

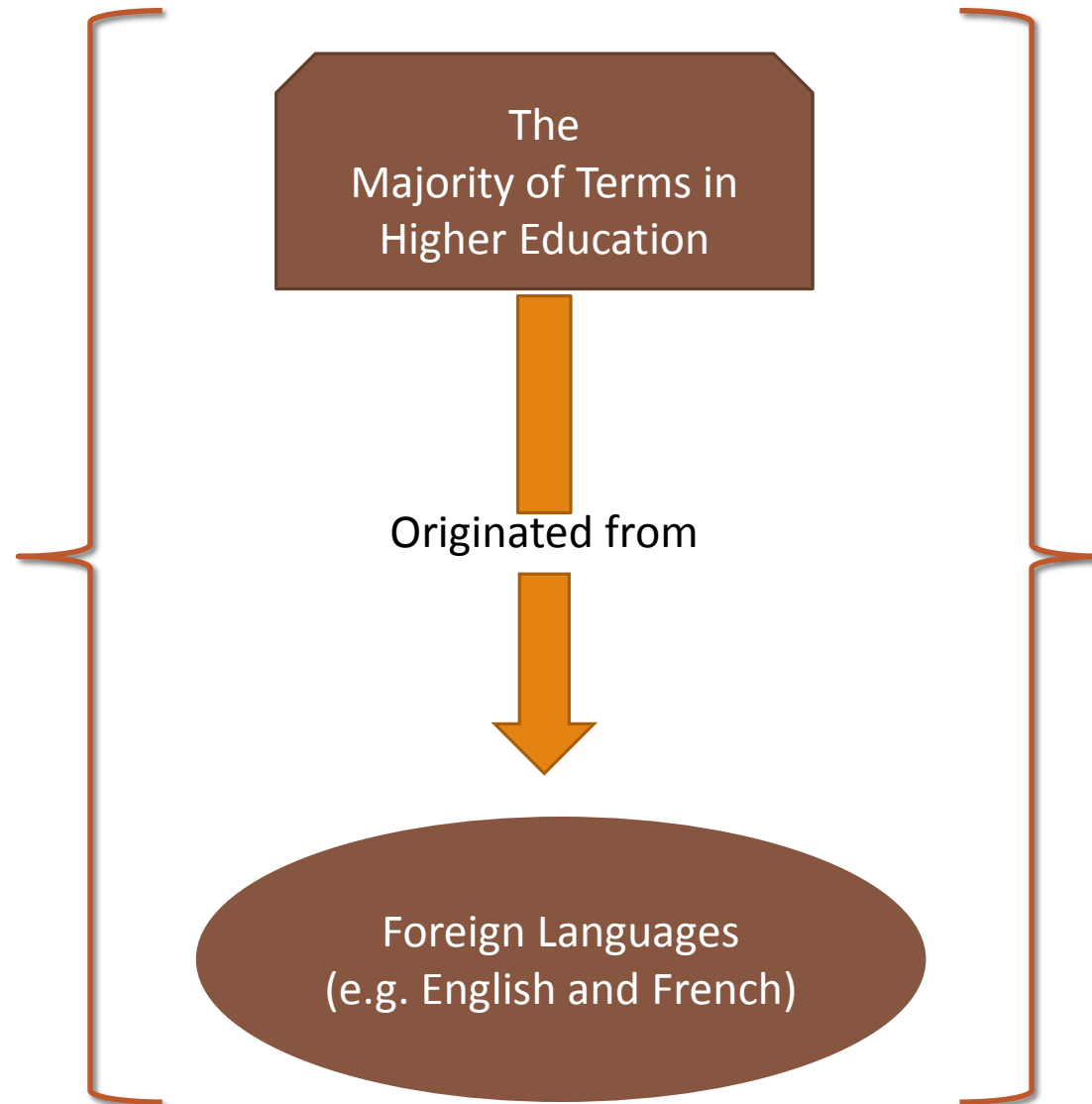


Understanding the Impacts of Multilingualism on the Turkish Language using Search Engines

Dursun AKASLAN & Umut KURAN

Harran University Engineering Faculty Computer Engineering Department
Sanliurfa Turkey





The Foreign Word Equivalence Manual

published by

The Turkish Language Organization

indicates

The utilisation of foreign originated terms are prevalent in subjects

Philosophy

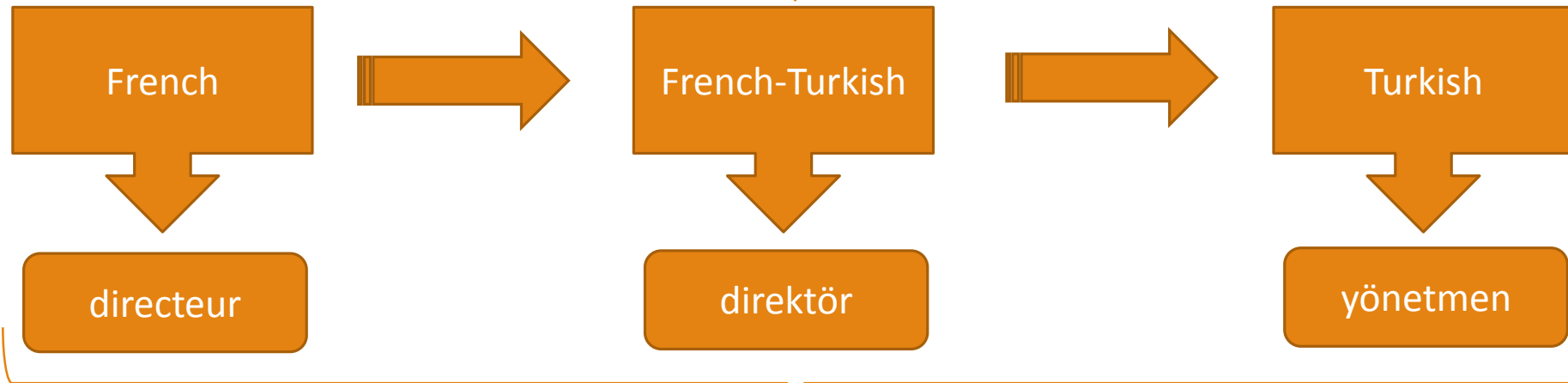
Physics

Medicine

Sport

Economy

The majority of foreign originated terms have their own equivalence

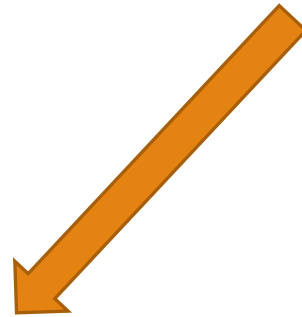


The simultaneous utilisation of foreign originated terms and their equivalence leads to

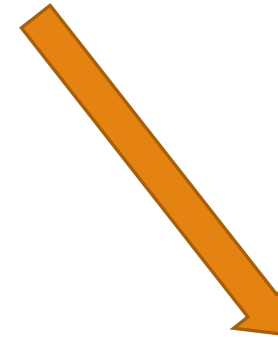
multilingualism



Hence, the goal of this research is to understand the impacts of multilingualism on the Turkish language in terms of the utilization of terms



Understanding the prevalence of the foreign originated terms and their synonyms using search engines



Investigating significant differences among terms in various disciplines using inferential statistics

The use of the Turkish
equivalence of the foreign
originated terms

has been perceived as

An important solution to
deal with the impacts of
multilingualism on the
Turkish language

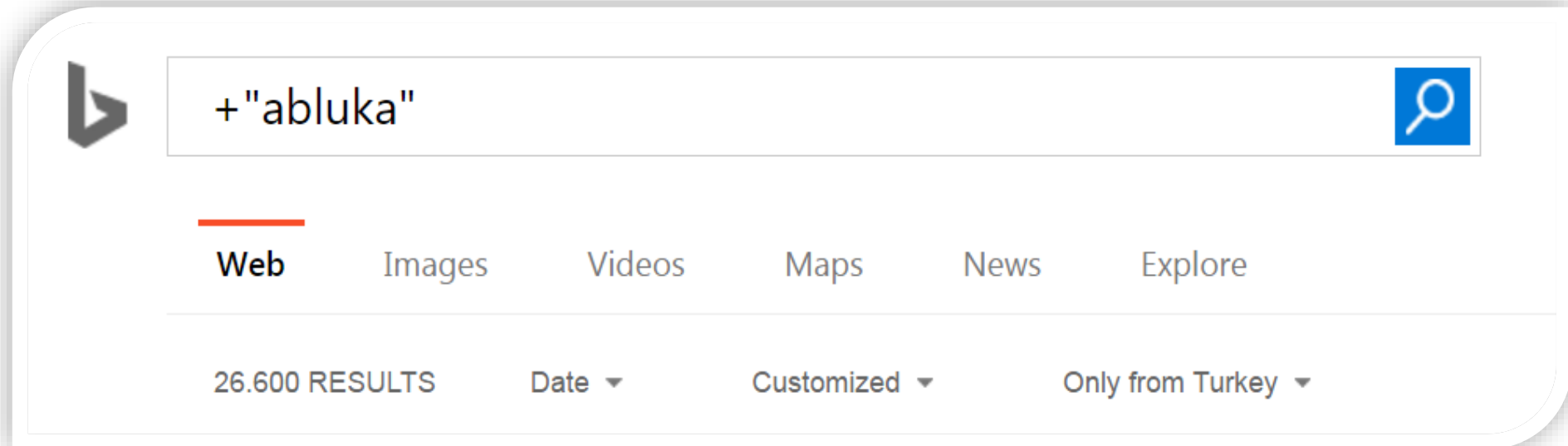
The Turkish Language Organisation made available the
equivalence of 1948 foreign originated terms in 2008 using
the Foreign Word Equivalence Manual (FWEM)

The FWEM will be used a dataset to
understand the impact of foreign
originated terms

DATASET					
id	equivalence	originated	root	original	discipline
1	sayı boncuğu	abaküs	French	abacus	mathematics
2	sapınç	aberasyon	French	aberration	astronomy
3	sapınç	aberasyon	French	aberration	sprit
4	çıkma durumu	ablatif	French	ablatif	grammer
5	kuşatma	abluka	Italian	abloco	military
6	sürdürümcü	abone	French	abonné	daily
7	sürdürüm	abonman	French	abonnement	daily
8	soğurma	absorbe	French	absorbé	physics
...

The FWEM made available the equivalence of foreigned originated terms with their disciplines and languages.

Search engines might be used to measure the number of the foreign originated terms and its synonyms on the Turkish language.



It is possible to filter and customize search results such as region (e.g. Turkey) and language (e.g. Turkish) using search engines such as Yandex and Bing

The number of foreign originated terms and its synonyms appearing in the search engines must be evaluated to understand the impacts of multilingualism

id	Turkish Equivalence			Foreign Originated		
	equivalence	Bing Search	Yandex Search	originated	Bing Search	Yandex Search
1	sayı boncuğu	12.700,00	12.000,00	abaküs	128.000,00	136.000,00
2	sapınç	1.560,00	15.000,00	aberasyon	10.700,00	21.000,00
3	sapınç	1.560,00	15.000,00	aberasyon	10.700,00	21.000,00
4	çıkma durumu	2.720,00	3.000,00	ablatif	6.030,00	14.000,00
5	kuşatma	31.300,00	535.000,00	abluka	27.100,00	492.000,00
...

An impact factor might be described as a measure reflecting the importance or power of a term within the Turkish language using search results

IMPACT FACTOR (IF)

The impact factor of a term might be calculated considering the number of search results returned by a search engine during all the time using its foreign and Turkish counterparts at the same time.

R is the ratio of the number of search results in foreign languages to the one in Turkish language or vice versa

$$IF = \frac{R}{C} * 5$$

where

The R should be multiplied with -1 or with +1 if vice versa due to its positive or negative effect on the Turkish language

C is a constant in order to standardize the impact factor of terms between -5 and +5

Constant (C)	Ratio (R)
10	-10 < R < +10
100	-100 < R < +100
1,000	-1,000 < R < +1,000
10,000	-1,0000 < R < +10,000
100,000	-100,000 < R < +100,000
1,000,000	-1,000,000 < R < +1,000,000

adaptasyon

229,000
times in Yandex
Search Engine

The word “adaptasyon” is a foreign originated term and is derived from the word “adaptation” in the French language and its equivalence is written as “uyarlama” in the Turkish language.

uyarlama

396,000
times in Yandex
Search Engine

$$IF = \frac{R}{C} * 5 = \frac{(396000 / 229000)}{10} * 5 = \frac{+1.7293}{10} * 5 = \mathbf{0.8646}$$

The R = +1.7293 because the Turkish counterpart of the term is higher than its foreign one, C = 10 because R is between -10 and +10

The impact factor might be used as an indicator for the relative importance of a term within the Turkish language.



However, the impact factor does not indicate its weight within the Turkish language. Hence, it is relatively important to classify the terms using levels namely A1, A2, B1, B2, C1 and C2.



Level	R	C
A1	$-10 < R < +10$	10
A2	$-100 < R < +100$	100
B1	$-1,000 < R < +1,000$	1,000
B2	$-1,0000 < R < +10,000$	10,000
C1	$-100,000 < R < +100,000$	100,000
C2	$-1,000,000 < R < +1,000,000$	1,000,000

The number and mean scores of the impact factors of terms within various languages are presented as illustrated below.

No. and Disciplines of Terms		A1		A2		B1		B2		C1		C2	
		Y	B	Y	B	Y	B	Y	B	Y	B	Y	B
<i>German</i>	23	-0,33 (14)	-1,35 (12)	0,70 (7)	0,91 (5)	0,71 (2)	0,00 (4)	-	0,75 (2)	-	-	-	-
<i>French</i>	1836	-0,25 (1092)	-0,09 (949)	-0,56 (567)	-0,50 (606)	-0,53 (154)	-0,63 (216)	-0,68 (22)	-0,50 (62)	-	-	-0,09 (1)	0,00 (3)
<i>English</i>	313	-0,04 (165)	0,16 (156)	-0,80 (100)	-0,84 (80)	-1,29 (42)	-1,08 (52)	-1,15 (6)	-0,99 (23)	-	-	-	0,00 (2)
<i>Spanish</i>	4	-0,02 (4)	1,40 (2)	-0,36 (16)	1,80 (2)	-	-	-	-	-	-	-	-
<i>Italian</i>	43	-0,49 (18)	-0,65 (16)	-1,28 (11)	-0,59 (15)	-0,74 (9)	0,19 (6)	-	-1,95 (6)	-	-	-	-
<i>Latin</i>	23	-0,82 (11)	-0,50 (13)	-	-2,02 (8)	-	-0,57 (1)	-	0,64 (1)	-	-	-0,07 (1)	-
<i>Romaic</i>	2	-	-	-1,91 (2)	-2,76 (1)	-	-3,02 (1)	-	-	-	-	-	-
<i>Russian</i>	3	-4,71 (2)	0,70 (1)	1,85 (1)	-1,51 (1)	-	1,58 (1)	-	-	-	-	-	-
<i>Greek</i>	18	0,00 (14)	0,05 (14)	-1,29 (4)	-2,27 (3)	-	-0,64 (1)	-	-	-	-	-	-
overall	2265	-0,22 (1320)	-0,07 (1163)	-0,59 (708)	-0,55 (721)	-0,68 (207)	-0,69 (282)	-0,78 (28)	-0,69 (94)	0,00 (0)	0,00 (0)	-0,08 (2)	0,00 (5)

The mean scores of the impact factors of the terms associated with their disciplines were also illustrated in the following table.

No. and Disciplines of Terms		A1		A2		B1		B2		C1		C2	
		Y (1320)	B (1163)	Y (708)	B (721)	Y (207)	B (282)	Y (28)	B (95)	Y (0)	B (0)	Y (2)	B (5)
<i>anatomy</i>	30	-1,00	-0,54	-1,60	-2,37	-	-	-	-	-	-	-	-
<i>anthropology</i>	1	-0,50	1,68	-	-	-	-	-	-	-	-	-	-
<i>architecture</i>	3	0,72	-1,02	-	0,59	-	-	-	-	-	-	-	-
<i>astronomy</i>	29	-0,66	-1,56	-1,05	-1,08	-	-1,23	-	-0,21	-	-	-	-
<i>biology</i>	35	-1,08	0,10	-0,75	-1,60	1,05	-0,55	-	1,17	-	-	-	-
<i>botany</i>	18	-0,28	-0,40	-0,76	-0,33	-	-	-	-0,63	-	-	-	-
<i>chemistry</i>	36	-1,09	-0,96	0,10	-0,85	-2,13	0,92	0,50	1,19	-	-	-	-
<i>cinema</i>	21	0,14	-0,91	0,16	0,36	-0,79	-3,53	-	1,69	-	-	-	-
<i>common</i>	1217	0,01	0,06	-0,56	-0,54	-0,62	-0,50	-	-0,88	-	-	-0,07	0,00
<i>economy</i>	70	-0,24	-0,50	-0,96	-0,73	-1,07	-1,48	-0,73	-0,64	-	-	-	0,00
<i>education</i>	4	-1,39	-0,85	-0,68		-1,88	-2,90	-1,31	-	-	-	-	-
<i>geography</i>	22	-0,94	-0,63	0,30	0,07	-0,70	-0,59	-	-	-	-	-	-
<i>geology</i>	17	-0,91	-0,92	-0,69	-0,31			-	-	-	-	-	-
<i>geometry</i>	1	0,83		-	-4,10			-	-	-	-	-	-

Conclusion

- The main goal of this study was to understand the impacts of the multilingualism on the Turkish language using the search results of the Yandex and Bing.
- The usage of the terms originated from foreign languages, especially English and French, and their Turkish synonyms were investigated associated with various disciplines and languages.
- A model for measuring the impact factor of the terms were proposed in the study.
- Furthermore, the terms were classified into levels namely A1, A2, B1, B2, C1 and C2 to specify their weight within the Turkish language.
- However, the dataset used in the study was limited with the 1948 terms and 1265 synonyms.
- Findings from this study can be further analysed to very significant differences between languages and disciplines.
- The impact factor and classification are also generalizable to other languages and disciplines.

Acknowledgement

I would like to express our gratitude to the Harran University for providing valuable support for the research. Turning these ideas and concepts given through the paper would not have been possible without the generous support of Haran University for the Scientific Research Project (No: 14169) on the Relational Database Design for Alleviating the Impacts of Foreign Words.

Assist. Prof. Dr. Dursun AKASLAN

Harran University Engineering Faculty Computer Engineering Department

E-mail: dursunakaslan@harran.edu.tr

Tel: +90 414 318 30 00 Ext: 1378

