0000-0002-2045-1912

***Abstract*— Huge range of principles can be applied in various disciplines. To illustrate, the paradigm concept has been extended from the scientific sciences to the structural linguistics. The Pareto principle of economic sphere may also be used to language theory in order to distinguish qualitative and quantitative features within language layers. Pareto principle states that 20 percent of an effort produces 80 percent of the output, and 20 percent of an effort is efficient for 80 percent of the results. The study investigates how the Pareto principle is mirrored in the Uzbek language's morphologic and morphemic structures. It also demonstrates that the 80/20 concept may be applied to lexical categorization and word creation research. The word form suffixes had been carefully analyzed while speech context and word usage frequency observations were conducted utilizing the Uzbek educational corpus. It has been established that the most essential form suffixes influence 80 percent of morphologic structures as well as language vocabulary.Keywords—component, formatting, style, styling, insert**

***Keywords: Uzbek educational corpus, morphemic, morphology, formative suffixes, pareto principle***

## I. INTRODUCTION

Uzbekistan, a nation with diverse cultural backgrounds, officially designates the Uzbek language as its sole state language [1]. A.Pulatov, M.Ayimbetov, S.Mukhammedova, S.Karimov, A.Babanarov, D.Urinbayeva, N.Abdurakhmanova [2], A.Norov on computer linguistics Khamroyeva's thesis on "Linguistic bazes of formation of the author's corpus of the corpus of the Uzbek language", A.Eshmuminov's "The corpus of the Uzbek national language Synonym base" is one of the practical steps in computer linguistic [3]. We are to achieve this goal, we need to design elaborate taggest, wich wil encompass all linguistics properties of written and spoken texts wich a proper focus onsounds, segments, characters, words, morphemes, compounds, collocations, multiword units, phrases, sentences, idioms, provervs, and others linguistic properties [4]. To document this language and to determine how it is used in academic texsts, a corpus database needs to be collected together with a corpus platform to explore the corpus [5].

Grammar is description of the sentence construction methods of a particular language; a mathematical system that defines language. While describing grammar, we show the show the rules for emergence of a chain of characters belonging to that language. This is whu grammar can be called a language chain generator [6]. Linguistic modeling of polysemous morphem and words for a semantic analyzer lays the foundation for the development of theyr mathematical algoritms and the creation of a semantic analyzer in the furure [7]. In this study, we will consider the reflection of morphemic units in the educational corpus of the Uzbek language and the influence of its volume and analysis on the theoretical views that have developed in practice so far.

## II. METHODS

This article presents theoretical solutions in the Uzbek educational corpus of the Uzbek language using automatic analysis and statistical approach to express the position of the 80/20 Pareto rule in such morphotypes as nouns, adjectives, numbers, and verbs in the morphological structure of Uzbek linguistics

## III. RESULT

Morphemic polysemy is widespread within morphemic elements. This relates to the morpheme's polyfunctionality. As an example, consider the suffix -chilik. Textbooks identify three meanings for this suffix:

1. A noun indicating a field that produces what is understood from the bottom up: paxtachilik, urug'chilik, chorvachilik, uzumchilik (cotton growing, seed growing, cattle breeding, viticulture.)

2. A noun describing the state of a phenomenon or a thing: pishiqchilik, mo'lchilik, arzonchilik (abundance, prosperity, cheapness.)

3. A term indicating an action related with a fundamental principle: ulfatchilik, tirikchilik, dushmanchilik (friendship, livelihood, enmity) [8].

Polysemous morphemes do not have the same number or proportion of meanings as the words they are composed of. If any within the above meanings are extensively and actively employed, the pattern of word production becomes less fruitful. We shall check the frequency of use of some of the listed examples in the Uzbek language corpus.

TABLE I. THE FREQUENCY WITH WHICH THE MEANINGS OF THE POLYSEMIC MORPHEME -CHILIK ARE USED IN THE UZBEK EDUCATIONAL CORPUS

| № | Word | Search results | A link to the search result in the corpus |
|---|---|---|---|
| 1 | *paxtachilik* | 1236 | https://uzschoolcorpara.uz/uz/Search?c=1&q=paxtachilik |
| 2 | *urug'chilik* | 1841 | https://uzschoolcorpara.uz/uz/Search?c=1&q=urug%27chilik |
| 3 | *chorvachilik* | 7143 | https://uzschoolcorpara.uz/uz/Search?c=1&q=chorvachilik |
| 4 | *arzonchilik* | 32 | https://uzschoolcorpara.uz/uz/Search?c=1&q=arzonchilik |
| 5 | *pishiqchilik* | 128 | https://uzschoolcorpara.uz/uz/Search?c=1&q=pishiqchilik |
| 6 | *tirikchilik* | 1305 | https://uzschoolcorpara.uz/uz/Search?c=1&q=tirikchilik |
| 7 | *ulfatchilik* | 85 | https://uzschoolcorpara.uz/uz/Search?c=1&q=ulfatchilik |
| 8 | *dushmanchilik* | 11 | https://uzschoolcorpara.uz/uz/Search?c=1&q=dushmanchilik |

As shown in the Table I, the suffix -chilik "the noun representing the field that grows what is understood from the base" is widely employed in speech (cotton farming 1236 times, seed farming 1841 times, and cow breeding 7143 times). The usage meaning "noun denoting the work related to the concept understood from the base" is quite uncommon. This meaning-making pattern appears infrequently: patriotism 85 times and enmity 11 times. If we use the average arithmetic approach to compute the total number of productive and unproductive types of formation, we can estimate the probability of polysemous compound word formation as 80/20. "The task of morphological analysis is to automatically determine which category each word in the text belongs to; is to determine which lexical-grammatical class the words belong to [6].

Hojiev's claims that notion of "the majority of morphemes are polysemic in nature" is not supported by Uzbek language materials . To back up his claim, he provides examples of morphemes that generate personal nouns. "In the modern Uzbek language, the morphemes -chi, -shunos, -kor, -soz, -furush and -xo'r produce personal nouns, which all mean "person": dutorchi (dutar player), huquqshunos (lawyer), tadbirkor (businessman), yo'lsoz (road constructor), do'ppifurush (hat seller), norinxo'r (norin lover). So, the primary function of these morphemes is to generate words, and what distinguishes them from other word-forming morphemes is the ability to form personal nouns from the noun family. They have no other responsibility than that. It is not feasible to defend and establish that they are polysemantic using this function [9]. A.Hojiev adopts the absolute opinion that "Most morphemes are polysemic in nature". However, the word "majority" appears in this conclusion as well. The word "most" signifies "in most cases (80 percent) it shows the nature of ambiguity." The rationale advanced by A. Hojiev relates to the 20 precent share in the 80/20 law. As stated in the previous section, it is not need to be 80/20. The fact that "most" linguistic phenomena bear this symbol shows the prevalence of polysemic phenomena in Uzbek word-forming suffixes. So the 80/20 rule applies here as well. The Uzbek language follows a regular word order. The arrangement of morphemes in word forms varies depending on word group. This issue will be discussed more below.

### *A. Composition of noun word forms*

According to A. Hojiev, the origin of artificial words can be a root or an artificial term. The word form (the basis for its development) can also be derived from root or manufactured words, such as terimni (terim+ni) and terim-chini (terimchi+ni). However, because word forms of the noun family are not produced from word forms, word forms are not employed to form new forms. For example, the basis for forming a form in the word "terimchi" (picker), with formators terimchi+lar (pickers), terimchlar(i)miz (our pickers).

Also, in the word books, the portion -lar produces the plural form of the word book, -(i)m is the possessive form, and - is the accusative form - all are form-makers, and the book is the foundation for constructing a form, but not a word. That is why, first and foremost, the plural, possessive, and agreement forms of my books' word form (the makers of these forms) have no effect on one another's meaning. Second, if any form inside it is removed, the precise meaning of that form is lost: kitoblarim, kitoblar, kitob, kitobim, kitobni... [9]

According to the scientist, any derived term is formed based on its dictionary definition. The word's form is determined by its categorical meaning, "subject": plural, possessive, and agreement forms are produced from nouns. A word is made up of one basis of word formation and one word builder, whereas word forms related to the noun group might have one base of formation and one or more form builders. When there is more than one formative, each one creates a distinct, independent form of the noun from the formative base. For instance, in word form kitoblarimni each form -lar, -(u)m, -ni builder creates a distinct noun form, including number, possessive, and agreement forms. As a result, it is impossible to generate a categorical form of a word from a noun form [9].

The noun group's word form is composed in the following order: noun, plural, possessive, and agreement. This scenario applies to practically all noun forms. However, there is another order in Uzbek nouns (which is rarely used): Noun + Possession + Plural + Agreement: I saw my sisters. In this situation, the possessive shifts position and precedes the plural suffix, implying respect rather than ownership. Compare my sisters (she has numerous sisters) with my sisters (one sister is used for sizzling). We shall do a statistical analysis of the amount of this application using the corpus materials.

TABLE II. THE FREQUENCY OF THE WORD FORM "OPALARIM" AND "OPAMLAR" IN THE EDUCATIONAL CORPUS OF THE UZBEK LANGUAGE

| № | Word | Search results | A link to the search result in the corpus |
|---|---|---|---|
| 1 | Opalarim | 55 | https://uzschoolcorpara.uz/uz/Search?c=1&q=opalarim |
| 2 | Opamlar | 13 | https://uzschoolcorpara.uz/uz/Search?c=1&q=opamlar |

As seen in the Table II, the form "Noun + plural + possessive + agreement" (my sisters) was used 55 times, whereas the form "Noun + possessive + plural + agreement" was used 13 times. Looking at the percentage, there are several application situations. Grandmothers (plural grandmothers) appear 58 times, with the form grandmothers (respective form) appearing three times . In this situation, the form "Noun + plural + possessive + agreement" is used several dozen times more frequently than "Noun + possessive + plural + agreement." However, this is not the case with the related terms dad and mom. As a result, the rule that "the order of word forms in Uzbek language nouns is fixed" is not absolute: the 80/20 rule applies to this order as well.

The order of word forms in the verb phrase. The composition of the verb form is usually as follows: root + word form + lexical form + syntactic form. For example: jon+lan+tir+il+di (reanimated) In most circumstances, the lexical form generator follows the word-former, but this order is not essential. The order of stem + lexical form generator + word-former may also be seen in nouns generated from the verb stem: o‘tkazgich, isitkich (transmitter, heater) So, the 80/20 principle also applies to word order.

## IV. DISCUSSION

Word classification concepts. Words are classified into groupings such as nouns, adjectives, and verbs based on their lexical status (consisting of an indivisible unity of sound and meaning) and grammatical traits (morphological and syntactic function). Each word combines lexical and grammatical features. Thus, such groups are referred to as lexical-grammatical categories. The categorization of lexical-grammatical categories (word groups) of words is based on three principles: morphological (1), lexical-semantic (2), and syntactic (3) [10]. “The years in the usage of linguistic corpus made it much easier for the linguists to learn the presence of observation, sample / evidence collection, comparison, and conclusions facilitated this study” [11].

Relying on a morphological feature entails categorizing words according to their formal-grammatical signals and morphological characteristics. These are morphological indications, such as getting or not receiving word-changing affixes, change-invariance (whether number, possession, agreement, or person-number affixes change), type and declension, and so on. Words in the noun category, for example, contain properties like as number change, possessive affixes, and agreement modification. Verbs are inflected by adding person-number affixes (pure verbs only accept short versions of adjectives: I went. Conjugate: I went, I'm a student [12].

In speech, using one instead of the other is a regular normative scenario that arises as a result of the context, grammatical position, and does not alter the appearance. This form of transfer is called a conversion. For example, there is always one excellent and one terrible, yor-yor (song). Good and terrible words (their usual syntactic purpose is to appear as a determiner or participle) bir yaxshiga bir yomon har yerda bor, yor-yor (song). Words are classified into categories using all of the indicators listed above. As a result, words are classified into lexical-grammatical groups based on a certain set of indicators. Limiting a distinction to one side contradicts the inclusive character of language.

For example, if we are only based on the signification feature (semantic side), the adjective and adverb remain the same (the adjective indicates the sign of the object, the adverb indicates the sign of the sign), while the formation of adverbs from words in the rigid form (we can learn this by their use in speech) in some cases, lexical-grammatical signs such as facultatively, partial variation, different degrees of separation from the paradigm, formation from the adve [13].

If we only rely on meaning, we would add words like suvoq (the building is almost finished: only the plaster remains) and o‘roq (mowing) to verbs because they are related to the meaning of action, but these are nouns (action nouns), because a verb must also have a formal sign associated with it. Formally, based only on the sign of change-invariance, only on the morphological side, noun and pronoun (in the form of a noun), adjective and pronoun (this, similar pronouns), adjective and auxiliary remain the same (according to signs such as variation, lack of formal indicators), but they differ sharply from each other. The relationship to the syntactic situation is particularly striking in that it is difficult to confine oneself to a one-sided indication. For example, the word oy (in the meaning of "luna") is a noun (the moon is floating in the sea of the sky). Its conventional grammatical role is to become a possessive, complement, or object. However, it has become a determiner in the conjunction.

1) oy yuz (moon-faced creature) appeared in the regular grammatical role of the adjective;

2) In the example "I'm lonely, the front is gone, let the moon go and survive" (Song), the ravish presented a task.

If we merely examine the grammatical context, the first of them should be classified as a noun, the second as an adjective, and the third as adverb. However, this term is a noun, and its qualities are distinct from those of adjectives and adverbs. As a result, the syntactic sign cannot be the determining factor in word categorization. The 80/20 concept also applies for classifying words. A word's morphological and syntactic properties are used to categorize it. Correct classification is only feasible if the 80/20 principle is applied when classifying words.

While researching the primary and secondary functions of words, A. Botirova stated, "These criteria are connected with the problem of word development in the context of the emergence, formation, and development of the language, on the one hand, and with the polyfunctionality of words, on the other." In other words, diachronically, in addition to the occurrence of certain words leaving their categories and transferring to other categories, the potential of entering at least two paradigms in a given synchronic stage of the language demonstrates that these criteria cannot always be applied concurrently. As a result, it is appropriate to focus on the primary and secondary character of word groupings .

As A. Botirova correctly said, the semantic, syntactic, and morphological criteria for categorizing words in linguistic materials are insufficient to place any word in a category. While the linguistic system's rigidity increases its effectiveness, it also creates a variety of scenarios that impede interpretation and description. First, a homonymous word might fall into more than one group. For example:

1. "What should I do now?" Otabek questioned himself (self-pronoun).
2. The small girl began dancing spontaneously (adverb).
3. Everyone is interested in the country's fate.

Initially, kasal, issiq, sovuq, salqin, nam, ko‘k (sick, hot, chilly, cool, wet, and blue) are adjectives as well as nouns. Adjective+noun collocations: hot bread, cold air, cool place, humid climate, green tea; noun: issiq ta’sir qilgan, sovuq urgan, salqin tushdi, nam tushdi, ko‘k yuzini bulut qopladi (affected by heat, struck by cold, fell cool, fell wet, a cloud covered his blue face). One of the most serious issues in linguistics is the homonymy or polysemy of these terms.

Second, there are occasions where a term transitions from one category to another with one or more of its current meanings. For example, when the term road is used repeatedly, it becomes an adjective. The imperfect verb has joined the group of conjunctions. A substantial number of auxiliary and modal words are derived from words that have shifted to this category and have the same meaning .

Third, word groupings can execute functions for one another. For example, in the combination of iron gate, cast-iron stove, wooden spoon, golden autumn, the noun serves as an adjective, but in the combination of cry blood, repetition, terror is prepared.

Fourthly, while fulfilling its job in the speech process, the word may momentarily take the role of another word group. This is riding. For example: The noun that is linked to a noun gets dropped. A good (man) has his head, whereas a bad (man) has his. Nouns appear to possess all noun properties. These include grammatical form, syntactic function, and semantic characteristics.

## V. Conclusion

As a result, in certain circumstances, the word group's primary purpose is 80 percent, while the secondary function is 20 percent. A number may only be employed as a predicate; nevertheless, the conjunction can occasionally be used instead of the superlative, allowing it to accomplish its purpose. Auxiliaries bilan, -u, -yu, and -da serve as equal conjunctions. These demonstrate that word arrangement adheres to the 80/20 rule of meaning and function.

The 80/20 rule in grammatical forms. The plural suffix denotes the plural in 80precent of occurrences and the singular in 20 precent of cases. For example: Dadam keldilar, oyim aytdilar, Navoiy Hazratlari aytadilar, Usmon Hazratlari marhamat qiladilar (My father arrived, my mother announced, His Holiness Navoi states, His Holiness Usman blesses, and the -lar suffix expresses respect rather than pluralism. Bolalar o‘qiyaptilar, ishchilar dam olyaptilar, talabalar ta’tildan qaytdilar Children are studying, workers are resting, and students have returned from vacation, there use the pure plural form is evident.

The single number indicates the singular in 80 precent of circumstances and the plural in 20 percent of cases. For example, Qushlar issiq o‘lkalarga uchib ketdi, paxtalar ochildi, mevalar g‘arq pishdi (Birds went to hot places, cottons opened, and fruits drowned, the participle is singular but the meaning is multiple. In 80 percent of situations, the principal agreement reflects its purpose, whereas 20 precent expresses the function of subsidiary agreements. For example, kitob o‘qidim, magazin bordim, Navoiy g‘azali – I read a book, looked through a magazine, and read Navoi's ghazal. In general, zero forms in all grammatical categories function as marked forms in 20 precent of the situations.
Ye kızma

## References


[1] G. Rahmanova, G. Ekşi, (2023). English-Medium Instruction in Higher Education in Uzbekistan: Views on Effectiveness, Career Prospects and Challenges. Jour. World Journal of English Language, 13(5), P. 458. doi:http://dx.doi.org/10.5430/wjel.v13n5p458J

[2] A. Gatiatullin, D. Suleymanov, N. Prokopyev, N. Abdurakhmonova, (2024) “Turkic Morpheme”: From the Portal to the Linguistic Platform Lecture Notes in Networks and Systems, 912 LNNS, pp. 181 - 189, DOI: 10.1007/978-3-031-53488-1-22

[3] M. Kholova, (2020). Peech genres of texts and passport mete-metrics in the dialect corpus (Baysun district on the example of the sound “J” dialects). Journal of Advanced Research in Dynamical and Control Systems, 12(6), 1103-1107.

[4] D. Niladri Sekhar, L. Ramamoorthy, (2019) Utility and Application of Language Corpora Pp. X. Pp. 308. Publisher: Springer Singapore, https://doi.org/10.1007/978-981-13-1801-6 9.

[5] A. Kwary. (2019) A corpus platform of Indonesian academic language, Jour.: SoftwareX. Vol. 9, ISSN 2352-7110, Pages 102-106. https://doi.org/10.1016/j.softx.2019.01.011.

[6] B. Mengliyev, S. Shahabitdinova, S.Khamroeva, S. Gulyamova, A.Botirova, A. (2021) The morphological analysis and synthesis of word forms in the linguistic analyzer. Journal of Language and Linguistic Studies, 2021, 17(1), p.p. 558–564

[17 B. Mengliev, S. Gulyamova, Linguistic Modeling of Polysemous Words for the Semantic Analyzer of The Uzbek Language "Proceedings - 7th International Conference on Computer Science and Engineering" UBMK 2022, 2022, Pp.518–522.

[8] R. Sayfullayeva, B. Mengliyev, G. Bakiyeva, (2007) Modern Uzbek language. Study guide. – Tashkent: UzMU – B.146.

[9] A. Hojiyev, (2010) Theoretical issues of morphology, morphemics and word formation of the Uzbek language. – Tashkent: Science. – P.61. - Hojiyev, A. (2010) O‘zbek tili morfologiyasi, morfemikasi va so‘z yasalishining nazariy masalalari. – Toshkent, Fan. – B.61.

[10] Grammar of the Uzbek language. Morphology. Volume I. – Tashkent: Science, 1976. – O‘zbek tili grammatikasi. Morfologiya. I jild. – Toshkent: Fan, 1976. – P 604.

[12] R. A. Karimov, B. R. Mengliev, The role of the parallel corpus in linguistics, the importance and the possibilities of interpretation International Journal of Engineering and Advanced Technology, 2019, 8 (5 Special Issue 3), P. 388.

[13] H. Nematov About the relation of verb proportions. Jour. Uzbek language and literature. – 1970. Vol. 3. P.37. - Неъматов Ҳ. Феъл нисбатларининг муносабати ҳақида. Жур. Ўзбек тили ва адабиёти. – 1970. No3. – Б.37-39.