TURKISH LOANWORDS IN THE CONTEMPORARY STANDARD ALBANIAN: A CORPUS-BASED STUDY

Author: Julia Linnik, undergraduate in Albanian studies, Saint Petersburg State University

Advisors: Alexander Rusakov, PhD, Maria Morozova, PhD

Introduction

When speaking about Balkan languages and cultures, one should not forget about the heritage of the Ottoman Empire, which is significant for all spheres in the Balkans.

The Ottoman Empire had brought to the Balkans a new administrative and military order, as well as religion and cultural particularities. In other words, new language and cultural context had appeared, new words were adopted and integrated into the structure of the Balkan languages.

Turkish loanwords in Albanian language can be divided into several groups: policy, administrative division of the state, military terms, kinship terms, etc. It is considered in fact that most of Turkish loanwords are used in conversational style and rarely occur in literary texts (Desnitskaya).

Purpose

Our research aims are:
(1) to survey the frequency of Turkish loanwords in modern standard Albanian;
(2) to define which connotation, or more precisely, shade of meaning do have some chosen ones.

Our final product is to be a database of Turkish loanwords found in the Albanian corpus, with their frequency and their distribution by semantic fields, based on classification of (Haspelmath, Tadmor 2009).

Hypothesis

We have formulated two hypotheses:

- Most frequent loanwords will probably be "cultural borrowings", related to cultural and religion spheres;

- Frequency of Turkish loanwords in modern Albanian texts is probably growing, because they are no longer intentionally excluded from the language.

Investigation

In the first stage we created a list of Turkish loanwords, extracted from (Latifi 2012), and corrected it on the basis of the dictionary (Dizdari 2005). The resulting database includes 892 words, divided into 24 semantic fields (Haspelmath, Tadmor 2009). Onward, we compared words' frequency and IPM (items per million) in different subcorpora of the Albanian corpus: fiction, non-fiction, and press.

At the second phase, we chose 10 Turkish loanwords which have non-borrowed synonyms in Albanian (see Table 2). We compared their frequency in subcorpora of texts created in two different periods: 1950-1989 and 1990-present time, in order to reveal possible changes in use of Turkish members of the doublet pairs.

Finally, we have checked translations of each word in the Dictionary of the Modern Albanian language so as to learn, which word in each pair has more extended meaning.

Conclusions

Firstly, distribution of loanwords by word class in Albanian has proved to be typologically not very surprising: like in most languages, nouns are borrowed most easily (in our database we have 806 nouns, only 3 verbs, 78 adverbs, 40 – words belonging to other parts of speech).

Secondly, the most frequent words turned out to relate to semantic groups "Religion and beliefs", "Social and political relations". This correlates with the conclusions made a half-century ago by Desnitskaya, and proves our first hypothesis. Among the most frequent groups are also "Food and drinks", "Emotions and values", while the least frequent are "Spatial relations" and "Miscellaneous Function Words" (see Tables 1 and 1a).

Thirdly, different trends are observed in the use of Turkish words and their indigenous synonyms. As we hypothesized, in some cases the frequency of the Turkish variant has grown in the texts from 1990-present time, in comparison with 1950-1989. However, the Albanian counterparts remain more frequent than borrowed words, because Turkish borrowings keep conveying conversational style, and their meaning is often much more specific and expressive than that of the indigenous words. In addition, we have found one exception from the mainstream: in the pair *odë - dhomë* (see Table 2) the Turkish variant's frequency is slightly lowering, along with limitation of its meaning to (1) chamber (office, section of Parliament), (2) guest parlor.

Word	Translation	Frequency 1950-1989	Modern frequency	IPM 1950-1989	Modern IPM
odë		27	12	69,41	9,25
dhomë	'room'	135	421	347,05	324,54
zarzavate		0	0	0	0
perime	'vegetables'	0	1	0	0,77
krevat		8	98	20,57	75,55
shtrat	'bed'	57	90	146,53	69,38
muhabet		2	22	5,14	16,96
bisedë	'talk'	97	167	249,36	128,74
llaf		3	24	7,71	18,5
fjalë	'word'	542	719	1393,34	554,26
mavi		0	0	0	0
blu	'blue'	19	23	48,84	17,73
sahat		2	17	5,14	13,1
orë	'hour'	235	388	604,12	299,1
behar		0	0	0	0
verë	'summer'	64	173	164,53	133,36
vakt		0	5	0	3,85
kohë	'time'	550	1155	1413,91	890,36
sabah		0	3	0	2,31
mëngjes	'morning'	120	266	308,49	205,05

Table 2

Perspectives of investigation

In the nearest future we are going to make the same research in the Corpus of Modern Greek and then compare our databases of Turkish loanwords in Greek and Albanian.

Perhaps, we will take the same text in Greek and Albanian translations and then compare the role of Turkish loanwords in each language.

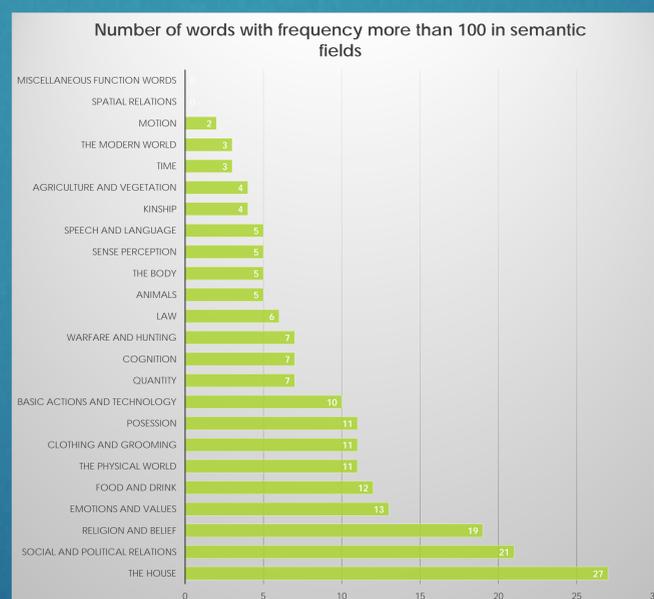| Semantic field	Example 'translation' (frequency in the corpus)
The house	kullë 'clock tower' (1332), kat 'floor' (1320)
Social and political relations	osman ' Ottoman' (4562), turk "Turk" (4086)
Religion and belief	allah 'Allah' (3163), mysliman 'Muslim' (2670)
Emotions and values	mall 'anguish' (1230), hall 'tough situation' (1014)
Food and drink	kafe 'coffee' (1635), futë 'apron' (495)
The physical world	xham 'glass' (904), bakër 'copper' (215)
Clothing and grooming	xhep 'pocket' (1003), çantë 'bag' (663)
Posession	borxh 'debt' (1581), pazar 'market' (1159)
Basic actions and technology	top 'ball' (2106), tel 'wire' (507)
Quantity	bile 'even more' (1426), çift 'pair' (1334)
Cognition	gjoja 'allegedly' (514), tamam 'excactly' (440)
Warfare and hunting	spahi 'a Turkish cavalryman: spahi' (240), pallë 'sword' (206)
Law	sanxhak 'province: sanjak' (568), saraj 'pallace' (369)
Animals	kara 'black horse' (177), majmun 'monkey' (174)
The body	topall 'lame' (406), damar 'vein' (160)
Sense perception	qejf 'enjoyment' (916), jeshil 'green' (258)
Speech and language	ja 'or..or' (1680), hajde 'let's go' (364)
Kinship	xhaxha 'uncle' (525), jetimë 'orphan' (286)
Agriculture and vegetation	çair 'pasture' (630), kafaz 'henhouse' (135)
Time	sahat 'hour' (209), behar 'summer' (153)
The modern world	dyqan 'shop' (935), ilaç 'shed' (384)
Motion	haxhi 'piligrim' (729), liman 'port' (203)

Table 1

Number of words with frequency more than 100 in semantic fields

Table 1a

Theoretical background:

1. M. Haspelmath, U. Tadmor. The Loanword Typology project and the World Loanword Database // M. Haspelmath, U. Tadmor (eds). Loanwords in the World's Languages: A Comparative Handbook, The Hague, De Gruyter Mouton, 2009, 1-34;

2. A. V. Desnitskaya. Albanskaya literatura i Albanskiy yazik [Albanian Literature and Albanian Language], Leningrad, Russia, 1987;

3. A. V. Desnitskaya. Albanskiy yazyk i ego dialekty [Albanian Language and its Dialects], Leningrad, Russia, 1968.

Data sources:

1. Albanian National Corpus (http://web-corpora.net/AlbanianCorpus/search/);

2. L. Xhanari (Latifi). Turqizmat dhe semantika e tyre në fjalorët e shqipes [Turkish loanwords and their semantics in Albanian dictionaries], Tirana, Albania, 2012;

3. T.F. Dizdari. Fjalor i orientalizmave në gjuhën shqipe [Dictonary of Orientalisms in Albanian language], Tirana, Albania, 2005.